

**USS | ESCUELA DE
POSGRADO**
UNIVERSIDAD SEÑOR DE SIPÁN

Escuela de Posgrado

TESIS

**“SISTEMA PREDICTIVO BASADO EN EL MODELO WEB
PARA LA PREDICCIÓN DE PREFERENCIAS DE USUARIO
EN EL SITIO WEB FOROS PERÚ”**

**PARA OPTAR EL GRADO ACADÉMICO
DE DOCTOR EN CIENCIAS DE LA COMPUTACION Y
SISTEMAS**

Autor:

Mg. García Merino Luis Santiago

ORCID: 0000-0001-9392-2474

Asesor:

PhD. López de Luise María Daniela

ORCID: 0000-0003-3130-873X

Línea de Investigación:

Infraestructura, Tecnología y Medio Ambiente

Pimentel – Perú

2021



UNIVERSIDAD SEÑOR DE SIPÁN

ESCUELA DE POSGRADO

DOCTORADO EN CIENCIAS DE LA COMPUTACION Y SISTEMAS

**“SISTEMA PREDICTIVO BASADO EN EL MODELO WEB PARA LA
PREDICCIÓN DE PREFERENCIAS DE USUARIO EN EL SITIO WEB FOROS
PERÚ”**

AUTOR

Mgtr. García Merino Luis Santiago

PIMENTEL – PERÚ

2021

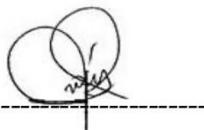
**“SISTEMA PREDICTIVO BASADO EN EL MODELO WEB PARA LA
PREDICCIÓN DE PREFERENCIAS DE USUARIO EN EL SITIO WEB FOROS
PERÚ”**

APROBACIÓN DE LA TESIS



CE 001170407

PhD. Juan Carlos Callejas Torres
Asesor Metodológico



Presidente del jurado de tesis
Dr. Jorge Luis Gutierrez Gutierrez



Secretario del jurado de tesis
Dr Christian Abraham Dios Castillo



Vocal del jurado de tesis
PhD. Maria Daniela Lopez de Luise

Dedicatorias

A Dios, por todas las bendiciones en lo personal, familiar y profesional que me brinda diariamente, y me ayuda estar en constante aprendizaje para procurar ser mejor persona y profesional en el día a día; y a mis queridos padres y hermanos por el apoyo brindado en todo momento en forma incondicional.

A mis padres por su esfuerzo, sacrificio que siempre han realizado con mucho cariño y amor para que pueda ser una buena persona y profesional, asimismo por todos sus consejos y sugerencias en todo momento.

Agradecimientos

A la PhD. María Daniela López de Luise mi asesora técnica especialista y al Dr. Juan Carlos Callejas Torres mi asesor metodólogo por su dedicación, paciencia y buenos consejos que siempre me han brindado y permiten el desarrollo del presente trabajo de investigación científica.

A mi colegas, amigos, amigas y promociones, por sus consejos y sugerencias para lograr concretizar el presente trabajo de investigación

RESUMEN

La presente tesis tiene el propósito brindar una alternativa de solución a las Insuficiencias en el procesamiento de lenguaje natural que limita la predicción de preferencias de usuarios en el sitio Web Foros Perú.

Su objetivo considera elaborar un sistema predictivo de preferencias de usuario en el sitio web Foros Perú, para extraer las palabras que predominan en un grupo de comentarios del sitio web Foros Perú.

El Modelo web para la predicción de preferencias de usuario apoya el procesamiento de lenguaje en el tiempo con el uso de nuevas tecnologías, para sistemas de mercadeo que incluyen variables socioculturales en el mercado. Entendiendo como mercado, el sistema total de actividades de negocios que planea, establece precios, promociona, distribuye productos y que satisfacen las necesidades de los clientes con el fin de lograr metas organizacionales. Gracias a la alta competencias nacionales e internacionales el uso de la tecnología proporciona una ventaja competitiva importante en la obtención de características de los perfiles de usuarios que generen un aprovechamiento de la oportunidad y un tiempo de respuesta acorde a la globalización; dando soporte a empresas permitiéndole aprovechar la continua interacción de los usuarios con redes sociales y centrándose en diferentes pautas que permiten predecir la intencionalidad de compra e interés basada en foros como es el sitio foros Perú; en tal sentido el proceso de validación de los resultados con el algoritmo de WEKA maximización esperanza, permiten comprender el significado de los mensajes de usuarios de todos los que desean vender ó comprar bienes, servicios, y ofrezcan una buena idea de realizar su negocio.

Como conclusión, modelo web permite que pequeñas y medianas empresas empleen el modelo basado en predicción de preferencias de usuario y tenga en cuenta la relación entre la clasificación binaria y la toma de decisión entonces se contribuye a conocer que comentarios escritos por un grupo de usuarios en internet y del sitio web foros Perú, son más resaltantes unos de otros respectivamente, a efectos de orientar la tendencia ó predicción de preferencias de usuario.

Palabras clave: modelo web, predicción de preferencias, procesamiento de lenguaje natural, maximización esperanza.

ABSTRACT

The purpose of this thesis is to provide an alternative solution to the shortcomings in natural language processing, limiting the prediction of user preferences in the website Foros Peru.

Its objective is to elaborate a predictive system of user preferences in the website Foros Peru, to extract the words that predominate in a group of comments of the website Foros Peru.

The Web Model for the prediction of user preferences supports the decision making process, in time and using new technologies, for marketing systems that include sociocultural variables in the market. Understanding as market, the total system of business activities that plans, sets prices, promotes, distributes products and satisfies customer needs in order to achieve organizational goals. Thanks to the high national and international competition, the use of technology provides an important competitive advantage in decision making and generates an advantage of opportunity and response time according to globalization; giving support to companies allowing them to take advantage of the continuous interaction of users with social networks and focusing on different patterns that allow predicting the intentionality of purchase and interest based on forums as is the site forums Peru; in such sense the validation process with WEKA algorithm of expectation-maximization of the results allow understanding the meaning of the messages of users of all those who want to sell or buy goods, services, and offer a good idea to do their business.

As a conclusion, web model allows small and medium companies to employ the model based on prediction of user preferences and take into account the relationship between binary classification and decision making then it contributes to know which comments written by a group of users on the Internet and Peru forums website, are more salient from each other respectively, in order to guide the trend or prediction of user preferences.

Keywords: web model, preference prediction, natural language processing, expectation-maximization.

ÍNDICE

(Página de Aprobación de tesis).....	¡Error! Marcador no definido.
Dedicatorias	iv
Agradecimientos	v
RESUMEN	vi
ABSTRACT.....	vii
ÍNDICE	viii
Indice de Figuras	xii
I. INTRODUCCIÓN.....	13
1.1. Realidad Problemática.	13
1.2. Antecedentes del Estudio: insuficiencias en el proceso de lenguaje natural, limita la predicción de preferencias de usuarios en el sitio Web Foros Perú.....	16
1.3. Teorías relacionadas al tema.....	19
1.3.1 Caracterización del Proceso de lenguaje natural y su dinámica en la predicción de preferencias de Usuario.....	22
➤ Etapa 2.- Evolución del Proceso de Lenguaje Natural (PLN), con la minería de opinión y preferencias, para lograr la predicción de preferencias de usuarios (aproximadamente 1980 a la actualidad).	25
1.3.2 Marco Conceptual.	35
1.4. Formulación del Problema.	47
1.4.1. Problema General.....	47
1.4.2. Problemas Especificos	47
1.5. Justificación e importancia del estudio.	48
1.6. Hipótesis General	49
1.7. Objetivos	50
1.7.1. Objetivos General.....	50
1.7.2. Objetivos Específicos.....	50
II. MATERIAL Y MÉTODO	50
2.1. Tipo y Diseño de Investigación.....	50
2.2. Población y muestra.	51
2.3. Variables.	52
2.4. Técnicas e instrumentos de recolección de datos, validez y confiabilidad.....	52
2.5. Procedimientos de análisis de datos.	52
2.6. Criterios éticos	53
2.7. Criterios de Rigor científico.....	53
III. RESULTADOS.....	54

3.1. Resultados en Tablas y Figuras	54
Diagrama de Bloques del Sistema Predictivo basado en la predicción de preferencias de usuarios en el sitio web Foros Perú.....	55
Mapeo de datos e información: (ver detalle en la Tabla 1 y Tabla 4).....	56
3.2. Discusión de resultados	69
Con respecto al Objetivos General.....	69
3.3. Aporte teórico.....	72
.....	72
3.3.1. Fundamentación del aporte teórico.....	72
3.3.2. Descripción Argumentativa del aporte teórico	74
3.4 Aporte práctico	75
3.4.1. Fundamentación del aporte práctico.	75
3.4.2. Construcción del aporte práctico	76
3.5. Valoración y corroboración de los resultados	77
CARACTERISTICAS DE LOS PERFILES DE USUARIOS ANALIZADOS CON EL SISTEMA PREDICTIVO BASADO EN EL MODELO DE PREDICCIÓN DE PREFERENCIAS DE USUARIOS DEL SITIO WEB FOROS PERÚ.	78
3.5.1 Valoración de los resultados (taller de socialización, criterio de expertos, etc.)	80
Valoración de los resultados por juicio de expertos.....	80
3.5.2 Ejemplificación de la aplicación del aporte práctico.....	84
3.5.3 Corroboración estadística de las transformaciones logradas	84
Significado de las palabras procesadas que corresponden a los comentarios procesados	85
IV. CONCLUSIONES.....	87
V. RECOMENDACIONES.....	88
VI. REFERENCIAS	89
Anexos.....	94
ANEXO 1.- MATRIZ DE CONSISTENCIA.....	95
Problema General	95
Problemas Especificos.....	95
ANEXO 2.- CUADRO DE OPERACIONALIZACION DE VARIABLES.....	102
ANEXO 3.- INSTRUMENTO.....	103
ANEXO 4.- INSTRUMENTO DE VALIDACION NO EXPERIMENTAL POR JUICIO DE EXPERTOS	104
ANEXO 5: FORMATO EPGUSS VISADO POR ASESOR DE TESIS ESPECIALISTA.....	110
ANEXO 6.....	111
RESULTADO CON EL ALGORITMO MAXIMIZACION ESPERANZA CON EL ARCHIVO EXPORTADO	111
EM -USUARIO.CSV	111

ANEXO 7	199
RESULTADO CON EL ALGORITMO MAXIMIZACION ESPERANZA DEL ARCHIVO EXPORTADO	199
EM – BLOG.CSV	199
ANEXO 8	203
Acta de Originalidad de la Tesis Doctoral EPGUSS	203

Indice de Tablas

Tabla 1.- Mapeo de Información y datos	56
Tabla 2.- Detalle de variables en archivo usuario	63
Tabla 3.- Datos agrupados según EM	65
Tabla 4.-Colores de las agrupaciones en la Figura 13.	66
Tabla 5.- detalle de variables en archivo blog	66
Tabla 6.- Detalle de Agrupaciones Naturales	68
Tabla 7.- Detalle de Variables en archivo Blog	77
Tabla 8.-Datos de los expertos	81
Tabla 9.- Resultados de calificación de expertos participantes	81
Tabla 10.- Calificación, Opinión, Evaluación de Expertos	82
Tabla 11.-. Colores de las agrupaciones en la Fig. 16	84
Tabla 12.-Colores de las agrupaciones a continuación en la Figura 17.....	86

Indice de Figuras

Figura 1.- Diagrama de Bloques del Modelo de predicción de preferencias de usuarios del sitio web Foros Perú	55
Figura 2.-Aplicación creada para analizar palabras.	57
Figura 3.- Interfaz de registro de temas de los comentarios analizado	58
Figura 4.-Interfaz de registro de comentarios analizados	58
Figura 5.-Registro y Control de Preposiciones de la aplicación	59
Figura 6.-Registro y Control de Artículos de la aplicación	59
Figura 7.-Registro y Control de pronombres de la aplicación.....	60
Figura 8.-Registro y Control de Terminaciones de parser (palabras analizadas)	60
Figura 9.-Registro y Control de Símbolos y Caracteres Especiales detectados en los comentarios analizados	61
Figura 10.-Interfaz procesar datos, muestra los resultados procesados y analizados de las palabras que componen los comentarios registrados en el prototipo desarrollado	61
Figura 11.-Prototipo para analizar las estadísticas WEKA, para realizar análisis estadístico inferencial	62
Figura 12.-Parámetros usados para el EM.....	65
Figura 13.-Asignaciones de agrupaciones	66
Figura 14.-Detalle de Parámetros empleados para EM.....	67
Figura 15.-Arquitectura propuesta para el sistema predictivo – Modelo Propuesto	72
Figura 16.-Modelo Propuesto – Aporte Práctico – Modelo Web	75
Figura 17.-Autor VS Cantidad de caracteres (tamaño comentarios) con espacios	79
Figura 18.-Autor VS Cantidad de palabras sin repetición	79

I. INTRODUCCIÓN

1.1. Realidad Problemática.

En la actualidad, a nivel mundial, el sistema económico va cambiando, y esto va ligado a las nuevas tecnologías, es por ello que los usuarios se van adecuando a los cambios tecnológicos que se presentan, Los modelos actuales cubren otros aspectos sin foco en el cliente, como, por ejemplo, la venta excesiva de sus productos sin existir alguna necesidad. Es por ello que se requiere desarrollar una alternativa enfocada a recuperar a los clientes potenciales que tienen tendencia a desertar.

Según AmirGandomi y MurtazaHaider, (2017) Indica que las técnicas de análisis predictivo se basan principalmente en métodos estadísticos. Varios factores requieren el desarrollo de nuevos métodos estadísticos para Big data. En primer lugar, los métodos estadísticos convencionales tienen una significación estadística: se obtiene una pequeña muestra de la población y se compara el resultado con la posibilidad de examinar la importancia de una relación particular.

Por otro lado, (Luis Rendeules, 2020) Muestra que la aplicación de moda en todo el mundo. Más de 800 millones de usuarios y en España, TikTok ya es la cuarta más descargada. Al menos tres millones de ciudadanos la tienen en sus móviles para crear y compartir vídeos divertidos y cortos con los que pueden obtener seguidores. TikTok es propiedad de una compañía china que prohíbe abiertamente comportamientos de acoso o abuso hacia los menores, tampoco tolera los vídeos de menores desnudos, pero la Policía española ya ha detectado que pederastas y pedófilos han comenzado a buscar allí nuevas víctimas. El problema no son los vídeos que cuelgan allí los jóvenes bailando, pasándose bien; y ellos no le dan importancia, no le ven un sentido sexual. el problema radica en la forma de pensar y sentir de algunos que lo están viendo. Por lo que es preciso tener en cuenta que jamás hay que mandar fotos o vídeos a desconocidos, teniendo en cuenta que los niños entregan su confianza muy rápido, y tienden a compartirlo todo por redes sociales, lo que algunas veces puede poner en riesgo su seguridad personal y de sus familias, por la emoción o sentimiento que da a los usuarios ser seguidores activos o menos activos a lo que ven o leen en los comentarios y mensajes de las redes sociales de internet, generando buena o mala conducta como personas

En tal sentido, el fenómeno de redes sociales, foros de internet, comunidades virtuales pasó de ser un medio de interacción, que demuestra sentimientos, emociones positivas o negativos (polaridad) para pequeños, jóvenes y adultos a ser un canal de comunicación que las empresas deben considerar seriamente. En el estudio titulado “Social Media Benchmark Study 2013” dice que el 67% de los consumidores utilizan los perfiles en redes sociales de una empresa para obtener información sobre los servicios y que el 43% de los jóvenes de entre 18 y 29 años, prefiere interactuar con las marcas a través de las redes sociales, que utilizar otro tipo de comunicación. En definitiva, las redes sociales son un mercado en que las organizaciones deben invertir tiempo y esfuerzo. (Abbasimehr, 2018)

Por otro lado, (Shahana & Omman, 2018) considera que todos los usuarios hoy en día cuentan con un mayor número de opciones para dejar sus opiniones, sus dudas, emociones y preocupaciones, lo cual para bien ó para mal, dejan constancia en el ciberespacio. Para las empresas estas acciones son de vital importancia porque así pueden medir su impacto en los medios sociales, pues si se analiza correctamente esta métrica, la información relevante indicará qué le gusta o qué no le gusta al usuario sobre la marca.

Asimismo, (Arango, 2017) en los foros de internet, se aprecia que son grupos de usuarios o redes sociales que vienen a ser como un escenario de comunicación por internet, donde se propicia el debate, la concertación y el consenso de ideas. Es una herramienta que permite a un usuario publicar su mensaje en cualquier momento, quedando visible para que otros usuarios que entren más tarde, puedan leerlo y contestar. Esta forma de comunicación se llama asincrónica dado su simultaneidad en el tiempo. Esto permite mantener comunicación constante con personas que están lejos, sin necesidad de coincidir en los horarios de encuentro en la red, superando así las limitaciones temporales de la comunicación sincrónica y demorando en el tiempo la interacción, lo cual, a su vez, favorece la reflexión y la madurez de los mensajes.

Por tal motivo, en la presente investigación se considera importante analizar los significados de los textos y comentarios de un grupo social, comunidad o foro peruano internet, puesto que su contenido muestra que usuarios son más activos que otros, representan un sentimiento y muestran una opinión que mostrará una preferencia que se puede inferir dando algún significado.

Por lo expuesto, por ejemplo el uso de Foros de debate y su participación en los Grupos de Perú Educa son importantes desde el punto de vista socio-cultural y constructivista del aprendizaje porque estas herramientas nos permiten interactuar, dialogar con todos los usuarios del Perú que tienen las mismas aspiraciones e intereses sobre determinados temas para construir conocimiento en espacios y tiempos diferentes fortaleciendo el trabajo en equipo, la tolerancia y solidaridad entre pares, donde es necesario aplicar estrategias enriquecedoras con intencionalidad y objetivos propuestos. (Witten et al., 2017b)

En los últimos años, Foros Perú es un foro de comentarios. No solo tiene utilidad como foro o comunidad virtual de intercambio de comentarios y opinión, sino también por su manera de compartir información o por las alucinantes historias que encuentras ahí, en diferentes temas. (EDUCA, 2020)

Por lo tanto, las manifestaciones del problema, se representan en los comentarios, mensajes de texto, que se encuentran en internet, como en los foros, tuis, etc. Que coadyuvan a que el mundo sea más rápido, moderno, más conectado y con los usuarios que son cada vez más rigurosos y exigentes, respondiendo a las coyunturas del mercadeo que es cada vez más incluyente, más social y más digital; como se observa en el sitio web Foros Perú.

De acuerdo a lo expuesto se considera que se requiere información sensible, para conocer adecuadamente el significado polaridad de los textos contenidos en los mensajes publicados en internet, una información dispersa y almacenada en diversos foros de debate, argumentación de opiniones, noticias, mensajes y comentarios en internet.

Por otro lado la calidad en los significados y polaridad de los mensajes se ve afectada y varía al interactuar todos los usuarios. En la cual, muchas veces en el sitio Foros Perú existen diversidad de mensajes que refiere los mismos temas y no llegan a conclusiones precisas. Por las diferentes ubicaciones de los usuarios de

internet, no todos poseen facilidades para realizar grupos y especificar sus opiniones, de acuerdo al tema que deseen exponer o compartir en internet. Se manifiesta o se observa que existen ciertos indicios que demandan el conocimiento del verdadero significado y polaridad para lograr un adecuado intercambio de información académica contextualizada en los comentarios del sitio web Foros Perú.

En tal sentido, las **causas** del problema planteado, serían los siguientes: Limitaciones prácticas en el procesamiento de lenguaje natural con técnicas de minería de opinión y preferencias, pocos referentes de técnicas y prácticas en el procesamiento de lenguaje natural con técnicas de minería de opinión y preferencias, pocas orientaciones metodológicas en el procesamiento de lenguaje natural con técnicas de minería de opinión, pocos referentes técnicos y prácticos en el procesamiento de lenguaje natural con técnicas de minería de opinión y preferencias y por último pocos práctica sistemática en el procesamiento de lenguaje natural con técnicas de minería de opinión y preferencias

- 1.2. Antecedentes del Estudio: insuficiencias en el proceso de lenguaje natural, limita la predicción de preferencias de usuarios en el sitio Web Foros Perú

A nivel internacional

Castaño (2018) en el estudio modelo predictivo para inferir en el próximo presidente de estado a través de un vocabulario ontológico en twitter. El objetivo fue determinar un modelo que permita predecir el próximo presidente de estado colombiano, basado en el contenido generado en twitter utilizando un vocabulario ontológico. Conclusión El modelo predictivo dio una favorabilidad al candidato actualmente electo Iván duque, estableciendo que a pesar de que han pasado varios meses una vez terminada la candidatura presidencial aún se conserva una buena imagen del mismo.

COMENTARIO

El autor considera que se debe implementar modelos predictivos para el contenido visual generado en Twitter u otras redes sociales y así obtener más datos y más tiempo en el uso de la misma, en la cual la muestra pueda tener un nivel de confianza como el mínimo, en la cual este pueda mejorar los servicios.

Bermúdez y Bermúdez (2019) en la tesis sistema de predicción de estadísticas de nacimientos y defunciones en Colombia soportadas por TIC, cuyo objetivo fue diseñar un sistema de información predictivo para las estadísticas de nacimientos y defunciones soportado por TIC que permita mejorar el subregistro y la oportunidad de las estadísticas de nacimientos y defunciones tan importante para la toma de decisiones en políticas públicas. La investigación es de tipo Explicativa, en la cual pertenece a la línea de investigación Gestión de Soluciones TI y se fundamenta en el Modelo de Investigación Cuantitativo., con un diseño de estudio experimental. Se concluye que se tiene en cuenta las características particulares de cada territorio el algoritmo SMOreg obtenido permite direccionar de manera oportuna políticas en salud al permitir disponer de estrategias para la recuperación oportuna de los hechos vitales y disponer de datos confiables que permita un mejor aprovechamiento de los recursos

COMENTARIO

Los autores consideran que el software para minería de datos y aprendizaje automático llamado WEKA, se demuestra ser una herramienta de software eficiente para ayudar a comprender el contenido de los datos buscando patrones, comportamientos, secuencias y tendencias que permitan generar un modelo estadístico preciso.

Marín y López (2019) en el estudio análisis de datos para el marketing digital emprendedor: Caso de estudio del Parque de Innovación Empresarial de Manizales. El objetivo fue analizar los datos recopilados y utilizarlos para optimizar procesos y recursos, La investigación fue de tipo exploratorio y descriptivo, con una población de 52 empresas, aplicándoles la técnica de encuestas con un intruemento del cuestionario. Conclusión Las empresas que aportaron al estudio pertenecen a múltiples sectores empresariales, y no sobrepasan los 12 empleados, se debe considerar que el 60% solo lleva entre 1 y 5 años en el mercado. De estas empresas, el 55% vende sus productos tanto online, como offline, y el 35% únicamente por medio físico. Se halló que más de la mitad (55%) de las empresas consideran muy importante, casi vital, el uso del marketing digital y solo 5% la consideran poco importante.

COMENTARIO

Los autores manifiestan que la falta de formación y entrenamiento en análisis de datos (en técnicas para descripción, diagnóstico, predicción y prescripción) se suman los aspectos de seguridad de la información y protección de datos, de manera que la calidad de los datos internos y externos y su uso sean garantizados tanto para las empresas como para los potenciales clientes.

A nivel nacional

Barrueta y Castillo (2017) en la tesis modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos. El objetivo fue implementar un Modelo de Análisis Predictivo que permita determinar clientes con tendencias a la deserción en los bancos en el Perú. La investigación es de tipo Explicativa, con estudio un enfoque cuantitativo. Conclusión el modelo diseñado cumple con el propósito de pronosticar clientes con tendencias a la deserción en Bancos en el Perú a través del análisis del comportamiento, alcanzando la mejor precisión (93.20%) con el algoritmo R-K-Means, teniendo también mejores resultados en los falsos/positivos y falsos/negativos en comparación con otros algoritmos.

COMENTARIO

Los autores consideran recolectar más datos a través del tiempo con el fin de enriquecer y seguir entrenando el modelo; asimismo, teniendo en cuenta los datos históricos de hasta seis meses de anterioridad, donde los futuros análisis se puede identificar las características de los clientes de los clientes desertores a mayor detalle.

Pollette y Hernández (2016) en el estudio implementación de un modelo predictivo basado en data mining y soportado por sap predictive analytics en retails, cuyo objetivo fue implementar un modelo predictivo basado en Data Mining y soportado por SAP Predictive Analytics en una empresa Retails. La investigación es de tipo explicativa, con estudio un enfoque cuantitativo Conclusiones al contar con una enorme cantidad de información se pueden lograr mejores resultados. Gracias a que contamos con información de ventas de los dos últimos años, los algoritmos utilizados arrojaron resultados bastante favorables e incluso mucho más precisos a una estimación matemática.

COMENTARIO

Los autores consideran que es importante contar con una enorme cantidad de información, ya que esto facilita a los algoritmos en encontrar patrones y/o tendencias que permitirán mejores resultados.

Ticona (2019) en la investigación Minería de opiniones basado en aprendizaje supervisado en la evaluación de destinos turísticos de la región de Puno, cuyo objetivo fue utilización de las técnicas y procedimientos necesarios para realizar Minerías de opiniones basado en aprendizaje supervisado con el fin de realizar una evaluación de las opiniones en línea elaboradas por los usuarios de diversos destinos turísticos de la región de Puno. Se utilizó una investigación de tipo aplicada y tecnológica, con una metodología CRISP-DM, de minería de datos integral y un modelo de proceso que proporciona a cualquier persona, Conclusión siguió los procedimientos necesarios en los cuales desarrollamos diferentes procesos de minería de Opiniones, para finalmente obtener cuadros que reflejen la clasificación de opiniones brindadas por los turistas, que en este caso resultó que la mayoría de los comentarios de los destinos turísticos de la región de Puno tienden a ser positivos.

COMENTARIO

El autor permitió analizar la importancia de la minería de opiniones basado en aprendizaje supervisado en la evaluación de destinos turísticos, en la eficiencia del proceso de abastecimiento de una empresa turística ya que las opiniones nos llevarán a identificar el servicio que presta la agencia de turismo evita la falta que la empresa evite tener pérdidas monetarias.

1.3. Teorías relacionadas al tema.

Esta investigación se sustenta en el lenguaje natural y la minería de datos; en la cual sostiene a las variables de la investigación.

Teoría del lenguaje natural

El documento publicado presenta por **objetivo** el modelo implementado de un robot conversacional llamado WIH (Word Intelligent Handler) y varias estadísticas que muestra cómo representa la información extraída de los diálogos en español. Sus **dimensiones y características**, refiere a los diálogos que se procesan con una interfaz por lotes. El prototipo tiene una arquitectura compleja con muchos componentes que permiten la predicción de preferencias de usuario, empleando el sentimental analysis como minería de opinión; ya que modela oraciones en español cuyo **aporte** implementa ondas lingüísticas morfosintácticas (MLW). Usando un mecanismo de representación del conocimiento, presenta el modelo WIH que brinda respuestas automáticamente tomando diálogos previos con o entre humanos; y que puede usarse como modelo para respuestas en futuros diálogos (**Lopèz de Luise, et al., 2017**).

Muestra que la estructura y el proceso de aprendizaje del robot WIH; indica que el enfoque es automático y se basa principalmente en la información extraída de diálogos reales procesados por una arquitectura autoadaptable, sin ningún tipo de predefinido diccionario. Las pruebas confirman hallazgos previos que muestran que ER se comporta de acuerdo con una adaptación general, para realizar la predicción de preferencias de usuarios. (**Lopèz de Luise, et al., 2017**).

Su **objetivo** es la Lingüística Computacional es modelar uno ó varios lenguajes naturales. Aunque es algo difícil de obtener, muchos ingenieros inspirados tomaron este camino, entre ellos, J. Weizenbaum, quien fue uno de los primeros bots de charla llamados ELIZA, otros son el Dr. K. Colby (creador de la charla Parry) (creador de ALICE), T. Winograd (con SHRDLU) y muchos otros más.

Se describe parte de la arquitectura principal de un prototipo llamado Word Intelligent Handler (WIH), su enfoque de modelado y algunas estadísticas de rendimiento y pruebas que prueban su solidez.

ANALISIS CRITICO

Probablemente, una de las tareas más difíciles es determinar el contexto de la oración y manejar con inflexiones variantes de las palabras que derivan en ambigüedad, polisemia, anáfora, etc;

El referido documento presenta un enfoque implementado como parte de un bot conversacional llamado WIH (Word Intelligent Handler). Tiene una arquitectura compleja con varios componentes. Algunos de ellos son la memoria ER, la memoria EP y otros módulos menores que proporcionan el prototipo de buen modelado del español frases. Constituyen el mecanismo de representación del conocimiento que utilizado por WIH para generar respuestas automáticas durante los diálogos con humanos. (Débora Hisgen and Daniela López De Luise-AIGroup, Universidad de Palermo, 2017)

Refiere que los componentes y algunas de sus interacciones, junto con casos de prueba y un análisis estadístico inferencial de los resultados obtenidos muestra que el prototipo WIH puede adaptar su comportamiento y la tasa de aprendizaje de su memoria de trabajo internas según el contenido de los diálogos. (**Bordignon y Panessi, 2016**)

Sus **características o dimensiones** muestra la estructura general y el enfoque de varios módulos de WIH. Los módulos descritos participan en el proceso de aprendizaje del prototipo y se complementan con otros módulos. ER-Memory, EP-Memory y MCT son parte del conocimiento modelado automático. Estos módulos no requieren interacción humana y solo usan información recopilada de diálogos reales procesados por el sistema. La arquitectura proporciona un modelado autoadaptable, ya que no involucran diccionarios predefinidos ni otro tipo de estructuras. Construye todas las estructuras internas utilizando información general morfológica y sintáctica contenida en oraciones procesadas, en el presente trabajo el análisis de las pruebas realizadas se demostró que el tamaño de la memoria indica cierta relación con la complejidad de las oraciones (podría medirse por el número de categorías en cada uno) y el tema (oraciones similares pero diferentes temas hacen cambio similar en el tamaño de la memoria), pero no con el número de oraciones.

Bordignon y Panessi (2017) Su **objetivo** manifiesta que el tratamiento de morfemas dependientes es una de las estrategias utilizadas con la finalidad de incrementar la

proporción de documentos relevantes recuperados en los sistemas de búsqueda de información. (Bordignon & Panessi, 2001)

Su **aporte o resultado** refiere que la técnica de stemming permite extraer sufijos y prefijos comunes, de tal forma que como un sólo término consiste en una adaptación del algoritmo de Porter, para procesar corpus de documentos en castellano. ((Bordignon & Panessi, 2001)

Dicho sistema se **caracteriza** utilizando Python; que muestra un análisis de sentimientos como un área de investigación actual en la minería de textos. Es el origen del procesamiento del lenguaje natural o de los métodos de aprendizaje automático. Son las fuentes importantes para la toma de decisiones y se pueden extraer, identificar y evaluar a partir de las revisiones de sentimientos en línea. El **objetivo principal** es conectarse en Twitter y buscar los tweets que contienen una palabra clave en particular y luego evaluar la polaridad de los tweets como positiva y negativa. En este documento, el **aporte o resultado** considera las palabras clave se recopilan de Twitter utilizando la API de Twitter y los datos sin procesar extraídos se preprocesan utilizando técnicas de Natural Language Toolkit. (Vadivukarassi et al., 2017).

ANALISIS CRITICO

Los sentimientos de los tweets en línea se evalúan en función de la selección de características de las palabras de puntuación. Para seleccionar las mejores características se utiliza la prueba de Chi Cuadrado y el clasificador Naïve Bayes se utiliza para entrenar y probar las características y también para evaluar la polaridad sentimental.

1.3.1 Caracterización del Proceso de lenguaje natural y su dinámica en la predicción de preferencias de Usuario

El presente trabajo de investigación determina las características del proceso del lenguaje natural y su dinámica; y observa que están dados 04 niveles de análisis, denominados wavelets morfosintácticas (mwl) denominado hoy en día minería temporal:

- Análisis morfológico (número ó cantidad de palabras)
- Análisis Sintáctico (número de palabras identificadas para formar la oración)
- Análisis Semántico (procesamiento del significado, si es positivo ó negativo)

- Análisis Pragmático (aportación significativa de los mensajes)

➤ **Determinar las tendencias históricas del proceso de lenguaje natural y su dinámica**

Su **objetivo** analiza el procesamiento del lenguaje natural (PLN) como que no es una ciencia nueva, la tecnología avanza con rapidez gracias a un mayor interés en la comunicación entre humanos y máquinas, además de la disponibilidad del big data, computadoras poderosas, algoritmos mejorados y como humano, usted puede hablar y escribir en inglés, español o chino. Pero el lenguaje nativo de una computadora – que se conoce como código o lenguaje de máquina – es mayormente incomprensible para la mayoría de las personas. En los niveles más bajos de su dispositivo, la comunicación se da no con palabras sino a través de millones de ceros y unos que producen acciones lógicas. Y los programadores utilizaban tarjetas perforadas para comunicarse con las primeras computadoras hace 70 años. Este proceso manual y arduo **el aporte** era entendido por un número relativamente reducido de personas. Ahora puede decir, “Alexa, me gusta esta canción” y un dispositivo que reproduce música en su hogar bajará el volumen y le responderá, “OK. Evaluación guardada”, en una voz similar a la humana. Luego adapta su algoritmo para reproducir esa canción – y otras como ésa – la próxima vez que escuche esa estación musical. **(SAS Software y Soluciones de Analítica, 2019)**

Demos un vistazo más de cerca a esa interacción. Su dispositivo se activó cuando lo oyó hablar, entendió la intención no hablada en el comentario, ejecutó una acción y proporcionó retroalimentación en un enunciado bien estructurado en inglés (o español), todo en un espacio de unos cinco segundos. La interacción completa la hizo posible el procesamiento del lenguaje natural, junto con otros elementos de inteligencia artificial como el aprendizaje basado en máquina y el aprendizaje a fondo. **(SAS Software y Soluciones de Analítica, 2019)**

(SAS Software y Soluciones de Analítica, 2019) El procesamiento del lenguaje natural permite que las computadoras a comunicarse con humanos en su lenguaje y escala otras tareas relacionadas con el lenguaje. Por ejemplo, NLP hace posible que las computadoras lean texto, escuchen la voz hablada, la interpreten, midan el sentimiento y determinen qué partes son importantes. **(SAS Software y Soluciones de Analítica, 2019)**

Las máquinas de hoy en día, pueden revisar exhaustivamente más datos basados en lenguajes que los humanos, sin fatigarse y de una forma consistente e imparcial. Considerando la asombrosa cantidad de datos no estructurados que se generan todos los días, desde registros médicos hasta medios sociales, la automatización será decisiva para analizar por completo datos de texto y habla con eficiencia. **(SAS Software y Soluciones de Analítica, 2019)**

El lenguaje humano es increíblemente complejo y diverso. expresa de maneras infinitas, verbalmente y por escrito. No sólo existen cientos de lenguajes y dialectos, pero en cada lenguaje existe un conjunto único de reglas gramáticas y de sintaxis, términos y palabras coloquiales. Cuando escribimos, a menudo se comete errores ortográficos o abreviamos palabras, o bien omitimos signos de puntuación. Cuando se habla, tenemos acentos regionales, y masculla, tartamudea o toma palabras prestadas de otros idiomas. **(SAS Software y Soluciones de Analítica, 2019)**

Aunque el aprendizaje supervisado y no supervisado, y específicamente el aprendizaje a fondo, se utilizan ahora ampliamente para modelar el lenguaje humano, se necesitan también entendimiento sintáctico y semántico, y conocimientos de dominio que no están necesariamente presentes en estos métodos de machine learning. NLP es importante porque ayuda a resolver la ambigüedad del lenguaje y agrega estructura numérica útil a los datos para muchas aplicaciones industriales, como el reconocimiento del habla o la analítica de texto. **(SAS Software y Soluciones de Analítica, 2019)**

El procesamiento del lenguaje natural tiene diferentes técnicas para interpretar el lenguaje humano, que van desde los métodos estadísticos y del aprendizaje basado en máquina hasta los enfoques basados en reglas y algorítmicos. Necesitamos una amplia variedad de métodos porque los datos basados en texto y en voz varían ampliamente, al igual que las aplicaciones prácticas.

Las tareas básicas de NLP incluyen la simbolización y el análisis sintáctico, lematización/derivación, etiquetado de la parte del habla, detección del lenguaje e identificación de relaciones semánticas. Si alguna vez creó diagramas de enunciados en la primaria, ya ha realizado estas tareas de forma manual antes.

En términos generales, las tareas NLP dividen el lenguaje en piezas elementales más cortas, intentan entender las relaciones entre las piezas y exploran cómo funcionan las piezas juntas para crear significado y poder hacer una adecuada predicción de preferencias de usuario.

Estas tareas supuestas utilizan a menudo en recursos NLP de más alto nivel, como:

- **Categorización de contenido.** Un resumen del documento basado en la lingüística, incluyendo búsqueda e indexación, alertas de contenido y detección de duplicación.
- **Descubrimiento y modelado de temas.** Capture con precisión el significado y temas en colecciones de texto, y aplique analítico avanzada a texto, como optimización y pronósticos.
- **Extracción contextual.** Extraiga automáticamente información estructurada de fuentes basadas en texto.
- **Análisis de sentimiento.** Identificación del estado de ánimo u opiniones subjetivas en grandes cantidades de texto, incluyendo minería de sentimiento y opiniones promedio.
- **Conversión de habla a texto y de texto a habla.** Transformación de comandos de voz en texto escrito y viceversa.
- **Sumarización de documentos.** Generación automática de sinopsis de grandes cuerpos de texto.
- **Traducción basada en máquina.** Traducción automática de texto o habla de un idioma a otro.

Por lo tanto, el fin de la presente investigación es emplear texto crudo del lenguaje y aplicar la lingüística y algoritmos para transformar o enriquecer el texto de tal forma que provea un mayor valor para conocer cual comentarios o grupos de mensajes tiene o predomina mas en la conversación realizada, y de esta forma se aplica el sentimental analysis y/o minería de opinión, para realizar una buena predicción de preferencias de usuario.

- Etapa 2.- Evolución del Proceso de Lenguaje Natural (PLN), con la minería de opinión y preferencias, para lograr la predicción de preferencias de usuarios (aproximadamente 1980 a la actualidad).

El algoritmo de Maximización-Esperanza (EM) es un algoritmo heurístico porque usa una secuencia de pasos que llegan a mostrar un valor aceptable, es decir no brinda una

convergencia absoluta, puesto que ese valor probabilístico no absoluto, por lo que el algoritmo no es determinista, sin embargo su valor siempre converge, porque emplea variables latentes, y utiliza frecuentemente para algoritmos de agrupamiento en aprendizaje automático y visión artificial, para aprender modelos ocultos de Márkov y Mixturas de Gaussianas, utilizadas en procesos de clasificación o reconocimiento.(Witten et al., 2011). El algoritmo EM, dada una distribución conjunta o sea de varias variables observables X, y se genera la variable Z, dado un parametro Q, el objetivo es maximizar la función de verosimilitud (loglikelihood), $p(E/Q)$ con respecto a Q.

Pasos que realiza el algoritmo EM (Witten et al., 2017a)

- 0. Se eligen valores iniciales para θ , denominado θ^{old} .
- 1. **Paso E** Evaluar $p(Z|X, \theta^{old})$.
- 2. **Paso M** Evaluar θ^{new} dado que

$$\theta^{new} = \arg_{\theta} \max Q(\theta, \theta^{old})$$

donde

$$Q(\theta, \theta^{old}) = \sum_z p(Z|X, \theta^{old}) \ln p(X, Z|\theta).$$

- ♣ Revisar la convergencia a través de la log-verosímilitud o la discrepancia entre los valores de los parámetros en cada iteración.
- ♣ Regrese al paso 1.

La idea de *Minería de Datos* no es nueva. Ya desde los años sesenta los estadísticos manejaban términos como *Data Fishing*, *Data Mining (DM)* o *Data Archaeology* con la idea de encontrar correlaciones sin una hipótesis previa en *bases de datos* con ruido. (Alfonso Cutro DATA PRIX, 2019)

A principios de los años ochenta, *Rakesh Agrawal*, *Gio Wiederhold*, *Robert Blum* y *Gregory Piatetsky-Shapiro* entre otros, empezaron a consolidar los términos de *Minería de Datos* y *KDD*. (Alfonso Cutro DATA PRIX, 2019)

Esta tecnología ha sido un buen punto de encuentro entre personas pertenecientes al ámbito académico y al de los negocios.

La evolución de sus herramientas en el transcurso del tiempo puede dividirse en cuatro etapas principales:

- *Colección de Datos (1960).*
- *Acceso de Datos (1980).*
- *Almacén de Datos y Apoyo a las Decisiones (principios de la década de 1990).*
- *Minería de Datos Inteligente. (finales de la década de 1990).*

Ahora, se pondrá un ejemplo básico cuando persona está en una tienda probándose un pantalón y de repente llega un cliente y tras mostrar la prenda rota afirma: “El tejido que ustedes venden aquí es de mala calidad” ¿Qué es lo que inmediatamente se le viene a la cabeza? Probablemente, casi con toda seguridad, esta situación le haga pensar o al menos reflexionar sobre la idoneidad de seguir o no comprando. La razón es que ha detectado una opinión claramente negativa hacia el producto.

En tal sentido, la minería de datos, en su evolución y análisis que realiza de los datos, emplea hoy en día las Redes Sociales, y van a considerarse importantes los comentarios sean positivos o negativos, y se magnifica alcanzando a millones de personas, pudiendo acabar con la reputación de cualquier persona o empresa.

Precisamente si hay algo que han conseguido los Medios Sociales es dar poder al ciudadano, al servir de plataforma donde poder mostrar sin tapujos su opinión hacia un producto, servicio, personaje o empresa en cuestión. ¿Cuántas veces un pequeño tweet ha generado un claro rechazo en la red traspasando el mundo online? Piense un segundo...

Pero hay más, ya que el uso creciente de estos medios ha traído a primer plano de una disciplina muy necesaria en la analítica de datos. Se trata del análisis de sentimientos o también conocido como la minería ó análisis de opinión que sirve para identificar, a través del procesamiento del lenguaje natural, análisis de texto y lingüística computacional, información que nos permita entender cuál es la intención exacta de un mensaje.

En resumen, se puede decir que la minería de opinión ó análisis de sentimiento. permite determinar la actitud respecto a un tema general o específico, frente a un producto o situación determinada.

Entonces el análisis de sentimiento o minería de opinión como se le conoce hoy en día, sirve como herramienta muy valiosa para las organizaciones porque permite obtener datos de calidad para mejorar las estrategias empresariales, facilitando la gestión de la reputación online y pudiendo tomar decisiones en tiempo real, por ejemplo: ante las temidas crisis de reputación online. **(PROMETEUS , 2019).**

Precisamente aquellas compañías que cuentan con el análisis de sentimiento dentro de su estrategia extraen conclusiones y toman decisiones concretas en base a la información que logran, pudiendo lanzar a posteriori acciones ante un problema en campos tan importantes como la experiencia de cliente o construir mensajes más acertados desde el área o departamento de marketing. **(PROMETEUS , 2019)**

También llamado Opinion Mining, el Análisis de Sentimientos tiene hoy una aplicación muy importante en el campo de las redes sociales y es en la actualidad uno de los temas más candentes dentro del campo de la información, por las aplicaciones que está generando. **(PROMETEUS , 2019)**

Las técnicas se centran principalmente en el procesamiento, búsqueda o extracción de información subjetiva, tratando de clasificar los textos de manera automática y catalogando éstos en función de la connotación positiva o negativa del lenguaje.

El Análisis de Sentimientos trata de traducir a indicadores más o menos medibles las emociones humanas inmersas en datos sociales, pero también se centra en los datos internos o propios de la empresa, permitiendo conocer de forma eficiente qué se dice sobre una marca o producto y pudiendo, seguir las opiniones de diferentes usuarios influyentes, o detectar, simplemente, tendencias en la red. **(PROMETEUS , 2019)**

Cuatro son las vías o enfoques en el análisis de sentimiento: localización de palabras clave, afinidad léxica, métodos estadísticos y técnicas a nivel de concepto, teniendo diferentes indicadores: Positivo, Negativo, Neutro.

En la actualidad existen numerosas herramientas capaces de detectar tendencias en Redes Sociales a través de datos de distintos sitios en la red que permita un mayor entendimiento de lo que se está comentando e identificar oportunidades. **(PROMETEUS , 2019)**

(Witten et al., 2017a) La minería textual, minería de opinión, lingüística computacional y recuperación de información, desde el punto de vista técnico, se refiere a la minería de datos que recoge técnicas usadas tradicionalmente en la recuperación textual, y en la lingüística computacional. Esta influencia llega a tal punto que resulta difícil poder afirmar que la minería textual haya incorporado técnicas propias.

Las organizaciones tienen una gran cantidad de información y datos. Solo por el hecho de producir productos y servicios y atender a usuarios, van acumulando registros de todo tipo. (Minería de Datos y su aplicación en el marketing , 2017).

Los datos de gran utilidad para los negocios, pero pocos tienen el conocimiento y las herramientas necesarias para transformarlos en conclusiones relevantes. Al no ser procesados, estos se pierden y las empresas siguen trabajando en base a estándares generales. (Minería de Datos y su aplicación en el marketing , 2017)

La minería de datos emplea información para personalizar y optimizar los productos y servicios. Se trata de un conjunto de tecnologías y técnicas computacionales para explorar las bases de datos de forma automática o semi automática.

Su objetivo es extraer conclusiones de un conjunto de datos, aplicando estructuras de análisis reutilizables. Para lograr esto, usa métodos de la inteligencia artificial, el machine learning, las estadísticas y los sistemas de bases de datos. (Minería de Datos y su aplicación en el marketing , 2017)

Las técnicas que se aplican pueden tener cualquiera de estos enfoques. Algunas de las más usadas son:

- **Clasificación:** Asigna una categoría a cada caso para organizar la información y entender las variables existentes.
- **Regresión lineal:** Determina una relación entre una o más variables para predecir el valor de un tercer dato. Se usa mucho para establecer la relación que existe entre grupos de datos.
- **Agrupamiento o clustering:** Agrupa los datos según criterios definidos. Para esto, intenta crear grupos naturales, buscando características comunes.

- **Resumen:** Reúne y compacta los datos para crear subconjuntos que resuman la muestra.
- **Dependencias:** Analiza las relaciones para establecer dependencias entre las variables.
- **Análisis de secuencias:** Estudia la evolución temporal de las variables para encontrar secuencias recurrentes.
- **Redes neuronales:** Imita la forma en que funciona el sistema nervioso para crear un sistema de interconexión de datos donde hay estímulos y salidas.
- **Árboles de decisión:** Crea construcciones lógicas para organizar en categorías las condiciones que ocurren de forma sucesiva y repetitiva.
- **Reglas de asociación:** Se utilizan para descubrir hechos que ocurren en común dentro de un conjunto de datos.

Normalmente, una empresa que podía perder cien clientes en una semana, conseguiría cien nuevos clientes en el mismo tiempo, por lo que consideraban que sus ventas eran satisfactorias. Estas empresas tenían la certeza de que existían bastantes clientes para reemplazar a los desertores. No obstante, la pérdida de clientes puede acarrear un costo muy alto, pues este es un indicador de que algo no está funcionando como debe ser.

1. **Cuál es el costo de los clientes que se pierden**

Todas las empresas deben vigilar de cerca cuál es el porcentaje de pérdidas de clientes y tomar las medidas necesarias para acortar estos números. Si la empresa tiene muchos desertores frecuentemente, esta empresa debe investigar el por qué, y además, determinar cuáles son las causas que llevan a los clientes a tomar esta decisión.

Cabe destacar que no se puede hacer mucho con los clientes que se pierden por causas de cambio de localidad o país en donde los productos de la empresa no están disponibles. Sin embargo, se puede hacer mucho cuando los clientes que desertan son por causas de poca calidad en los productos y servicios que la empresa ofrece, por lo tanto, la empresa debe tener la información de los porcentajes de deserción y motivos por los cuáles estos clientes se van para poder aplicar las estrategias necesarias para lograr la fidelización.

Supongamos que una empresa pierde la venta de 100 productos en una semana, porque 100 clientes decidieron no comprar su producto ya que están desmotivados por fallas en atención

al cliente o dichos productos tienen desperfectos. Calculando que cada producto tiene un costo de 1\$, serían 100\$ de pérdidas en ventas. Entonces, la empresa debería estimar cual es el costo de reducir el porcentaje de deserciones de estos clientes invirtiendo en nuevas estrategias de atención al cliente o control de calidad para mejorar sus productos y así evitar la pérdida de ventas. Aquí es dónde viene la interrogante ¿el costo de reducir este porcentaje de deserciones es inferior al costo que se tiene por perder las ventas sin necesidad? Probablemente es más bajo y no se correrá el riesgo de tener una mala reputación. ¿Crees que vale la pena aplicar nuevas estrategias e invertir en el mejoramiento de los productos para fidelizar a los clientes? Sin duda alguna es lo más aconsejable.

2. La necesidad de conservar a los clientes

Hoy en día, las empresas se diferencian del resto hacen un gran esfuerzo por conservar a sus clientes ya que muchos mercados que están en su etapa de madurez no tienen nuevos ingresos de clientes en ninguno de sus nichos. (Agencia de Marketing Digital EXPERTO SEO, 2017)

Un factor preocupante para muchas empresas es que la competencia sigue en aumento y los costos por atraer nuevos clientes son mucho más altos. En el mercado actual, tal vez sea cinco veces más difícil atraer a un cliente nuevo que mantener contento a uno existente.

Lamentablemente, antes era más importante atraer nuevos clientes, en lugar de conservar a los existentes. Eran mucho más importantes las transacciones que las relaciones y los análisis fueron dirigidos a las actividades antes de vender y durante la venta, pero nunca tomaron en cuenta las actividades después de la venta. (Agencia de Marketing Digital EXPERTO SEO, 2017)

Afortunadamente, hoy en día son más las empresas que reconocen la importancia de los clientes existentes, aunque la mayor parte de estas empresas no cuente con un sistema contable que muestren el valor de los clientes fieles.

A pesar de que gran parte de las organizaciones se enfoque en idear mezclas de mercadotecnia que produzcan ventas y clientes nuevos, la primera línea defensiva de las empresas será siempre conservar a sus clientes y la mejor forma de conservarlos es

proporcionándole total satisfacción post venta, de este modo conseguirán una sólida fidelidad.

3. **Cuál es la clave para fidelizar clientes**

Una de las estrategias más efectivas para la fidelización se trata de darle prioridad a las relaciones con los clientes y generar valor y compromiso. A esto lo llamaremos invertir en relaciones. En la actualidad, el punto focal de la mercadotecnia son las relaciones plenas de valor y redes de comercialización y no las transacciones individuales. Esta estrategia se enfoca en relaciones a largo plazo, la meta es proporcionar valor a los clientes y el éxito es su satisfacción total. (Agencia de Marketing Digital EXPERTO SEO, 2017)

Para enfocarse en las relaciones de valor con los clientes es necesario que todos los departamentos de la empresa trabajen en equipo con el de mercadotecnia para el servicio de los consumidores, lo que implica establecer relaciones en varios niveles (económico, social, técnico, jurídico) que redundarán en la fidelidad del cliente.

Existen cinco niveles o tipos de relaciones que se pueden establecer con los clientes que han comprado un producto en alguna empresa.

El básico: El vendedor de la empresa vende el producto al cliente, pero no hace un seguimiento de ningún tipo.

El interesado: Este es el vendedor que pide al cliente que lo llame si tiene algún problema con el producto o si tiene dudas con respecto a su funcionamiento.

El responsable: El vendedor se comunica con el comprador poco después de la venta para saber si el producto está cumpliendo con sus expectativas. Además, el vendedor le pide al cliente sugerencias para la mejora de los productos o servicios que ofrezca. Esta información le servirá a la empresa para mejorar su oferta de manera constante.

El proactivo: El vendedor o alguna otra persona del departamento de ventas llama al cliente de la empresa para ofrecerle nuevos productos que pueden ser de su utilidad.

El de la sociedad: La empresa trabaja en forma sostenida con todos sus clientes para encontrar la forma de proporcionarles más valor antes, durante y después de la venta.

Teniendo en cuenta esta información, vale la pena considerar las opciones y emplear el tipo de relaciones que más beneficios les traerá a las empresas. (Agencia de Marketing Digital EXPERTO SEO, 2017)

4. **Fidelización de clientes**

El Internet ha enseñado que nunca se agotarán las posibilidades creativas que de este pueden desprenderse, para ayudar a mejorar nuestra vida, o simplemente disfrutarla más. Aunque hoy es indispensable, en sus inicios tuvo momentos difíciles, la compañía que, mayoritariamente se encargó de lanzarlo al mercado, IBM, actualmente con una presencia limitada en este. Pero los esfuerzos de esta compañía no fueron en vano, hoy el Internet está incluido en la vida de todos y cada uno, lo requerimos para relajarse, para desestresarse, ver películas y series de televisión, escuchar música y conocer gente. También lo requerimos para trabajar evidentemente, para publicitarnos y dar a conocer nuestras marcas: hacer comercio, a lo largo de todo el mundo. (Uranga, 2019)

Hoy en día para la buena aplicación de comercio electrónico y marketing digital se tiene en cuenta el empleo de tecnología adecuada que brinde estadísticas no solo descriptivas sino también en el ámbito inferencial, por tal motivo en la presente investigación se considera el empleo del siguiente método y herramienta:

➤ **Descripción del algoritmo Expectation Maximization (EM) (algoritmo de maximización de expectativas ó algoritmo esperanza)**

EM asigna una distribución de probabilidad a cada instancia que indica la probabilidad de que pertenezca a cada uno de los conglomerados. EM puede decidir cuántos clústeres crear mediante validación cruzada, o puede especificar a priori cuántos clústeres generar.

La validación cruzada realizada para determinar la cantidad de clústeres se realiza en los siguientes pasos:

1. el número de clústeres se establece en 1
2. el conjunto de entrenamiento se divide al azar en 10 pliegues.
3. La EM se realiza 10 veces utilizando los 10 pliegues de la forma habitual de CV.
4. La loglikelihood se promedia sobre los 10 resultados.
5. si la loglikelihood ha aumentado, el número de conglomerados se incrementa en 1 y el programa continúa en el paso 2.

El número de pliegues se fija en 10, siempre que el número de instancias en el conjunto de entrenamiento no sea menor a 10. Si este es el caso, el número de pliegues se establece igual al número de instancias.

Los valores faltantes se reemplazan globalmente con `ReplaceMissingValues`.

OPCIONES: semilla: la semilla de número aleatorio que se utilizará.

- `numFolds`: el número de pliegues que se utilizarán al realizar la validación cruzada para encontrar el mejor número de clústeres (predeterminado = 10)
- `numExecutionSlots`: el número de ranuras de ejecución (subprocesos) que se utilizarán. Establecer igual al número de cpu / núcleos disponibles
- `debug`: si se establece en `true`, `clusterer` puede generar información adicional en la consola.
- `numKMeansRuns`: el número de ejecuciones de k-means para realizar.
- `displayModelInOldFormat`: usa el formato antiguo para la salida del modelo. El formato antiguo es mejor cuando hay muchos clústeres. El nuevo formato es mejor cuando hay menos grupos y muchos atributos.
- `minLogLikelihoodImprovementIterating`: la mejora mínima en la probabilidad de registro necesaria para realizar otra iteración de los pasos E y M
- `maximumNumberOfClusters`: el número máximo de clústeres a considerar durante la validación cruzada para seleccionar el mejor número de clústeres
- `minLogLikelihoodImprovementCV`: la mejora mínima en la probabilidad de registro con validación cruzada requerida para considerar el aumento del número de clústeres cuando se realiza la validación cruzada para encontrar el mejor número de clústeres
- `numClusters`: establece el número de clústeres. -1 para seleccionar el número de clústeres automáticamente mediante validación cruzada.

- doNotCheckCapabilities: si se establece, las capacidades del clusterer no se verifican antes de que se construya el clusterer (use con precaución para reducir el tiempo de ejecución).
- maxIterations - número máximo de iteraciones
- minStdDev: establece la desviación estándar mínima permitida

Para validar los resultados obtenidos se emplea la medida de log likelihood. Se entiende por log likelihood a la Probabilidad Logarítmica, es una función de verosimilitud mide la bondad de ajuste de un modelo estadístico a una muestra de datos para valores dados de los parámetros desconocidos.

Se forma a partir de la distribución de probabilidad conjunta de la muestra, pero se ve y se utiliza como una función de los parámetros únicamente, por lo que se tratan las variables aleatorias como fijas en los valores observados.

ANÁLISIS CRÍTICO

Según (Witten et al., 2017a) en el presente trabajo de investigación se considera el empleo de dos técnicas al que se le denomina heurístico puesto que utiliza el procesamiento de lenguaje natural en su análisis morfológico y sintáctico (PLN) con el empleo del algoritmo de maximización esperanza (EM), permiten conocer el significado de las palabras analizadas de los comentarios del sitio web Foros Perú, y en esa forma se logra conocer la polaridad que corresponde a determinar que autor de palabras escritas en los comentarios analizados se le puede considerar como seguidor activo y seguidor menos activo y también se permite conocer de los comentarios analizados, cuyo resultado los indica en dos grupos o cluster 0 el significado **“lo que se desea hacer y sus características”** y en el cluster 1 el significado **“estados de ánimo y apreciaciones, y sus características”**.

1.3.2 Marco Conceptual.

Se describen términos empleados en el desarrollo del trabajo de investigación, como son los siguientes:

1. **ANÁLISIS DE SENTIMIENTOS:** permite identificar a las personas más influyentes dentro de una red, es decir, aquellas cuyas opiniones tienen un impacto significativo sobre las opiniones de los demás.

2. **ORACIONES DE OPINIÓN** deben contener palabras clave que connoten una opinión.
3. **ADVERTORIAL**. Publicidad que se introduce en el encabezamiento del correo electrónico con un enlace o un hipervínculo a la página *web* del anunciante *e-mail*, tiene apariencia de documento informativo.
4. **AFFILIATE PROGRAM (Programas de afiliados)**. Software que permite utilizar el *website* de una empresa como plataforma de venta de otra compañía (normalmente a través de banners o botones). Entre las empresas que utilizan este tipo de venta destacan Amazon y Bol.
5. **ALIAS**. Nombre corto y fácil de recordar que se utiliza en sustitución de otro nombre largo y difícil de recordar.
6. **ASP**. Acrónimo de *Application Service Provider* (Proveedor de Servicios de Aplicaciones). Las ASP, también denominadas *netsourcing*, son empresas que alquilan aplicaciones informáticas a sus clientes a través de Internet.
7. **AUTO-RESPONDER**. *Software* utilizado para generar un mensaje previamente preparado y enviarlo como un *e-mail* en cuanto recibes un pedido o información. Estos *e-mails*, evidentemente, son todos iguales y, normalmente, no van personalizados principalmente avisan de que su mensaje ha sido recibido.
8. **BANNER (cartel, rótulo)**. Gráfico publicitario rectangular que puede ser fijo o animado, e incluso con sonido, que se incluye en las páginas *web* a modo de anuncio. Haciendo *click* sobre él, normalmente envía hacia el sitio *web* del anunciante.
9. **BANNER-EXTENSIBLE**. A diferencia del *banner* tradicional, el extensible puede hacerse más grande y ofrecer mayor espacio para dar información o plantear una oferta.
10. **BANNER-MICROSITE**. En apariencia, es un *banner* normal, pero con la diferencia de que, una vez que se pulsa, aparece en una ventana nueva un *microsite*.
11. **BARTERING**. Intercambio de servicios, productos o publicidad.
12. **BOTONES**. Son también *banners*, pero de tamaño más reducido. Pueden aparecer estáticos o dinámicos; fijos en secciones o en páginas de inicio. En muchos casos, se pueden seleccionar con el ratón y suelen colocarse en un lateral de la página.
13. **BROWSER (navegador)**. Programa utilizado para visualizar las páginas *web*. Los más utilizados son el Internet Explorer y Netscape.

14. **BRICKS.** Expresión que define a las empresas físicas cuyo objeto de negocio se da *off line*. Se usa para compararlas con las que trabajan en la red (**CLICKS**).
15. **BULK MAIL.** Envío masivo de correo electrónico. Es lo que realizan los *spammers*.
16. **BUSCADORES.** Herramientas de búsqueda de la red que permita, mediante palabras o combinaciones de palabras, encontrar documentos alojados en páginas *web*.
17. **BUSINESS PLAN (plan de negocio).** Este documento es el carné de identidad del emprendedor, el que debe presentar ante cualquier inversor. Un buen plan de negocio debe recoger la idea básica de la empresa, las previsiones de ingresos y rentabilidad futura y el plan de acción por el que se espera lograr ambos.
18. **BUY WIZARD.** Asistente de ayuda para la compra en Internet.
19. **CÁMARA WEB.** Cámara de vídeo que se conecta al ordenador y se utiliza para mantener videoconferencias.
20. **CHAT (tertulia, conversación, charla).** Comunicación simultánea entre dos o más personas a través de Internet.
21. **CIBERMALL.** Un gran almacén virtual.
22. **CLICK & BRICK (también conocidas como *brick & mortar*).** Empresas tradicionales que amplían su estrategia a realizar negocios también en la red.
23. **CLICK RATE (ratio de click).** La frecuencia con la que los visitantes de un sitio *web* pulsán sobre un anuncio, normalmente un *banner*, mostrado en una página.
24. **CLICK THROUGH.** Acción de pulsar sobre un *banner* publicitario. Ratios de *click through*, es la relación entre el número de visitantes de una página y los que pulsán sobre el *banner*.
25. **COOKIES.** Archivo de texto que se graba en el ordenador del visitante del cual se sirven los servidores *web* para guardar información acerca del cliente de un sitio. Sirve para identificar a visitantes recurrentes. Es un archivo de texto que se introduce en el disco duro al visitar un sitio *web*. La próxima vez que se vuelva a visitar ese mismo sitio, el *web* busca esta información que le ayudará a recordar quién eres tú, cuáles son tus preferencias, que has hecho otras veces que has visitado la *web*, que habías comprado... Se utiliza para personalizar la venta ya que saben lo que te gusta y te lo ofrecen.
26. **CRACKER (intruso).** Persona que intenta acceder a un sistema informático sin autorización.

27. **CRM o *Customer Relationship Management***. Básicamente consiste en centrar tu modelo de negocio en el cliente y dotar a tu empresa de las herramientas técnicas que permita prestar un servicio y comunicación a tus usuarios. Es una estrategia de negocio orientada a la fidelización de clientes. Permite a todos los empleados de una empresa disponer de información actualizada sobre los mismos, con el objetivo de optimizar la relación entre empresa/cliente. Los factores más valorados por los clientes son: rapidez de respuesta, comprensión, responsabilidad y accesibilidad. CRM facilita la gestión de todos estos factores englobando los procesos de márketing, ventas y atención al cliente en uno.
28. **CURSORES ANIMADOS**. Ideados por la compañía Comet Systems, permite introducir mensajes o animaciones del anunciante en el cursor del ordenador del usuario. El usuario debe instalar una pequeña aplicación de Comet Systems que hace que el cursor del internauta cambie su forma de entrar en las páginas que utilicen el *Comet Cursor*. El resultado es un cursor con el logotipo de la empresa anunciante.
29. **CYBERCASH**. Dinero digital en efectivo. Dinero electrónico. Este mecanismo de pago se creó con la intención de resolver los problemas de seguridad relacionados con el uso del número de tarjeta de crédito por Internet.
30. **CYBERSPACE (ciberespacio)**. Término creado por William GIBSON en su novela fantástica *Neuromancer* para describir el «mundo» de los ordenadores y la sociedad creada en torno a ellos.
31. **DNS**. Acrónimo de *Domain Name System* (Sistema de Nombres de Dominio). Sistema para traducir los nombres de los ordenadores en direcciones IP numéricas. Los dominios más importantes son: .com (comercial y empresas), .edu (educación y centros docentes), .org (organizaciones sin ánimo de lucro), .net (operación de la red), .gov (gobierno de Estados Unidos) y .mil (ejército de Estados Unidos). La mayoría de los países tienen un dominio propio. Por ejemplo, .us (Estados Unidos), .au (Australia), .uk (Reino Unido).
32. **DOMAIN (dominio)**. Dominio o nombre en propiedad que tiene una empresa para operar en la red. Hay elaborado un sistema internacional de dominios o DNS que los agrupa por jerarquías, simplificando su identificación. En España es la terminación: es.

33. **DOT.COM (punto com).** Forma de nombrar a aquellas empresas de Internet que se basan en la plataforma *web* como principal medio de negocio, comunicación e interacción con el público objetivo.
34. **DOWNLOAD (bajar, transferir).** Proceso de transferir información desde un servidor de información al propio ordenador personal.
35. **E.BIZ.** Expresión inglesa para definir los negocios en la red.
36. **e-BUSINESS.** Conjunto de líneas de negocio canalizadas a través de Internet.
37. **e-COMMERCE (comercio electrónico).** Comercio entre empresas y clientes que se realiza a través de Internet.
38. **e-CONOMY.** Economía electrónica. Parte de la economía relacionada con Internet.
- e-COUPONS.**— Cupones promocionales y de descuento a modo de publicidad.
39. **e-FULFILLMENT.** Término de comercio electrónico equivalente al *fulfillment* del marketing tradicional, es decir, los servicios integrados de logística (gestión de pedidos, tramitación de entregas, seguimiento y control...). El *e-fulfillment* es cada vez más un servicio a prestar por los grandes operadores de logística en su actividad en Internet -como *outsourcing* estratégico-, y exige una perfecta coordinación empresa-operador. Se pretende de esta manera que esta parte importante de la gestión comercial la lleve a cabo una empresa altamente especializada.
40. **e-MAIL (correo electrónico).** Sistema para enviar mensajes en Internet. El emisor de un correo electrónico manda los mensajes a un servidor y éste, a su vez, se encarga de enviárselos al servidor del receptor. Para poder ver el correo electrónico es necesario que el receptor se conecte con su servidor.
41. **e-MAIL MÁRKETING.** Las acciones de marketing que se realizan utilizando como soporte el correo electrónico.
42. **e-MAILING.** Uso del correo electrónico para el envío de mensajes publicitarios.
43. **e-MARKETER.** Persona encargada del marketing *on line*.
44. **ENGAGE.** Método por el cual dos empresas intercambian entre sí el contenido de sus bases de datos con los perfiles de los visitantes.
45. **ENLACES DE TEXTO.** Son líneas de caracteres alfanuméricos que pueden ser seleccionadas con el ratón y sirven para informar sobre eventos y noticias o para comunicar ofertas de ventas.
46. **e-SALES.** Ventas de bienes y servicios realizadas a través de Internet.

47. **ESCAPARATES.** También conocidos como *banner* escaparate, están formados por un conjunto de imágenes y texto sobre los que se puede pulsar. Los escaparates se suelen utilizar en las secciones de compras de los sitios y son una buena fuente de información sobre las ofertas de ventas.
48. **e-SHOPS .** Tiendas virtuales.
49. **e-TAILER.** Comerciante detallista que centra su actividad en Internet. Su medio de relacionarse con el consumidor es a través de e-mail.
50. **e-VALUE (valor electrónico).** Valor añadido que genera una compañía a los clientes de sus clientes.
51. **EZINES (*Electronic magazines*).** Revistas virtuales. Se trata de una revista que se distribuye a través del correo electrónico en formato texto o HTML. Los *e-zines* se reciben por suscripción. Dicha suscripción puede ser gratuita o previo pago de una cuota.
52. **FAQ (*Frequently asked questions*).** Lista de preguntas más frecuentes. Es una relación con las preguntas más comunes que alguien se puede hacer acerca de un bien o servicio y las respuestas a las mismas.
53. **FILE TRANSFER (transferencia de ficheros).** Copia de un fichero desde un ordenador a otro a través de una red de ordenadores.
54. **FIREWALL (cortafuegos).** Programa que sirve para filtrar lo que entra y sale de un sistema conectado a una red. Suele utilizarse en las grandes empresas para limitar el acceso de Internet a sus empleados, así como para impedir el acceso de archivos con virus.
55. **FREE BIES.** Cosas gratuitas que se ofrecen en un sitio *web*.
56. **FREE NETTING.** Define el hecho, realizado por empresas tradicionales, de entrar en Internet sin haber realizado previamente una auditoría de márketing que les pueda indicar dónde y cómo se encuentran y así valorar cuándo y cómo deben entrar con éxito en la red.
57. **FREWARE.** Aplicaciones de uso gratuito que pueden encontrarse en Internet.
58. **FTP.** Acrónimo de *File Transfer Protocol* (Protocolo de Transferencia de Ficheros). Protocolo que permite al usuario de un sistema acceder y transferir ficheros de un ordenador a otro a través de Internet.
59. **GPR (*Gross Rating Point*).** Unidad de medida del efecto de la publicidad tradicional.

60. **USERS** Persona que, con grandes conocimientos informáticos o ninguno, puede introducir información mediante el uso de ordenadores o redes informáticas de particulares, organizaciones o instituciones si están conectados a Internet.
61. **HIT.** Es cada uno de los archivos (gráficos, sonidos, texto...) que suministra el servidor donde está hospedada la sede *web* visitada. Una página con muchos gráficos puede generar muchos *hits* pero pocas visitas reales. La publicidad se paga por *impressions* por unidad de tiempo. Esto es una medida más cercana al número de visitas a un sitio *web*.
62. **HOME PAGE (página de inicio o principal).** Página inicial que aparece cuando se accede a un servidor de páginas *web*.
63. **HOST (sistema central).** Ordenador que permite a los usuarios comunicarse con otros sistemas centrales de una red.
64. **HOSTING.** Servicio de alojamiento de las páginas *web* que gestionan empresas especializadas. Las empresas que se dedican a este servicio son como los hoteleros de la red: ofrecen espacio para que otras compañías almacenen cualquier información que quieran que sea accesible por una red, desde sus páginas *web* hasta la información de su red interna o Intranet.
65. **HOTLINK.** *Link* caliente. Se denomina de esta forma al enlace muy visitado.
66. **HOT LIST.** Lista caliente. Listado de personas que han respondido favorablemente a una acción de *marketing*.
67. **HOUSE ADVERTISING.** Publicidad propia. Una empresa aprovecha su propio espacio *web* para anunciar productos o servicios propios.
68. **HOUSING.** Servicio de alojamiento, conexión, gestión y administración de equipos informáticos.
69. **HTTP.** Acrónimo de *HyperText Transport Protocol* (Protocolo de Transporte de Hipertexto). Protocolo para mover archivos de hipertexto a través de Internet.
70. **IMPRESSION.** Impresión. Cada impresión corresponde a un usuario que ve una página *web* con un anuncio o *banner* publicitario. Para que se contabilice una impresión es necesario que se muestre o descargue la página de manera completa.
71. **INDIVIDUALS EXPOSED.** Número de individuos distintos que han visto una vez la publicidad.
72. **INFOBANNER.** Es un *banner* que ofrece información en tiempo real al usuario.

73. **INFOMEDIARIO.** Lugar en la red que facilita la entrada a agentes o empresas que, anteriormente, no estaban posicionados en ese sector de actividad; por ejemplo, *franchipolis.com* dentro del sector franquicias. Los infomediarios, de esta manera, están en condiciones de establecer las bases para la creación de nuevos mercados en torno a los cuales compradores y vendedores, en muchos casos desconocidos entre ellos inicialmente, establecen relaciones comerciales de interés para ambos. El infomediario presenta como valor añadido, su capacidad y habilidad para estructurar la oferta que presenta la red.
74. **INTERSTITIAL.** Anuncio publicitario a modo de página *web* que aparece de forma repentina entre las páginas que esta consultando. Es una publicidad a pantalla completa. Intenta recrear el *spot* televisivo en Internet. El problema principal es la lentitud con la que se descarga. Puede ser considerado por el usuario como intrusivo.
75. **IP.** Acrónimo de Internet Protocol (Protocolo de Internet).
76. **JOINT-VENTURE.** Acuerdo de colaboración entre dos empresas para la creación de una tercera.
77. **KEYWORD.** Palabra clave para tener acceso a una parte restringida de una *web*.
78. **LAYER.** Botón a modo de *pop-up* que puede moverse por la pantalla para llamar la atención.
79. **LINK.** Enlace, hipervínculo. Conexión con otro documento *web* por medio de la dirección URL. Los enlaces aparecen en el texto de un documento *web* en forma de texto subrayado y de distinto color.
80. **LINK-EXCHANGES.** Intercambio de *links*. Fórmula mediante la cual se introduce un hipervínculo en tu *web* que remite a otra y, a cambio, esa otra *web* realiza la misma acción.
81. **LIST SERVER (servidor de listas).** Sistema automatizado de listas de distribución mediante correo electrónico.
82. **LOG.** Son archivos en los que se recogen las visitas que tienen las páginas de un sitio *web*.
83. **LOGIN.** Nombre que se usa para acceder a un sistema de ordenadores.
84. **LURKER.** Internauta que practica el *lurking*.
85. **LURKING.** Actividad consistente en leer mensajes y artículos de foros, grupos de *news* y listas de distribución sin jamás añadir comentario alguno. Es una forma de investigar el mercado utilizando la red.

86. **M-COMMERCE.** (*Mobile Commerce*) Comercio electrónico desde el teléfono móvil.
87. **MAILING LIST (lista de correo, lista de distribución).**— Lista de direcciones electrónicas utilizada para distribuir mensajes a un grupo de personas. Se usa habitualmente para discutir sobre un determinado tema. Puede ser abierta o cerrada y tener o no un moderador. Si es abierta, cualquier persona puede suscribirse a ella.
88. **MARKETPLACE.** — Término creado por John SVIOKLA, define el lugar en la red donde se producen intercambios comerciales entre diferentes empresas. Nacido por la asociación *on line* de varias grandes empresas, que pueden llegar a competir *off line*, por eso es tan importante que desde los inicios del proyecto, todos los socios estén de acuerdo sobre sus objetivos. Los *marketplaces* son más importantes en diversos sectores que en otros.
89. **MARKETING ONE-TO-ONE.**— Una de las enormes posibilidades que proporciona Internet es la interactividad con el usuario, lo que permite conocer de primera mano sus impresiones y adaptar la oferta de sus productos y servicios a su demanda precisa. Acciones de *márketing* personalizadas.
90. **MÁRKETING VIRAL.**— Una estrategia de *márketing* viral consiste básicamente en conseguir que los usuarios se transmitan a otros un determinado mensaje, noticia, promoción, evento..., obteniendo de este modo un crecimiento exponencial.
91. **MERCHANDISING DIGITAL.** — Material promocional sobre el que se puede aplicar la imagen corporativa. En Internet, bajo este término, quedarían englobados los salvapantallas, fondos de escritorio, postales...
92. **METAMERCADO.**— Son todas las aportaciones y soluciones complementarias que rodean al producto, mercado o servicio principal. Un ejemplo en el sector del automóvil, una *web* de una determinada manera puede complementar su oferta con servicios financieros, seguros, talleres...
93. **META-TAGS.**— Conjunto de códigos del lenguaje *html* que proporcionan una información esencial acerca de una página *web*. La mayoría de los motores de búsqueda leen las *meta-tags* para catalogar las páginas *web*.
94. **MICROSITES.**— Son pequeñas *webs* independientes, dentro de un sitio, dedicadas a un producto o tema concreto.

95. **MINERÍA DE OPINIÓN**, se denomina también análisis de sentimiento que viene a ser la aplicación del proceso de lenguaje natural, lingüística computacional y minería de textos; y que permite extraer información subjetiva a partir de contenidos creados por los usuarios,
96. **MÓDEM**.— Término que proviene de las palabras Modulador-Demodulador. Consiste en un dispositivo que se conecta al ordenador y a una línea telefónica y que permite poner en contacto dos ordenadores.
97. **NAVEGADOR**.— Programa utilizado para visualizar las páginas *web*. Los más utilizados son el Internet Explorer y Netscape.
98. **NETIQUETTE**.— Conjunto de normas de comportamiento que rigen una conducta adecuada en Internet.
99. **NETMARKETERS**.— Profesionales de márketing exclusivos de la red.
100. **NETWORK (red)**.— Grupo de ordenadores y otros dispositivos periféricos conectados unos a otros para comunicarse y transmitir datos.
101. **NEWBIE**.— Principiante o novato en Internet.
102. **NEWS (noticias)**.— Tablón electrónico de anuncios. Permite al usuario participar en grupos de discusión mediante el envío de mensajes, o bien sólo acceder a estos grupos para obtener la información.
103. **NEWSGROUPS (grupos de noticias o news)**.— Anuncios clasificados por grupos y que son accesibles mediante correo electrónico.
104. **NODO**.— Ordenador o conjunto de ordenadores que reciben la llamada del usuario y la dirigen hacia el servicio solicitado allá donde se encuentre.
105. **OFF LINE**.— Cuando no se está conectado a Internet. Todo aquello que tiene que ver con la vida real fuera de la red.
106. **ON LINE**.— Ordenador conectado a la red.
107. **OPT-IN**.— Alta voluntaria en una lista de distribución. Es además la acción de notificarle a una empresa u organización que se desea recibir vía *e-mail* la información que vayan generando.
108. **OPT-OUT**.— Baja voluntaria en una lista de distribución.
109. **PALABRA CLAVE**.— Palabra contratada por los anunciantes en los buscadores para que cuando el usuario introduzca este término el buscador muestre una referencia correspondiente al sitio *web* del anunciante.
110. **PASSWORD**.— Contraseña alfanumérica necesaria para acceder a un espacio o información restringida.

111. **PATROCINIO.**— En el soporte financiero de una página *web* a cambio de la mención del nombre o producto de patrocinador o la inserción del logotipo. El patrocinio mejora la imagen de marca del patrocinador.
112. **PERMISSION MARKETING.**— Es un tipo de márketing basado en conseguir el permiso de los consumidores para enviarles información sobre nuestra empresa, productos y servicios.
113. **POP-UP.**— Ventana que aparece sobre la que se está visualizando.
114. **PRECISION MARKETING.**— Técnica de márketing que consiste en ampliar el conocimiento sobre los clientes y, en función de dicho conocimiento, personaliza la oferta el máximo posible.
115. **PROTOCOLO.** Lista de comandos estandarizada a la que responde un servidor.
PUERTO.— Lugar donde la información entra o sale de un ordenador, o ambas cosas.
116. **QUICK LINK.**— Enlace que permite trasladarse de forma muy rápida. Suele ser una palabra que figura resaltada en una página *web* y que tras «clickear» sobre ella lleva a otro sitio *web* o a un micrositio relacionado.
117. **RDSI.** Acrónimo de red Digital de Servicios Integrados.
118. **REACH.** Porcentaje de personas que han visto el *banner*.
119. **ROAD SHOW.**— Cuando una empresa quiere salir a Bolsa o vender un proyecto; es el hecho de realizar una ronda de visitas a compañías de inversión y analistas por toda la nación, o incluso por varios países.
120. **SEARCH ENGINES.**— Motores de búsqueda o buscadores.
121. **SERVER (servidor).** Sistema que proporciona recursos (por ejemplo, servidores de ficheros, servidores de nombres). En Internet, este término se utiliza a menudo para designar los sistemas que proporcionan información a los usuarios de la red.
122. **SET (Secure Electronic Transaction).** Sistema que garantiza la seguridad en las transacciones electrónicas.
123. **SITE (sitio).** Lugar en la red. Sitio o página *web*.
124. **SPAM o correo basura.** Cualquier tipo de *e-mail* no solicitado.
125. **SPAMMER.** Persona que roba o compra direcciones de correo electrónico sustraídas y remite *e-mails* no solicitados.
126. **SPAMMING.** Consiste en el envío indiscriminado de mensajes de correo no solicitados, normalmente con mensajes publicitarios.
127. **SPONSORSHIP.** Patrocinio, auspiciador.

- 128.**SSL** (*Secure Socket Layer*). Sistema que permite que la información (normalmente datos económicos) viaje encriptada evitándose que puede ser leída. Es el método que permite garantizar una alta seguridad en el comercio electrónico.
- 129.**START-UP**. Empresa de nueva creación nacida para hacer negocios sólo a través de Internet.
- 130.**TRACKING**. Seguimiento de la navegación de un usuario al que se le ha marcado con una *cookie*.
- 131.**TRAFFIC**. Tráfico. Cantidad y tipo de personas que acuden a un sitio *web*.
- 132.**TENDENCIAS ACTUALES DEL E-COMMERCE**, viene a ser la cantidad de datos de los usuarios disponibles en la red, con el fin de entender más fácil a los clientes y poder así satisfacer sus necesidades.
- 133.**USABILITY**. Todos los aspectos relacionados con la navegabilidad, diseño de páginas, presentación de los contenidos... de una *web*.
- 134.**UMTS** (*Universal Mobile Communication Systems*). Es la tercera generación de teléfonos móviles basados en multimedia e Internet. Permitirán la navegación por Internet y la transferencia de vídeo, imágenes, sonido y texto.
- 135.**URL**. Acrónimo de *Universal Resource Locator* (Localizador Universal de Recursos /Identificador Universal de Recursos). Sistema unificado de identificación de recursos en la red. Es el modo estándar de proporcionar la dirección de cualquier recurso en Internet.
- 136.**VENTANAS FLOTANTES**. También conocidas como *up windows*, son ventanas que se presentan en un navegador nuevo de manera inesperada mientras se está cargando una página solicitada. Hay *por ups* interactivos y otros que simplemente ofrecen un hipervínculo.
- 137.**VENTURE CAPITALIST**. Inversor de capital riesgo.
- 138.**VIRUS**. Programa que se duplica a sí mismo en un sistema informático incorporándose a otros programas que son utilizados por varios sistemas. Estos programas pueden causar problemas de diversa gravedad en los sistemas que almacenan.
- 139.**VORTAL**. Un portal vertical. Contracción de las palabras inglesas: *vertical* y *portal*.
- 140.**WAP** (*Wireless Application Protocol*).— Protocolo de Aplicaciones Inalámbricas. Permite recibir en los teléfonos móviles la información exclusivamente de texto de Internet. Sistema de navegación por Internet desde el móvil GSM.

141. **WEB.** Servidor de información WWW. Se utiliza también para definir el universo WWW en su conjunto.
142. **WEBMASTER (administrador de web).** Persona responsable del mantenimiento de un sitio *web*.
143. **WEBSITE.** Conjunto de páginas *web* que dependen del mismo dominio.
144. **WIRE (Web Internet Resource Executive).** Responsable del comercio electrónico.
145. **WYSIWYG.**— Acrónimo de *What You See is What You Get* (lo que ves es lo que tienes) . Técnica que ofrece la reproducción exacta en la pantalla de un texto tal como aparecería después en formato impreso.
146. **WWW.**— Acrónimo de *World Wide Web* (telaraña o malla mundial). Sistema de información distribuido con mecanismos de hipertexto. Es el universo de servidores *http*, que permiten mezclar texto, gráficos y archivos de sonido juntos.

1.4. Formulación del Problema.

1.4.1. Problema General

- ¿Cómo mejorar el procesamiento de lenguaje natural que limita la predicción de preferencias de usuarios en el sitio Web Foros Perú?

1.4.2. Problemas Especificos

- ¿Como se caracteriza epistemológicamente el procesamiento de lenguaje natural y minería de opinión mediante el sentimental analysis?
- ¿De que manera se diagnostica el estado actual de la dinámica del procesamiento de lenguaje natural NLP con la minería de opinión y preferencias en el sitio web Foros Perú?
- ¿Cómo elaborar un modelo de predicción de preferencias de usuario, para mejorar el intercambio de información en el sitio web Foros Perú?
- ¿De que manera se propone estructurar un sistema predictivo basado en el modelo de predicción de preferencias de usuario, que verifique que usuarios son más activos en el sitio web Foros Perú?
- ¿Cómo se validará y corroborará los resultados alcanzados?

1.5. Justificación e importancia del estudio.

Justificación Científica:

La presente investigación se justifica por cuanto pretende demostrar que el proceso de lenguaje natural, limita la predicción de preferencias de usuarios en el sitio Web Foros Perú. Teniendo en cuenta que esto permitirá tener un adecuado proceso para entender los significados y polaridad de los mensajes contenidos en la página de internet FOROS PERÚ, permitiendo conocer que usuarios es mas activo que otro.

Justificación Teórica:

Busca obtener información de argumentos sólidos de conocimientos científicos comprobados para la sustentación del estudio; por otro lado, en la metodología, se tiene en cuenta la estadística inductiva en la cual se encarga de extraer conclusiones a partir de una muestra al total de la población con un pequeño margen de error; en la práctica, si, se demuestra en cuanto al sistema predictivo basado en el modelo de preferencias para la predicción de preferencias con prototipo de usuarios en el sitio web Foros Perú de estudio se logrará sentar las bases para una próxima investigación propositiva que en la práctica conllevará a plantear lineamientos estratégicos en el modelo de preferencias para la predicción. Esto permitirá poder aplicarse como medio de inteligencia de negocios, prevenir el abandono de clientes, usarse como componente en sistemas de recomendación, compararse con la competencia, analizar la opinión del público, predecir el futuro, usarse como medio para analizar las citas y descubrir líderes de opinión para mercado. Por otro lado en la justificación social permitirá mostrar estadísticas como:

Justificación Económica:

El 71% de las personas en Internet (65% de estadounidenses) han realizado investigaciones en línea sobre un producto al menos una vez, entre los lectores de comentarios en línea, entre el 63% y el 77% informa que los comentarios tuvieron una influencia significativa en su compra (usuarios más activos en las conversaciones revisadas); los consumidores dicen estar dispuestos a pagar entre 25% y 79% más por un ítem que tenga 5 estrellas de calificación a uno con 4 estrellas; el 38% de los

usuarios ha aportado una clasificación sobre un producto, servicio o persona a través de un sistema de clasificación en línea.

1.6. Hipótesis General

H1: Un sistema predictivo basado en el modelo web mejora la predicción de preferencias de usuarios.

1.6.1. Hipótesis Específicas

H2: El sentimental de analysis resalta, caracteriza epistemológicamente en el proceso de lenguaje natural y minería de opinión

H3: Si se diagnostica el estado actual de la dinámica del procesamiento de lenguaje natural NLP con la minería de opinión y preferencias en el sitio web Foros Perú, mediante etapas entonces la comprensión y precisión será adecuada para verificar el perfil de los usuarios que interactúan en foros Perú.

H4: Si se elabora un modelo de predicción de preferencias de usuario, se mejora significativamente el intercambio de información en el sitio web Foros Perú.

H5: Si se estructura un sistema basado en el modelo de predicción de preferencias de usuario, entonces se verificará que los usuarios son más activos en el sitio web Foros Perú.

H6: Si se valida y corrobora los resultados alcanzados, empleando herramienta estadística inferencial entonces se demostrará que los usuarios son mas activos y/o pueden tener mas seguidores en los comentarios que escriben.

Contrastación de Hipotesis:

HO: El procesamiento de lenguaje natural NO limita significativamente la predicción de preferencias de usuarios en el sitio Web Foros Perú

1.7. Objetivos

1.7.1. Objetivos General

- Evaluar el sistema predictivo basado en el modelo de preferencias para la predicción de preferencias con prototipo de usuarios en el sitio web Foros Perú.

1.7.2. Objetivos Específicos

1. Comprobar las características epistemológicas del procesamiento de lenguaje natural y minería de opinión mediante clasificación masiva de datos de manera automática.
2. Determinar el diagnostico del estado actual de la dinámica del procesamiento de lenguaje natural NLP con la minería de opinión y predicción de preferencias en el sitio web Foros Perú.
3. Verificar el modelo de preferencias de usuario, para mejorar el intercambio de información en el sitio web Foros Perú
4. Verificar la estructura de un sistema web basado en el modelo de predicción de preferencias de usuario en el sitio web Foros Perú.
5. Demostrar la validación y corroboración de los resultados alcanzados en el estudio.

II. MATERIAL Y MÉTODO

2.1. Tipo y Diseño de Investigación.

La presente investigación es de tipo descriptivo proyectivas, así lo menciona Jacqueline Hurtado (2014), manifestó en su libro “Metodología de la investigación” que el tipo de investigación proyectiva tiene como objetivo diseñar o crear propuestas enfocadas a resolver ciertas situaciones siempre que estén sustentados en un proceso de investigación (p.135)

Además, es de enfoque cuantitativo, ya que se realizará un procesamiento estadístico de los datos registrados mediante el instrumento elaborado, cuyos resultados permitan elaborar la propuesta de acuerdo a los objetivos trazados. Así lo mencionan Hernández y Mendoza (2018), dice que un estudio cuantitativo se usa para la consolidación de las suposiciones de manera exacta Cordova Baldeón I. (2015)

El diseño del estudio es no experimental porque no se alteró, manipuló o intervino de alguna forma en el comportamiento de las variables al ser analizadas, ya que estas se observaron tal y como se presentan dentro del contexto abordado. Así como lo mencionan (Hernández y Mendoza, 2018)

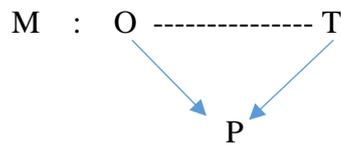


Figura 1: Diseño de estudio

Donde:

M: Muestra

O: Observación

T: Teoría

P: Propuesta

2.2. Población y muestra.

Población:

Según, Arias, F. (2015) considera que la población que es el conjunto de personas, objetos que constituyen el universo total del estudio de investigación; por lo cual la población estuvo conformada por todos los comentarios dispuestos en el sitio web www.forosperu.net en internet. Siendo un total de 2500 comentarios.

Muestra:

El muestreo es un estudio no probabilístico, porque la elección de los elementos no depende de la probabilidad, si no de causas atañidas a los tipos de la investigación o de quien hace la muestra. En este estudio la manera de realizarla no es mecánico, ni con base en fórmulas de probabilidad, si no que depende de la causa en la forma de operación y toma de decisiones del investigador o investigadores, por lo tanto, las muestras elegidas acatan a otros juicios de elección (Hernandez Sampieri, 2014)

Por lo tanto, la muestra se tomó por muestreo probabilístico aleatorio, debido a que se realiza al azar, ya que se ha logrado cumplir con el objetivo. En la cual queda

establecida por 1000 comentarios de la sección universidades del sitio web Foros Perú.
Cordova Baldeón I. (2015)

2.3. Variables.

VARIABLE INDEPENDIENTE:

“Sistema Predictivo basado en el modelo web que mejora la predicción de preferencias de usuarios”

VARIABLE DEPENDIENTE:

Predicción de preferencias de usuario del sitio web foros Perú.

2.4. Técnicas e instrumentos de recolección de datos, validez y confiabilidad.

En el actual estudio se aplicaron las técnicas de gabinete que valió para procesar la información teórica tanto de la variable Sistema basado en el modelo de predicción, como de la variable Preferencias de usuario del sitio web foros Perú. También, se trabajó la técnica de campo, para evaluar sobre el conocimiento d en el modelo de predicción asimismo para la recolección de la información se utilizó el instrumento de análisis documental.

2.5. Procedimientos de análisis de datos.

Para el análisis y recolección se optó por los siguientes estudios estadísticos:

Se utilizó La estadística descriptiva, en la cual, permitió analizar y describir e interpretar en forma objetiva los valores de los datos para así dar paso a un nuevo descubrimiento de nuevos hechos, basándose en los principios del carácter científico, asimismo, estos datos permitieron realizar un análisis técnico de datos para obtener una información confiable y oportuna. Por otro lado, la estadística a partir de los datos se efectuó estimaciones y predicciones de datos de confianza.

Empleando para realizar el análisis respectivo el procesos de wavelets morfosintácticos y el algoritmo de maximización-esperanza, asi como el grado de verosimilitud (likelookhood) respectivamente.

2.6. Criterios éticos

En este proyecto de investigación, se tuvo como sujeto de estudio a usuarios del sitio web foros Perú., en la cual se garantizó la protección de sus derechos. Por ello, en este trabajo se tuvo cuidado de asegurar el respeto a los usuarios del sitio web que participaron en la investigación,

Asimismo, toda investigación en que participen seres humanos debe realizarse de acuerdo con cuatro **principios** éticos básicos, como son: el respeto por las personas, la beneficencia, la no maleficencia y la justicia. Se respetó al elegir la participación voluntaria, guardando confidencialidad.

2.7. Criterios de Rigor científico.

En el presente trabajo de investigación científica se ha considerado los siguientes criterios de rigor científico, según se indica a continuación:

CREDIBILIDAD O VALOR DE VERDAD,

El rigor científico en torno a la credibilidad implica la valoración de las situaciones en las cuales una investigación pueda ser reconocida como creíble, para ello, es esencial la pesquisa de argumentos fiables que pueden ser demostrado en los resultados del estudio realizado.

TRANSFERIBILIDAD O APLICABILIDAD

Los resultados de este estudio, no son transferibles ni aplicables a otros contextos y/o ámbitos de acción, criterio del cual se tiene razón plena, en tanto la naturaleza social y compleja del fenómeno estudiado.

DEPENDENCIA

Este criterio implica el nivel de consistencia o estabilidad de los resultados y hallazgos del estudio.

CONFIRMABILIDAD

El grado de implicación del investigador en el estudio, no se ha eludido, en todo caso se extiende la garantía suficiente sobre el proceso de la investigación, producto de la información arrojada por los instrumentos aplicados, donde los datos no están sesgados, ni responden a ningún tipo de manipulación de naturaleza personal

III. RESULTADOS

3.1. Resultados en Tablas y Figuras

El resultado en el presente trabajo de investigación es que se llega a conocer el significado y polaridad (refiere a que usuarios o seguidores son más activos y menos activos) de los 1000 comentarios que se han analizado del sitio web Foros Perú; mediante el empleo de un prototipo para analizar las palabras de todos los comentarios que corresponden a un tema específico. con la técnica algoritmo maximización-esperanza de WEKA que permite hacer la clasificación de aprendizaje automático, minería de datos, con la técnica de procesamiento de lenguaje natural (wavelets morfosintácticas), con la que se ha realizado el procesamiento de palabras de los comentarios (instancias).

Por lo tanto, se logra conocer cuál es el significado de las palabras de los comentarios analizados, es decir que usuarios han escrito mensajes que predominan unos más que otros en sus opiniones, y determinando que usuarios son más activos en los comentarios escritos en el sitio web indicado, según (Brun & Senso, 2004) a eso se le conoce como una forma de inferencia a partir de las (variables latentes) ó metadatos creados y procesados con el prototipo, generando archivos ó tablas en formato texto, que en el presente trabajo se le ha denominado: usuario y blog.

Es preciso indicar que en la tabla 6 de la sección valoración y corroboración de resultados, describe el resultado obtenido indicando en las filas los usuarios que escriben comentarios, y en las columnas su media aritmética de palabras analizadas y luego en la tabla 5 muestra los clusters (grupos), es decir en la fila los usuarios que escriben comentarios y en las columnas, weka en forma automática hace la clasificación y asigna 10 grupos que los muestra con los números del 0 al 9, indicando la cantidad de elementos o instancias analizadas.

Por lo expuesto, se muestra un diagrama de bloques que muestra el procesamiento de los comentarios, con el algoritmo EM y las wavelets morfosintácticas (análisis

lexico, morfologico y sintactico); en cuanto al significado y polaridad se verifica con el analisis de sentimiento el significado de los mensajes contenidos en los comentarios que corresponden a un tema especifico procesado, indicando que significado expresan todas las palabras y que usuarios tiene polaridad positiva y por ende cuales son los mensajes predominan e indican que usuario es mas activo que otro,

Diagrama de Bloques del Sistema Predictivo basado en la predicción de preferencias de usuarios en el sitio web Foros Perú.



Figura 1.- Diagrama de Bloques del Modelo de predicción de preferencias de usuarios del sitio web Foros Perú

Mapeo de datos e información: (ver detalle en la Tabla 1 y Tabla 4)

IdComentario	Tema	Autor	Palabra	Stem	Frecuencia	TamPalabra	TermPalabra	Mayuscula	EsTitulo	TiPalAnt	Resaltada	Tipopagina (URL)	Tip.Palabra
2	Los thread de los estudiantes de economia	jldark89	debatir	debat	1	Media	ir	NO	NO	verbo	NO	contenido	verbo

Tabla 1.- Mapeo de Información y datos

Características de las variables o metadatos determinados para realizar el procesamiento de palabras de los comentarios.

- **Palabra:** palabra seleccionada del comentario que quedo del proceso realizado.
- **Raíz o STEM de la palabra:** es el stem o la raíz de la palabra seleccionada y es extraída con el algoritmo de porter este es la raíz cierto?
- **tema:** tema del que trata el comentario y opinión
- **palAntTipo:** tipo de palabra que tiene la palabra anterior a ésta (sustantivo, verbo, otra), Si no hay anterior se coloca null y con esto además se determina si es la primera palabra procesada en la sentencia.
- **tipoPagina:** índice o contenido
- **terminacion:** si es ar, er, ir, or, ur, ra, re, ri, ro, ru, s, m, sa, se, si, so, su, an, en, in, on, un,cion, ciones, null.
- **empiezaMay:** es SI cuando la palabra empieza con mayúscula y NO cuando empieza con minúscula.
- **resaltada:** es SI cuando la palabra entera está entre comillas, apóstrofes o toda en mayúsculas
- **esTitulo:** es SI cuando la palabra está contenida en el título (en nuestro caso sería en el tema)
- **catLongPal:** se cuenta la cantidad de caracteres en la palabra original y luego se determinan tres categorías (corta, media, larga) según tenga menos de 5 caracteres, menos de 10 o más.

Dimensión: Comprensión.-

Con respecto al objetivo general: Evaluar el sistema predictivo basado en el modelo de preferencias en los usuarios en el sitio web Foros Perú.

Para evaluar o analizar el modelo de predicción de preferencias en los usuarios en el sitio web Foros Peru, en la presente investigación se ha empleado en primer lugar el programa de computadora para hacer el analisis lexico, morfologico y sintactico es decir todo el procesamiento de lenguaje natural (PLN) a los 1000 comentarios de la pagina web Foros Peru.

Interpretación:

PROTOTIPO: se le denomina a la unión de dos técnicas que permite realizar estadística, tecnica de algoritmo de maximización – esperanza EM de WEKA, y técnica PLN que permite evaluar o analizar las palabras de los comentarios procesados teniendo en cuenta las variables o metadatos que forman las tablas ó grupos de datos usuario y blog.



Figura 2.-Aplicación creada para analizar palabras.

Interpretación: La figura muestra la captura de pantalla de inicio del programa elaborado para realizar el proceso de lenguaje natural basado en las waveletes morfosintácticas (es decir el PLN, su analisis lexico, morfologico, sintactico):



Figura 3.- Interfaz de registro de temas de los comentarios analizado

Interpretación: La figura muestra la captura de pantalla donde se registran los temas de los 1000 comentarios analizados.

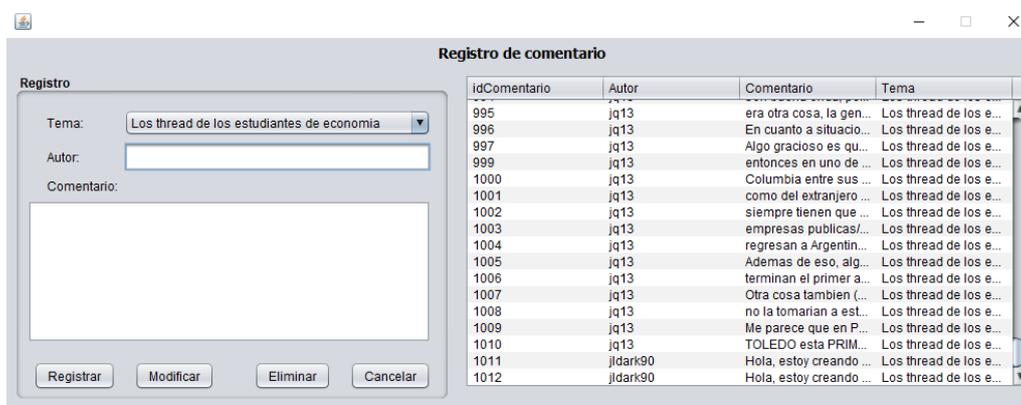


Figura 4.-Interfaz de registro de comentarios analizados

Interpretación: La figura muestra la captura de pantalla que visualiza los 1000 comentarios analizados en el presente trabajo de investigación.

Dimensión: Comprensión

Con respecto al objetivo específico 1: Comprobar las características epistemológica del procesamiento de lenguaje natural y minería de opinión mediante el sentimental analysis.



Figura 5.-Registro y Control de Preposiciones de la aplicación

Interpretación: La captura de pantalla del prototipo muestra la característica del proceso de lenguaje natural que es la realización del análisis léxico, morfológico y sintáctico que en este caso verifica la sintaxis de cada comentario y en la figura muestra el registro de preposiciones extraídas de los comentarios analizados, fin de poder viabilizar el análisis.



Figura 6.-Registro y Control de Artículos de la aplicación

Interpretación: La captura de pantalla del programa elaborado muestra la característica del proceso de lenguaje natural que es la realización del análisis morfológico, sintáctico, que en la figura muestra el registro de artículos extraídas de los comentarios analizados. (waveletes morfosintactios).

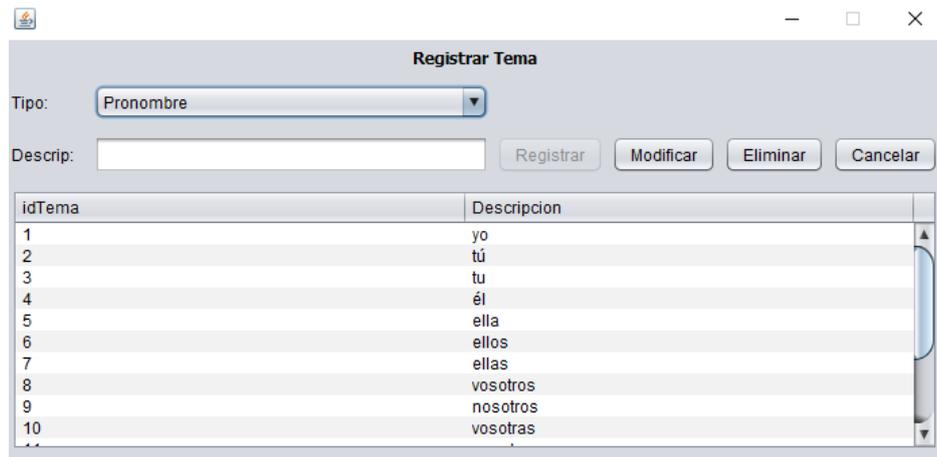


Figura 7.-Registro y Control de pronombres de la aplicación

Interpretación: La captura de pantalla del programa elaborado muestra la característica del proceso de lenguaje natural que es la realización del análisis sintáctico y morfológico que en la figura muestra el registro de pronombres extraídos de los comentarios analizados. (waveletes morfosintactios).

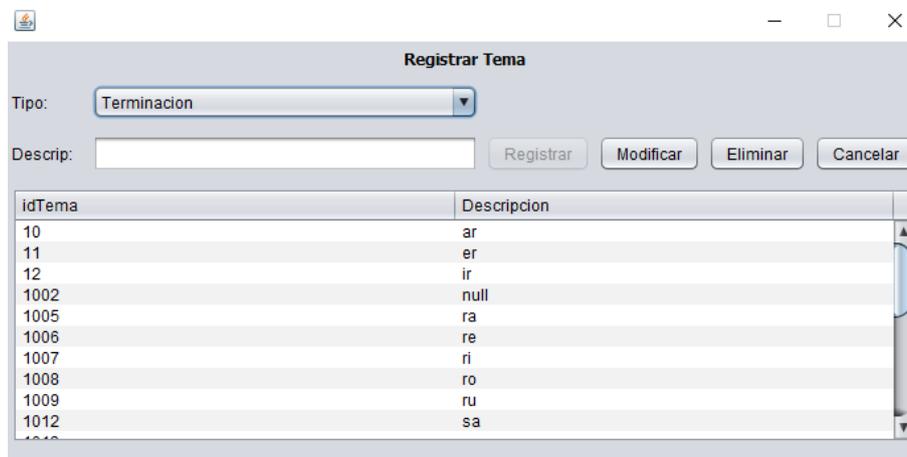


Figura 8.-Registro y Control de Terminaciones de parser (palabras analizadas)

Dimensión: Comprensión:

Con respecto al objetivo específico 2: Determinar el diagnostico del estado actual de la dinámica del proceso de lenguaje natural PLN con la minería de opinión y preferencias en el sitio web Foros Perú.

El diagnostico del estado actual se refiere al grupo de comentarios de la pagina foros Peru, que corresponden a un tema especifico universitarios que se analiza con las waveletes morfosintácticas (PLN) y el algoritmo esperanza-maximización (EM).

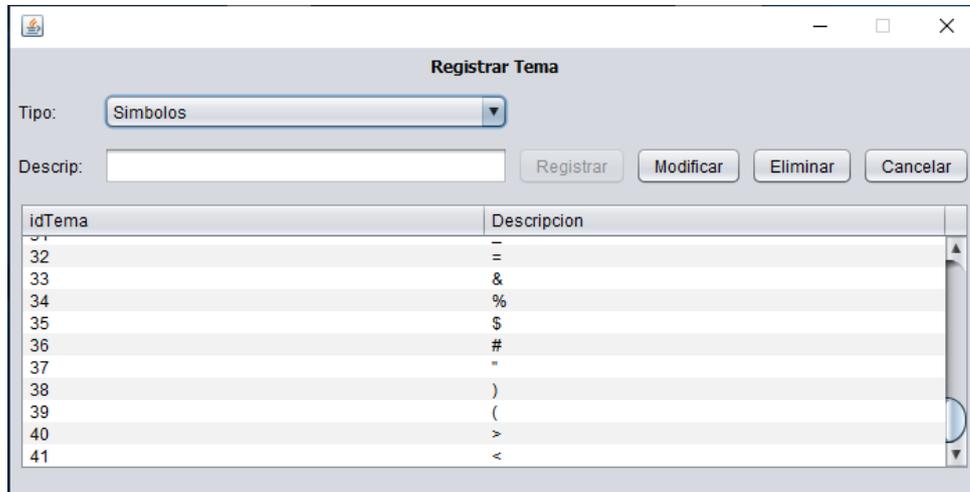


Figura 9.-Registro y Control de Símbolos y Caracteres Especiales detectados en los comentarios analizados

Interpretación: La captura de pantalla del programa elaborado muestra el estado actual de la dinámica del procesos de lenguaje natural PLN con la minería de opinión y preferencias en el sitio web Foros Peru, determinando que los comentarios examinados no se conoce explícitamente lo que significan, para lo cual se inicia el registro y control de símbolos, caracteres especiales detectados en los comentarios analizados por el programa informatico elaborado.

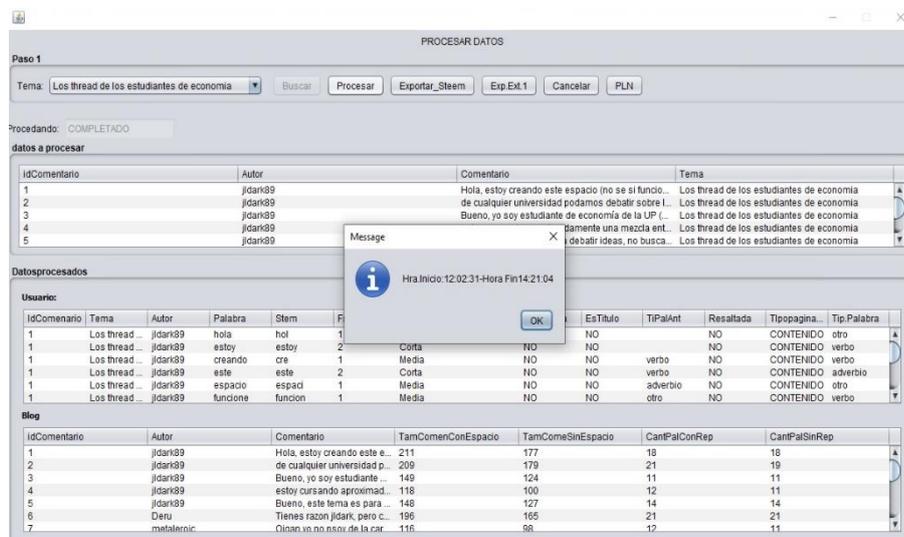


Figura 10.-Interfaz procesar datos, muestra los resultados procesados y analizados de las palabras que componen los comentarios registrados en el prototipo desarrollado

Interpretación: La captura de pantalla del programa elaborado muestra el estado actual de la dinámica del procesos de lenguaje natural PLN con la minería de opinión y preferencias en el sitio web Foros Peru, mostrando los comentarios analizados desagregados en dos tablas denominado usuario y blog sobre la dinámica del proceso de lenguaje natural se realiza con la minería de opinión y preferencias de los comentarios analizados del sitio web Foros Perú.

Dimensión Precisión:

Con respecto al objetivo específico 3: Verificar el modelo de preferencias de usuario, para mejorar el intercambio de información en el sitio web Foros Perú.

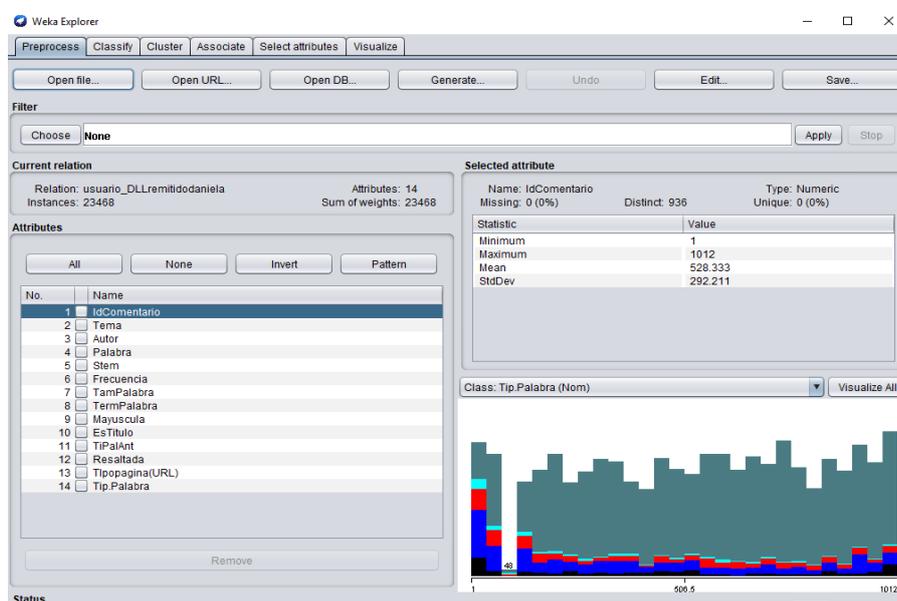


Figura 11.-Prototipo para analizar las estadísticas WEKA, para realizar análisis estadístico inferencial

Interpretación: La captura de pantalla del programa elaborado muestra estadísticamente el modelo de preferencias de usuario que mejora el intercambio de información en el sitio web foros Perú.

Dimensión: Precisión

Con respecto al objetivo 4.- Verificar la estructura de un sistema web basado en el modelo de preferencias de usuario en el sitio web Foros Perú.

Interpretación: A continuación se muestra la estructura de la tabla usuario y blog respectivamente que contiene los metadatos implementados para poder hacer el sentiment analysis, y en esa forma se verifica la estructura del sistema web basado en el modelo de preferencias de usuario en el sitio web Foros Perú, y sus respectivas estadísticas.

1.Sentiment Análisis de usuario

En esta sección, los 1000 comentarios analizados de la pagina Foros Perú, al realizar su proceso estadístico, se obtiene y evalúa el archivo usuario.csv, que tiene 23468 registros con las variables ó metadatos de la Tabla 1.

variable	tipo	Valores distintos	Valores únicos	Valores faltantes
IdComentario	Numérico (type)	936	0	0
Tema	Nominal	1	0	0
Autor	Nominal	37	0	0
Palabra	Nominal	3398	0	0
Stem	Nominal	2316	0	0
Frecuencia	Numérico	2	0	0
TamPalabra	Numérico	3	0	0
TermPalabra	Nominal	21	0	17109
Mayúscula	Nominal	2	0	0
Estitulo	Nominal	2	0	0
TipPalAnt	Nominal	5	0	825
Resaltada	Nominal	2	0	0
TipoPagina	Nominal	2	0	0
Tip.Palabra	Nominal	5	0	824

Tabla 2.- Detalle de variables en archivo usuario

2. Descripción de las variables ó metadatos analizados por el prototipo y que están alineados a las dimensiones de tiempo, predicción, comprensión:

En este archivo, el prototipo USUARIO analizado, los metadatos que se va extraer son las siguientes variables:

1. **IDComentario** es el identificador único de cada comentario capturado de la página WEB Foros Perú (sección universidades). Es un número entero correlativo que

- enumera a partir de la aplicación generada como parte del prototipo de implementación del modelo que aquí se presenta.
2. **Tema**, es el título general de los comentarios. En este caso es un único tema: “Los thread de los estudiantes de economía “
 3. **Autor**, es el alias del autor dentro del sitio en estudio. Se han procesado 37 autores diferentes.
 4. **Palabra**, es la palabra completa original que compone cada comentario. La misma, ha sido pre-procesada por el prototipo, eliminando caracteres especiales, convirtiendo todo a minúsculas, y elimina números. Se considera palabra a toda sucesión de caracteres entre espacios.
 5. **Stem**, es la partícula de Porter, tal como se describió en la sección de “Materiales y Métodos”. El algoritmo fue implementado como parte del prototipo, para el idioma español, y aplicado a las palabras (Palabra)
 6. **Frecuencia**, es la cantidad de veces que aparece Palabra dentro del archivo. Esto lo implementa el prototipo.
 7. **TamPalabra**, es una categorización que determina la cantidad de palabras según la cantidad de caracteres. Los valores posibles son: {corta, media, larga}. Esto lo implementa el prototipo.
 8. **TermPalabra**, es la caracterización sistemática de la terminación de las Palabra. Entre los valores: ar, er, ur, an, un, ra, ru, n, m, etc. Esto lo implementa el prototipo.
 9. **Mayúscula**, indica si la palabra originalmente comienza o no con mayúsculas. Sus valores posibles son Si o No. Esto lo implementa el prototipo.
 10. **Estitulo**, el prototipo aquí determina si la palabra es parte o no del título del comentario.
 11. **TipPalAnt**, aquí el prototipo determina si es {adjetivo, vrbo, adverbio, sustantivo, otros}
 12. **Resaltada**, consiste en determinar si originalmente la palabra ha sido encerrada entre comillas dobles o simples, o resaltada por otros medios. El prototipo lo deriva a raíz del contexto.
 13. **TipoPagina**, es un campo legacy donde se define si la página de donde se extrajeron las palabras son índices o no. Esto determina si tiene muchos hipervínculos, lo que estaría reflejando muchas referencias cruzadas.
 14. **TipPalabra**, al igual que **TipPalAnt**, consiste en el tipo de palabra analizada. El prototipo sigue los mismos lineamientos.

Se comienza evaluando las características comunes de los usuarios que realizan sus comentarios en el foro analizado de la página web Foros Perú, que se han extraído y puesto en un archivo usuario.csv; según su comportamiento textual.

En tal sentido, se ejecuta el tipo de agrupamiento denominado Maximización Esperanza (EM) en WEKA ©. Que da por resultado la tabla que se muestra en la tabla 2 (Ver Anexo A para más detalles).

ID Agrupamiento	Cantidad de registros (porcentaje)
0	13702 (58%)
1	9766 (42%)

Tabla 3.- Datos agrupados según EM

El valor de Log likelihood es -11.74261 y el tiempo que tardó la plataforma para construir el modelo es de 33.17 segundos (time taken to build model – full training data).

Los parámetros del EM que se usaron son los que se indica en la tabla 1:

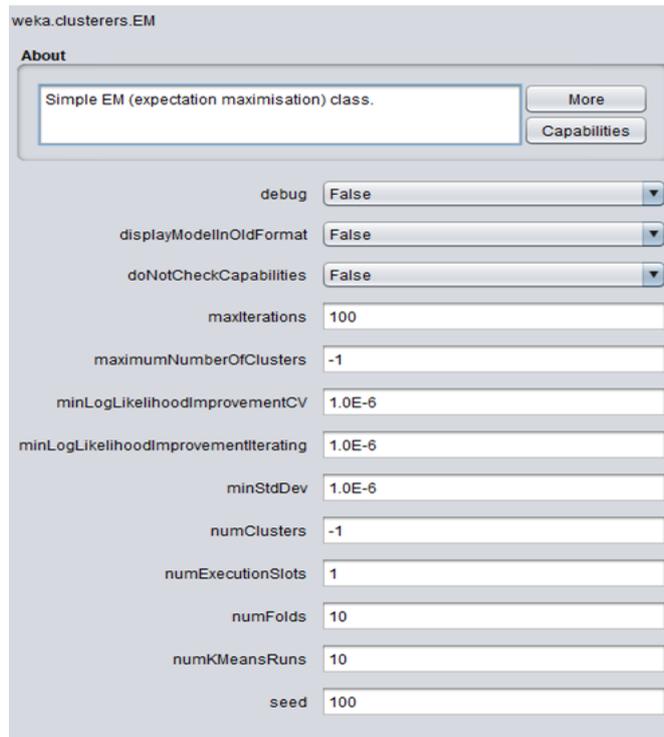


Figura 12.-Parámetros usados para el EM

Por los resultados expuestos, se consideran las asignaciones por **STEM** (eje Y de ordenadas) y **AUTOR** (eje X de abscisas), según se muestran en la figura 12.

En cuanto a la relación de la clase Tip.palabra, se observa en forma grafica:

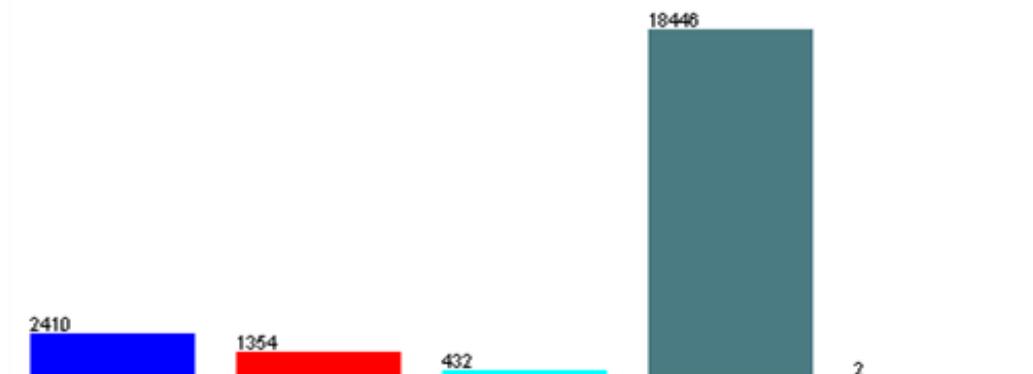


Figura 13.-Asignaciones de agrupaciones

A continuación, observamos los colores de las agrupaciones TipPalabra en la tabla 3.

Tabla 4.-Colores de las agrupaciones en la Figura 13.

Nº	Agrupación (TipPalabra)	Color
1	Sustantivo	Azul
2	Verbo	Rojo
3	Adjetivo	Celeste
4	Otros	Gris oscuro
5	Adverbio	Gris

Interpretación: Muestra las estadísticas de los comentarios analizados del modelo de preferencias de usuario mediante la aplicación del algoritmo EM de maximización – esperanza, para conocer que tipo de palabras ha detectado respectivamente.

3. Sentiment Análisis de blog

Esta sección corresponde al proceso estadístico de 1000 comentarios analizados, específicamente al archivo blog.csv tiene 936 registros y las variables de la Tabla 4.

Tabla 5.- detalle de variables en archivo blog

Variables	Tipo	Valores Distintos	Valores unicos	Valores faltantes
IDComentario	Numérico	936	936	0
Autor	Nominal	37	8	0
Comentario	Nominal	893	851	0
TamComenconEspacio	Numérico	204	45	0
TamComesinEspacio	Numérico	173	28	0
CantPalConRep	Numérico	26	3	0
CantPalSinRep	Numérico	25	3	2

Nota.- son los metadatos o variables establecidas para la tabla blog.

4. Descripción de los metadatos analizados por el prototipo y que están alineados a las dimensiones de tiempo, predicción, comprensión:

En este archivo, el prototipo BLOG analizado los metadatos que se va extraer son las siguientes variables:

1. **IDComentario**, es el mismo que en el archivo usuarios.csv
2. **Autor**, es el mismo que en el archivo usuarios.csv
3. **Comentario**, es el mismo que en el archivo usuarios.csv
4. **TamComenconEspacio**, es la cantidad de caracteres del comentario al que pertenece Palabra (del archivo usuarios.csv), contando los espacios en blanco intermedios.
5. **TamComesinEspacio**, es la cantidad de caracteres del comentario al que pertenece Palabra (del archivo usuarios.csv), sin contar los espacios en blanco intermedios.
6. **CantPalConRep**, es la cantidad de veces que aparece cierta **Palabra** en el comentario al que pertenece Palabra. Se cuentan las repeticiones.
7. **CantPalSinRep**, es la cantidad de veces que aparece cierta **Palabra** en el comentario al que pertenece Palabra. No se cuentan las repeticiones.

Se realiza un análisis de agrupamiento por *Expectation Maximization* (EM), con los parámetros de la tabla 4.

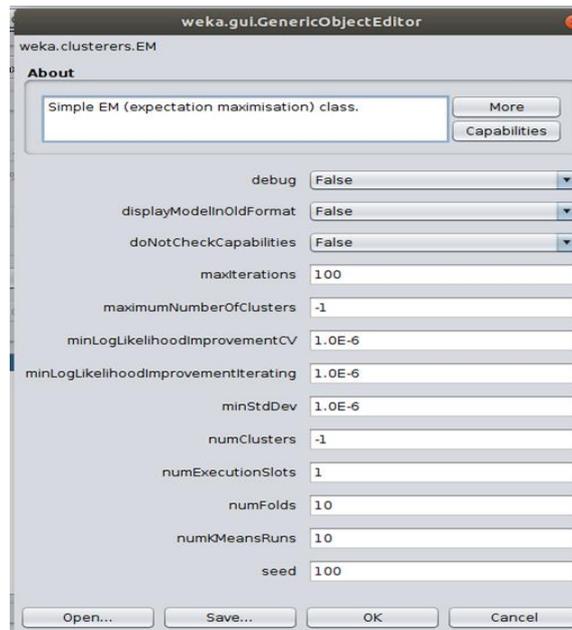


Figura 14.-Detalle de Parámetros empleados para EM

Nota.- parámetros que emplea algoritmo maximización esperanza de Weka para procesar el conjunto de datos blog con sus parámetros.

El tiempo total tomado para el entrenamiento es de 737.57 segundos. Y el modelo entrega 10 agrupaciones naturales tal como se muestra a continuación en la tabla 5, lo que significa que los comentarios analizados los agrupa en 10 cluster o agrupaciones y le pone un numero de orden, y su respectivo promedio de cuantos palabras de los comentarios analizados los ha asignado a cada cluster respectivamente.

#cluster (nro de orden del grupo de los comentarios)	cantidad instancias (%)
0	146 (16%)
1	170 (18%)
2	39 (4%)
3	39 (4%)
4	115 (12%)
5	222 (24%)
6	65 (7%)
7	34 (4%)
8	38 (4%)
9	68 (7%)

El Log likelihood: -14.53056, y el detalle de los eigenvectores son los que muestra la tabla 9 (para mayores detalles ver el Apéndice B).

Tabla 6.- Detalle de Agrupaciones Naturales

Dimensión Precisión:

Con respecto al objetivo 5: Demostrar la validación y corroboración de los resultados alcanzados en el estudio.

Interpretación: La validación, corroboración y demostración de los resultados alcanzados en el presente estudio se verifica en lo descrito a continuación, debido a que la presente investigación trata los 1000 comentarios analizados de la pagina web Foros Perú, la validación se muestra indicando que las estadísticas de WEKA y el programa elaborado para realizar el proceso de lenguaje natural (wavelets morfosintacticas), refieren que la predicción anónima que hace el computador, comparación real de mensajes en el sitio web Foros Peru, y análisis de diferencias entre precisión, comprensión y significado, permite realizar la inferencia para conocer que grupo de usuarios tiene polaridad positiva, precominan mas el significado de sus mensajes o negativa cuando predominan menos el significado de sus mensajes.

3.2. Discusión de resultados

Según (Witten et al., 2017), en la presente sección se compara y contrasta el pasado con el presente, y de acuerdo a los estudios previos referidos en la parte de antecedentes y los hechos referidos en la realidad problemática; se ha establecido las diferencias así como las coincidencias a partir del reconocimiento de las debilidades (limitaciones) y fortalezas (aportes) del estudio realizado, logrando analizar y explicar los resultados según se menciona a continuación:

Con respecto al Objetivos General

Evaluar el sistema predictivo basado en el modelo de preferencias para la predicción de preferencias con prototipo de usuarios en el sitio web Foros Perú.

- La validez interna y externa de los resultados permite evaluar el sistema predictivo y compara el pasado, el presente, evitando el sesgo del investigador y esta fase está controlada por la capacidad que presenta el modelo de predicción de preferencias de usuario implementado a partir de las expresiones escritas o mensajes comentados en la página web Foros Perú.

Estos datos se han transformado en resultados en tres fases:

- a) Predicción anónima que realiza el computador con el prototipo de los comentarios.
- b) Comparación real de mensajes en el sitio web Foros Perú, verificando los metadatos o variables de la tabla 1 y tabla 4 respectivamente.
- c) Análisis de diferencia entre precisión del modelo práctico y la realidad. Por lo que, analizados los resultados, la aplicación se ejecuta como herramienta de procesamiento de lenguaje que vienen a ser los wavelets morfosintaticos dentro del sitio web Foros Perú según la modificación de los procesos de entendimiento y significado respectivamente.

Con respecto al objetivo, comprobar las características epistemológicas del procesamiento de lenguaje natural y minería de opinión mediante el *sentimental analysis*.

- Del Prototipo implementado para realizar el “*Sentiment Analisis*”, se han obtenido dos archivos de datos intermedios:

1. usuario.csv

2. blog.csv

Con respecto al objetivo, determinar el diagnóstico del estado actual de la dinámica del procesamiento de lenguaje natural NLP con la minería de opinión y predicción de preferencias en el sitio web Foros Perú, expresado por el desinterés de los especialistas de TI en las organizaciones debido a que se requiere conocimiento matemático, estadístico detallado, el intercambio de información entre usuario es grande ya que todos hoy en día hacemos mensajes escritos (textos), existe el temor de que la información sea mal interpretada por algunos usuarios maliciosos, deficiencia en la implementación de políticas para entender los mensajes escritos por los usuarios, así como protocolos de seguridad que no garantizan la autenticación y confidencialidad de la interpretación de mensajes.

- Se conoce el estado actual que las palabras sin realizar el análisis no indican nada y a partir de los análisis realizados sobre *usuario.csv*, se estudiaron las agrupaciones naturales por medio del algoritmo Heurístico EM, y se observa un *likelihood* de -11.74261, que, si bien es el menor entre los alternativos esquemas de agrupación, es superior al que se obtiene para su contraparte en el estudio de *Sentiment Analisis: blog.csv*

Con respecto al objetivo verificar el modelo de preferencias de usuario, para mejorar el intercambio de información en el sitio web Foros Perú

- Es notorio que exista una distribución similar en los dos clúster resultantes, pero el clúster 0 denota autores, como usuarios predominantes en las conversaciones.
- El significado de esta distribución se ha precisado en la sección correspondiente a las estadísticas y aquí cabe destacar que demarca autores con mayor expresividad gramatical (mayor extensión de palabras). También se ha hallado que dicho cluster

(en el análisis denotado como cluster 0) se caracteriza por palabras de deseos y caracterizaciones de hechos y cosas. Su contraparte, en cambio denota estados de ánimo y apreciaciones.

- Como consecuencia, se puede decir que los autores de los comentarios con prevalencia en la agrupación ó cluster 0, aportan información propia acerca de sus **necesidades ó peticiones**. En cambio, los de su contraparte, (agrupación o cluster 1), estarían evaluando o aportando información **con apreciaciones personales ó estados de ánimo**.
- Con esto, se puede decir que el prototipo es capaz de determinar con el eigenvector del cluster 0 a los comentarios de interés, y con el del clúster 1 los comentarios que son superfluos.
- Del *Sentiment Analysis* surge entonces, que el prototipo es capaz de detectar a partir del contenido textual, información de potencial valor e interés para estudiar el comportamiento de individuos opinando en la web Foros Perú.

Con respecto al objetivo verificar la estructura de un sistema web basado en el modelo de predicción de preferencias de usuario en el sitio web Foros Perú.

- Se verifica según las estadísticas se denota la posibilidad de identificar comentarios repetidos o que parafrasean otros aportes.
- Del estudio de blog.csv, se obtuvieron 10 agrupaciones naturales, con un log likelihood de -14.53056, que es menor. Lo que da a pensar que posiblemente la muestra sea más representativa de la agrupación obtenida, o bien que el modelo es más ajustado que en usuarios.csv.

Con respecto al objetivo demostrar la validación y corroboración de los resultados alcanzados en el estudio.

- Es importante precisar que a partir de esta otra información extraída por el prototipo se detentan individualmente cuáles fueron los usuarios (para el estudio

realizado, los identificados como flecsdiego y jq13) cuyos aportes marcan tendencia en cuanto a contenido. Esta información, cruzada con la obtenida del EM en usuarios.csv, permite obtener no sólo al influencer, sino también conocer cómo opina y sobre qué está opinando: basta con evaluar los eigenvalores correspondientes al cluster 0 del EM correspondiente a usuarios.csv

- Asimismo, el prototipo es capaz de detectar otros usuarios (Lovin y Deru), como **seguidores activos**, y los puede discriminar del resto de los usuarios que son **seguidores menos activos**.

3.3. Aporte teórico

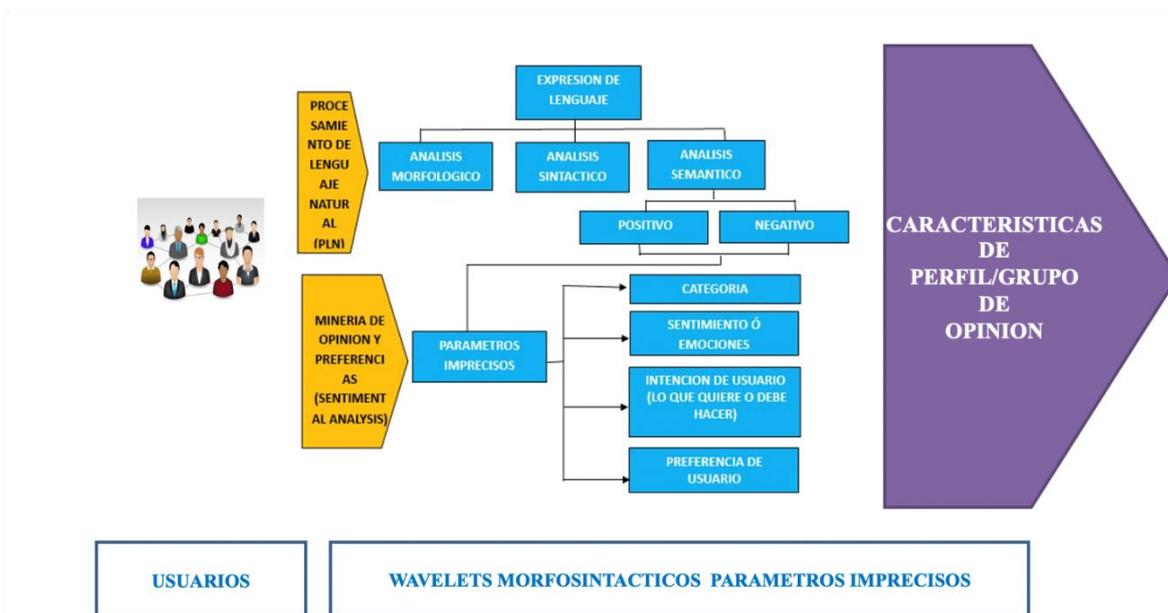


Figura 15.-Arquitectura propuesta para el sistema predictivo – Modelo Propuesto

3.3.1. Fundamentación del aporte teórico

El Modelo de predicción de preferencias de usuarios, esta basado en la teoría de análisis de palabras propuesta (Wavelets morfosintacticas) que se muestra en la presente investigación al emplear el modelo (Ver Figura 1); basado en la

Predicción anónima, Comparación real de mensajes en el sitio web Foros Perú, Análisis de diferencia entre precisión del modelo práctico, la realidad y la etapa final del aporte que consolida la inclusión de las nuevas tecnologías como el procesamiento funcional mediante la medición, predicción y calificación inteligente empaquetadas en un sistema predictivo fundamentado en las dimensiones de tiempo, predicción (inferencia) y comprensión (significado de mensajes en base a los valores numericos mostrados en la tabla 6); relacionados de acuerdo a los wavelets morfosintácticos del proceso de lenguaje natural, que permiten un adecuado proceso moderno de convertir texto, palabras en vectores numericos que permiten el aprendizaje automático con inteligencia artificial.

En la experimentación se evidencia unas ligeras variaciones fruto de un probable sesgo humano; y como conclusión se ha mostrado los fundamentos, las metodologías para realizar análisis de palabras sobre un foro de comentarios denominado Foros Perú; teniendo en cuenta que la expresión del lenguaje natural utiliza el análisis morfológico (palabras), análisis sintáctico (oraciones y frases), y finalmente el análisis semántico (significado) que identifica la polaridad de los textos descritos en internet a fin de sugerir ciertas predicciones (vectores numericos o coeficientes wavelets obtenidos con el algoritmo esperanza maximización) sobre los gustos, preferencias o perfiles de influencia de usuarios.

El modelo presentado en la presente tesis, muestra la predicción de preferencias de usuarios y estará fundamentado en la capacidad de aprendizaje que se valida con el algoritmo (Maximización esperanza) y la aplicación del sentimental analysis para que se pueda realizar predicciones de preferencias de usuario, de acuerdo al análisis realizado en un grupo de mensajes escritos por un grupo de usuarios, donde se considera los siguientes parámetros imprecisos: a) nivel de comprensión. b) Clasificación de mensajes de usuarios que predominan más.

En tal sentido, como las variables obtenidas con el prototipo mostrado, son numéricas se utilizará para validar la estadística inferencial, con el algoritmo de EM Expectation Maximization de Weka, y para el proceso de lenguaje natural de los comentarios analizados el prototipo emplea y determina su precisión, comprensión, basado en la aplicación de las wavelets morfosintácticas, logrando el resultado como se denota en la **Tabla 1**.

3.3.2. Descripción Argumentativa del aporte teórico

El prototipo para poder realizar la predicción de preferencias de usuarios, capta parámetros con las wavelets morfosintácticas de los comentarios registrados que son los que permiten la predicción de preferencias, con el empleo de la minería temporal (mlw) y minería tradicional (por ejm: clustering), estos son los siguientes variables o también llamados metadatos:

- **Palabra:** palabra seleccionada del comentario que quedo del proceso realizado.
- **Raíz o stem de la palabra:** es el stem o la raíz de la palabra seleccionada y es extraída con el algoritmo de porter ¿este es la raíz cierto?
- **tema:** tema del que trata el comentario y opinión
- **palAntTipo:** tipo de palabra que tiene la palabra anterior a ésta (sustantivo, verbo, otra), Si no hay anterior se coloca null y con esto además se determina si es la primera palabra procesada en la sentencia.
- **tipoPagina:** índice o contenido
- **terminacion:** si es ar, er, ir, or, ur, ra, re, ri, ro, ru, s, m, sa, se, si, so, su, an, en, in, on, un,cion, ciones, null.
- **empiezaMay:** es SI cuando la palabra empieza con mayúscula y NO cuando empieza con minúscula.
- **resaltada:** es SI cuando la palabra entera está entre comillas, apóstrofes o toda en mayúsculas
- **esTitulo:** es SI cuando la palabra está contenida en el título (en nuestro caso sería en el tema)
- **catLongPal:** se cuenta la cantidad de caracteres en la palabra original y luego se determinan tres categorías (corta, media, larga) según tenga menos de 5 caracteres, menos de 10 o más.

Asimismo, la predicción de preferencias de usuario ó también se le puede denominar la mejora de las insuficiencias del proceso de lenguaje natural, se comprueba con el log likelihood y kappa statistics; indicadores que se encuentran calculados estadísticamente, con el Weka (**ver anexo 6 y 7 cluster – archivos usuario y blog**).

Por lo expuesto, el prototipo representa el modelo de predicción de preferencias de usuario y viene a ser la **novedad** que se propone, puesto que mediante este aplicativo implementado ver (**Figura 1**); se puede validar el resultado procesado con minería temporal y minería tradicional, se puede indicar que predice un cluster (**ver cluster en anexo 6 y 7**); para un comentario del foro: dice si el comentario es de un influencer y el tipo de comentario.

Por lo expuesto, con la combinación de Wavelets Morfosintácticos MLW del algoritmo Expectation Maximization (Esperanza - Maximización) y técnicas tradicionales del PLN

para obtener perfiles de opinión, y con ayuda de la estadística se logra realizar inferencias, encontrando estimadores de máxima verosimilitud de parámetros en el modelo probabilístico que dependen de variables no observables (modelo de predicción de preferencias de usuarios).

3.4 Aporte práctico

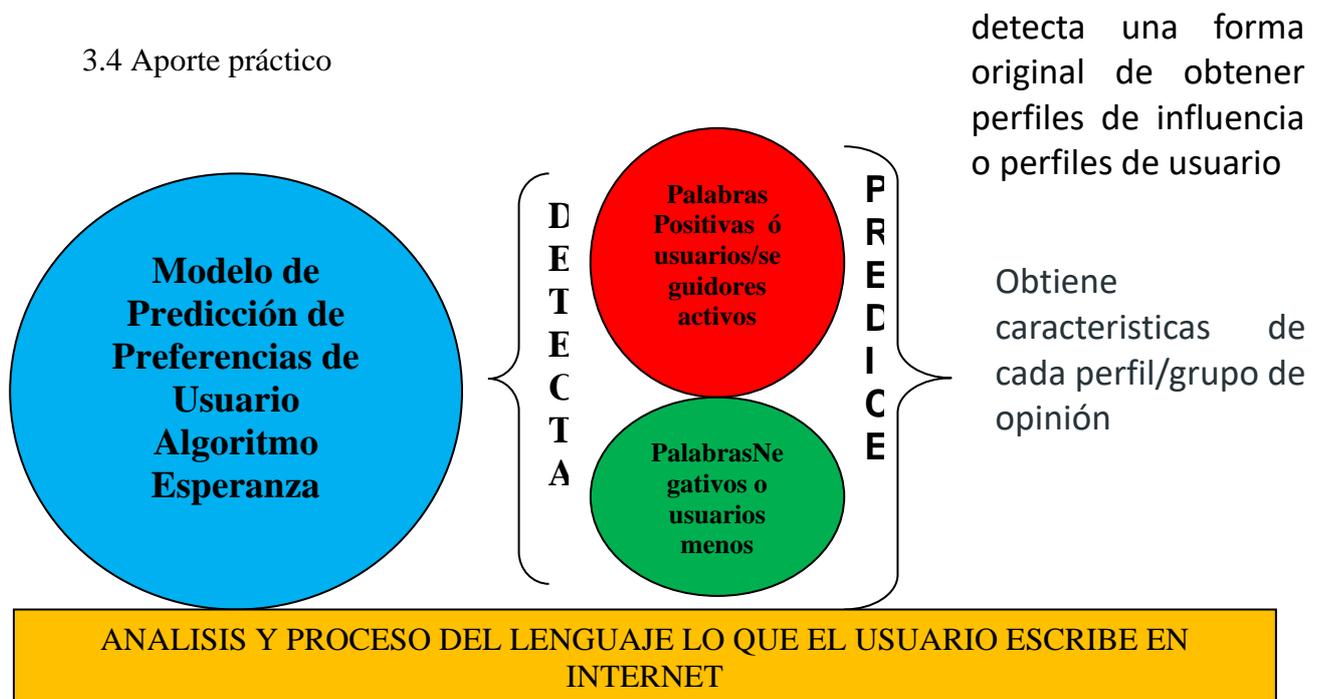


Figura 16.-Modelo Propuesto – Aporte Práctico – Modelo Web

3.4.1. Fundamentación del aporte práctico.

Con la minería de opinión y preferencias, (en la investigación se hace mediante el análisis de palabras) se observa que el denominado modelo de predicción de preferencias de usuario, inicia a partir del procesamiento de lenguaje natural que utiliza la nueva forma de procesamiento wavelets morfosintácticas (que implica los tipos de análisis léxico, morfológico, sintáctico y semántico respectivamente).

En la presente investigación se considera llegar a conocer el significado (análisis semántico) de las palabras extraídas de cada comentario con el fin de verificar si tiene polaridad positiva o negativa, es decir que palabras predominan más que otras en los comentarios analizados, para lograr una mejor **COMPRESION** en un tiempo determinado.

Cuando se tienen en cuenta todos los parámetros imprecisos (metadatos) de los comentarios seleccionados, entonces se podrá verificar el significado (análisis semántico) de las palabras

descritas en dicho comentario y de esa forma podrá elegir si la palabra tiene polaridad positiva o negativa, logrando de esta manera inferir (PREDICCIÓN) en un tiempo determinado, que permita tomar una decisión adecuada para que el usuario realice su compra, venta o interacción que sea de su preferencia.

La transformación que se realiza con el modelo presenta tres fases como a) Predicción anónima, b) Comparación real de mensajes en el sitio web Foros Perú, y el c) Análisis de diferencia entre precisión del modelo práctico y la realidad.

Por lo que, analizados los resultados, la aplicación se ejecuta como herramienta moderna convertir palabras textos en vectores numéricos o coeficientes wavelets extraídos del sitio web Foros Perú teniendo en cuenta las dimensiones de tiempo, comprensión y significado respectivamente.

3.4.2. Construcción del aporte práctico

El modelo propuesto realice la derivación de reglas y relaciones funcionales entre el contenido del archivo blog y (**Ver Tabla 4 y Tabla 7**); que representa el estado de opinión de usuarios; para conocer que usuario es más activo y/o menos activo; y la derivación se fundamenta en las estadísticas y sus métricas; Por tal motivo, los pasos a seguir que se tienen en cuenta son:

1. Registro de comentarios en el prototipo usando las wavelets morfosintácticas (mlw) (**Ver Figura 12**)
2. Proceso del prototipo consiste en generar una lista de palabras con sus raíces sin preposiciones, símbolos ni caracteres especiales en una planilla que vienen a ser los cluster y eigenvectores (**Ver Anexo 6 y 7**).
3. Ahora los eigenvalores (es decir los valores de las variables actuales) se combinan con el eigenvector para obtener la predicción. Si $A =$ vector de datos=1 renglón de tu archivo y $B =$ eigenvector de cluster 0 entonces con el producto interno $A \cdot B$ da el resultado de la predicción (**Ver Anexo 6 y 7**)

Por lo expuesto, se considera que con la sistematización del modelo a través de un prototipo, de la presente investigación; se representa el sistema predictivo del modelo de predicción de preferencias que analiza los comentarios palabra por palabra, empleando los wavelets morfosintácticos, algoritmo de Porter, algoritmo de maximización esperanza, extrae sus raíces morfológicas ó STEM, y que actualmente es la propuesta para realizar análisis de

parser ó palabras (minería de opinión y preferencias), por lo que toda innovación en este campo aportará a dinamizar dicho proceso (**Ver figura 12, Anexo 6, Anexo 7**).

El impacto para la sociedad, y la **significación práctica** de la presente investigación consiste que el modelo de predicción de preferencias de usuario busca lograr perfiles de influencia ó características de cada perfil de opinión, con respecto al uso de nuevas tecnologías, para que los sistemas de mercadeo incluyan variables socioculturales en el mercado; en generar un impacto en el análisis de palabras y comentarios en el sitio web Foros Perú, teniendo en cuenta ciertos parámetros de referencia, que ayuden a reconocer mejor el comportamiento de los usuarios y que coadyuve a obtener las características de un formador de opiniones enfocados en la dimensión de comprensión ó significado, con precisión y tiempo respectivamente; permitiéndole a los usuarios aprovechar la continua interacción con redes sociales y centrándose en diferentes pautas que permiten predecir la intencionalidad de compra, venta, aceptación, preferencia, ó demostración de interés basada en foros como es el sitio foros Perú en la sección universidades. (**Ver Tabla 7**)

3.5. Valoración y corroboración de los resultados

Tabla 7.- Detalle de Variables en archivo Blog

Atributos	0	1	2	3	4	5	6	7	8	9
Autor	0.15	019	0.04	0.03	0.12	0.24	0.07	0.03	0.04	0.07
Jldark89	10.3315	21.4243	2.042	5.5897	13.5498	40.722	8.1048	2.4674	2.7558	24.012
Deru	12.4814	11.4384	8.3437	3	11.9609	16.2483	15.8791	3.1648	4.8573	9.6263
metalerojc	2.0377	3.9985	1.0179	1	1.0015	4.9623	1.9821	1	1	1
Tribolon	1	3.0488	1	1.9512	1	1	1	1	1	1
gsuz	1	3.0032	2	2.996	1.0178	1.0149	1	1.8329	1	1.1352
Fiescdiego	15.3373	32.4918	2.0003	7.1659	4.0065	27.2492	2.414	11.2724	9.9946	11.0682
Ofiblion	2.0096	1.5414	1.0013	3.4606	6.0907	4.7072	1.9992	2.0023	1.1661	2.0217
Jq13	77.4991	93.1864	14.2126	13.2457	61.697	116.5591	15.5719	7.8367	10.8352	21.3562
Pet924	1.4664	3.0007	1.0002	1	1.9611	2.895	2.5379	1.0752	2.9885	1.075
szf	5.8187	3.9315	3.0132	2	6.4869	3.2072	1.9965	1.9153	2.0234	2.6074
Ofiblion	2.0258	1.0012	1.0037	1	3.8955	2.6813	2.9706	2.1119	1.0361	2.2739
lovin	3.045	2.0405	1	1.9605	3.9849	1.9537	2.0013	1.0137	1	1.0003
Luchex2710	1.9953	1	1	1	1	1.0047	1	1	1	1
deru	1	1.9622	3	1	1.0378	1	1	1	1	1
mickly	1.8147	1	1	1	1.5381	1.1853	1	1.0018	1	1.4601
nauttico	1	2	1	1	1.2295	2.9954	1	1.7558	1	1.0193
cera	1.9823	1	1	1	1	1	1.0177	1	1	1
Cera	1.0154	1	1	1	1	1	1.9846	1	1	1
Redo	1	1.9189	1	2	1	1	1	1.0811	1	1
estrelladefuego	1.0192	1.9743	3.0094	1.0259	1.8677	1.9948	2.9714	1.1315	1	1.0059
Lovin	15.4884	5.2408	8.0113	1.0286	7.2213	7.588	10.7047	4.3841	2.2775	2.0548
Blake	1	1	1	1	1.0253	1.0005	1	1.0074	1	1.9669
Jldrak89	1.0001	1.9985	1.0055	1	1.0015	1	1.9945	1	1	1
sir.kane	1.0126	2	2.9995	1	1	1.9878	1.0005	1.006	1.9905	1.0031
Nova00	1.1863	1	1	1	1	1.0579	1.8137	1.0157	1.9263	1
lp1	1	1	1	1	1.0225	2	1	1.0145	1	1.963
jklm92	1.9955	1	1	1	1	3.0138	1.0047	1.0184	1.9654	1.0022
la unica	1	2.905	1	1	1.095	1	1	1	1	1
Deru	1.0086	1	1	1	1	1	1.9914	1	1	1
Metalerojc	1	1.0002	1	1	1.9983	2	1	1.0002	1	1.0013
Jorge	1	1.9987	1	1.0013	1	1	1	1	1	1
DiegoD	1	1	3	1	1	1	1	1	1	1
jldark	1	1	1	1	1	1.9591	1	1.0141	1	1.0268

Ofibilion	1	1.0001	1	1	1.5601	1.9263	1	1.43452	1	1.0784
blake	1	1	1.006	1	1	1	1.994	1	1	1
fernando123	1	1.0008	1	1	2.9321	1.0032	1	1.0083	1.969	1.0065
jldark90	2.9983	1	1.0017	1	1	1	1	1	1	1
Total	178.4903	219.1856	78.6664	69.4271	153.1819	264.9168	99.0158	66.5652	69.7856	106.7651

TamComenConEspacio

Mean	102.53 98	193.6982	42.2743	230.0522	164.32 24	133.34 26	74.536 6	149.4691	139.8305	147.4368
Std dev	10.8832	13.0788	15.772	9.9907	8.1029	14.175 5	8.914	6.6767	4.7014	4.7781

El tamaño en caracteres con espacios de cluster 0, 1 y 5 son:

cluster 0 (0.15%) entre [102.5398 – 10.8832, 102.5398 + 10.8832] = [916566, 916566]

cluster 1 (0.19%) entre [193.6982 – 13.0788, 193.6982 + 13.0788]

cluster 5 (0.24%) entre [133.3426 – 14.1755, 133.3426 + 14.1755] = [1,191,671, 1,191,671]

➤ Explicación de la tabla 6, archivo blog:

CARACTERISTICAS DE LOS PERFILES DE USUARIOS ANALIZADOS CON EL SISTEMA PREDICTIVO BASADO EN EL MODELO DE PREDICCIÓN DE PREFERENCIAS DE USUARIOS DEL SITIO WEB FOROS PERÚ.

Los resultados muestra que los 1000 comentarios procesados principalmente tienen longitudes bastante extensas, siempre promediando todos los caracteres, se observa en los coeficientes o vectores numéricos obtenidos que: **LOS COMENTARIOS MÁS CORTOS APARENTEMENTE SON ACCESORIOS y EL CONTENIDO DE LAS COMUNICACIONES ESTÁN EN LAS SENTENCIAS AGRUPADAS EN LOS 3 CLUSTERS SELECCIONADOS (0, 1 Y 5).**

Observando la tabla 6, los usuarios **resaltados (FLECS DIEGO Y JQ13)** son los que más aportes han tenido a estos clusters, y por lo tanto los que más tendencia están fijando en cuanto a contenido.

Asimismo, otros usuarios subrayados (**LOVIN Y DERU**), serían **seguidores activos**, y el resto de los usuarios **simplemente seguidores menos activos**.

En conclusión, **LOS USUARIOS QUE MARCAN TENDENCIA, ES DECIR HAN MOSTRADO OPINIONES QUE PREDOMINAN O SON USUARIOS MAS ACTIVOS EN LOS COMENTARIOS ANALIZADOS SON FLECS DIEGO Y JQ1.** Observando sus comentarios (ver discusión final) probablemente se pueden inducir la tendencia de opinion del conjunto de usuarios.

Las figuras que se muestran a continuación corroboran los resultados:

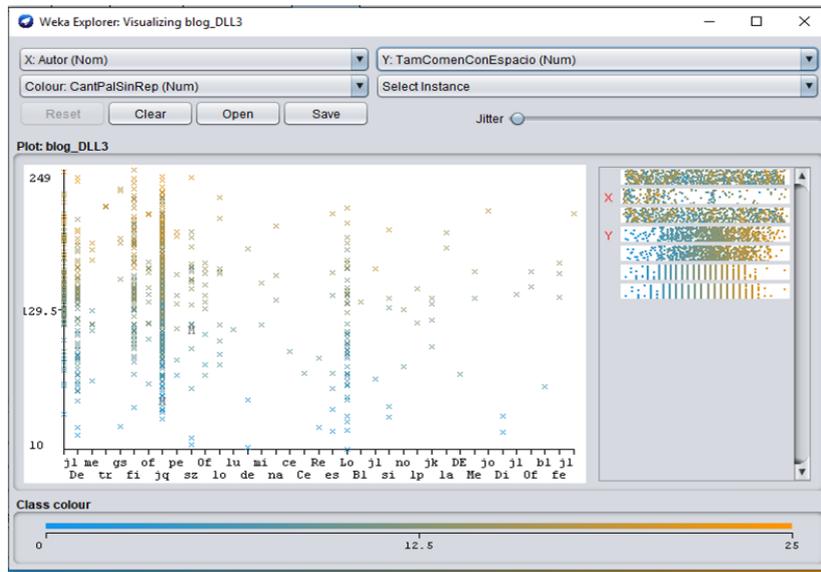


Figura 17.- Autor VS Cantidad de caracteres (tamaño comentarios) con espacios
 Nota.- muestra la distribución de datos autor vs cantidad de caracteres procesados incluyendo espacios en blanco

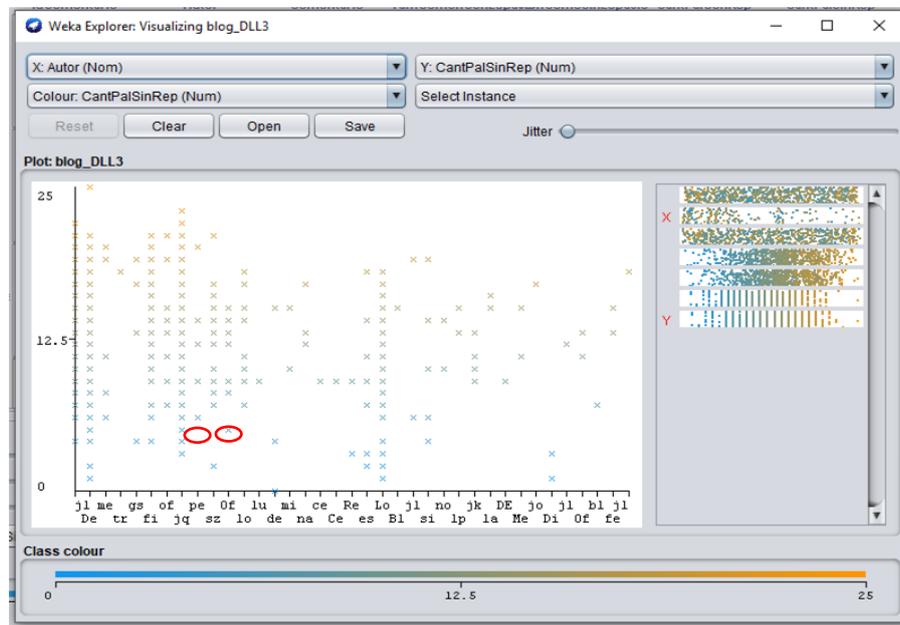


Figura 18.- Autor VS Cantidad de palabras sin repetición
 Nota.- muestra la distribución de datos autor vs cantidad de palabras que no se repiten en el dataset procesado

Una vez predichos los resultados por del algoritmo EM aplicado se procede a obtener la comparativa del nivel de precisión y significado de los mensajes descritos en los comentarios

analizados del sitio web FOROS PERU. Es así como se puede corroborar las manifestaciones, comportamientos de los usuarios que opinan con sus comentarios así como también con los resultados mostrados de estadística inferencial con el programa WEKA.

Por lo expuesto es así como se demuestra el cumplimiento de estimación probabilística para obtener la precisión y significado de los mensajes, que aporta con la inferencia realizada por el algoritmo de maximización de la esperanza de Weka, a convertir palabras, textos, en valores numéricos o coeficientes waveletes a partir del tratamiento de mensajes del procesamiento de lenguaje natural PLN.

3.5.1 Valoración de los resultados (taller de socialización, criterio de expertos, etc.)

Valoración de los resultados por juicio de expertos

Se evalúa la pertinencia del modelo de predicción de preferencias de usuario en el sitio web Foros Perú. La propuesta se basa en la investigación exploratoria, mediante la observación y el análisis efectuado del Proceso de Lenguaje Natural PLN, que sustentan el aporte teórico

teniendo como base el método inductivo – deductivo, dado que se basa en el análisis de mensajes de textos descritos por usuarios en el sitio web Foros Perú, que son el elemento principal de análisis y evaluación con el prototipo, todo ello permitirá lograr la integración necesaria en el aporte práctico. La valoración de los resultados del aporte teórico y práctico se realiza con expertos, a los cuales se ha validado su competencia respecto a la línea de investigación de la propuesta. Se seleccionaron 4 expertos teniendo en cuenta ciertos criterios como la Experiencia Profesional en Tecnologías de la Información y Comunicaciones, Título profesional Ingeniero de Sistemas y Grado Académico de doctor logrado.

Después de aplicar el instrumento de evaluación de competencias de los expertos se llegó a los siguientes resultados:

N°	Experto	Título Profesional	Grado Academico
1	Dra. Sofia Irene Delgado Wong	Licenciado en Administración e Ingeniera Informática y de Sistemas	Doctora en Administración
2	Dr. Wilton Vidauro Carpio Campos	Ingeniero Informatico y de Sistemas Licenciado en Administración Licenciado en Educación	Doctor en Ciencias de la Educación
3	Dra. Karla Oblitas De las Casas	Ingeniero Químico Licenciada en Educación	Doctora en Ciencias de la Educación

Tabla 8.-Datos de los expertos

	Kc	Ka	$K=1/2(Kc-Ka)$	Competencia
1	1.0	1.0	1.00	Alta
2	0.9	1.0	0.95	Alta
3	0.9	1.0	0.95	Alta

Tabla 9.- Resultados de calificación de expertos participantes

Como se observa el resultado de la competencia de los expertos es Alta; para el análisis y evaluación de los resultados de la valoración del juicio de experto Sistema Predictivo basado en el modelo de predicción de preferencias en usuarios en el sitio web Foros Perú, y se consideraron los siguientes criterios:

5.- Muy adecuado 4.- Bastante Adecuado 3.- Adecuado (Regular) 2.- Poco Adecuado

1.- No Adecuado

Pregunta	Experto N° 1	Experto N° 2	Experto N° 3
N° 1	5	5	5
N° 2	4	4	5
N° 3	4	4	5
N° 4	4	4	5
N° 5	4	4	5
N° 6	4	4	5
N° 7	4	4	5
N° 8	4	4	5
Puntaje Total	33	33	40

Tabla 10.- Calificación, Opinión, Evaluación de Expertos

1. En cuanto a la novedad científica del Modelo de Predicción de Preferencias de Usuarios, los cuatro expertos indicaron que es **MUY ADECUADO**.
2. Referente a la pertinencia de los fundamentos teóricos del Modelo de Predicción de Preferencias de Usuario, tres expertos expresaron que **ES BASTANTE ADECUADO Y UN EXPERTO MANIFESTÓ QUE ES MUY ADECUADA**.
3. El nivel de argumentación de las relaciones fundamentales aportadas en el modelo tres expertos expresaron que **ES BASTANTE ADECUADO** y un experto **MANIFESTÓ QUE ES MUY ADECUADA**.
4. En cuanto a la correspondencia entre el aporte teórico desarrollado y el aporte práctico en el modelo tres expertos expresaron que expresaron que **ES BASTANTE ADECUADO** y un experto **MANIFESTÓ QUE ES MUY ADECUADA**.
5. En el ítem referido a la claridad en la finalidad de cada una de las acciones del Sistema Predictivo basado en el modelo de predicción de preferencias de usuarios en el sitio web Foros Perú, tres expertos expresaron que **ES BASTANTE ADECUADO** y un experto **MANIFESTÓ QUE ES MUY ADECUADA**.
6. Tres expertos manifiestan que la posibilidad de aplicarse Sistema Predictivo basado en el modelo de predicción de preferencias de usuarios en el sitio web Foros Perú, **ES BASTANTE ADECUADO** y un experto **MANIFESTÓ QUE ES MUY ADECUADA**.
7. En el ítem referido a la concepción general del Sistema Predictivo basado en el modelo de predicción de preferencias de usuarios en el sitio web Foros Perú, tres expertos expresaron que **ES BASTANTE ADECUADO** y un experto **MANIFESTÓ QUE ES MUY ADECUADA**.
8. En cuanto a la significación práctica de la estrategia para la pertinente predicción de preferencias de usuarios tres expertos expresaron que **ES BASTANTE ADECUADO** y un experto **MANIFESTÓ QUE ES MUY ADECUADA**.

3.5.2 Ejemplificación de la aplicación del aporte práctico

La puesta en marcha del prototipo tiene como objetivo ejecutarlo, exportar a un manejador de predicción en formato CSV, como WEKA que realiza análisis estadístico inferencial, para lograr realizar la evaluación para fines de clasificación e interpretación de los datos procesados con el algoritmo Expectation Maximization EM.

3.5.3 Corroboración estadística de las transformaciones logradas

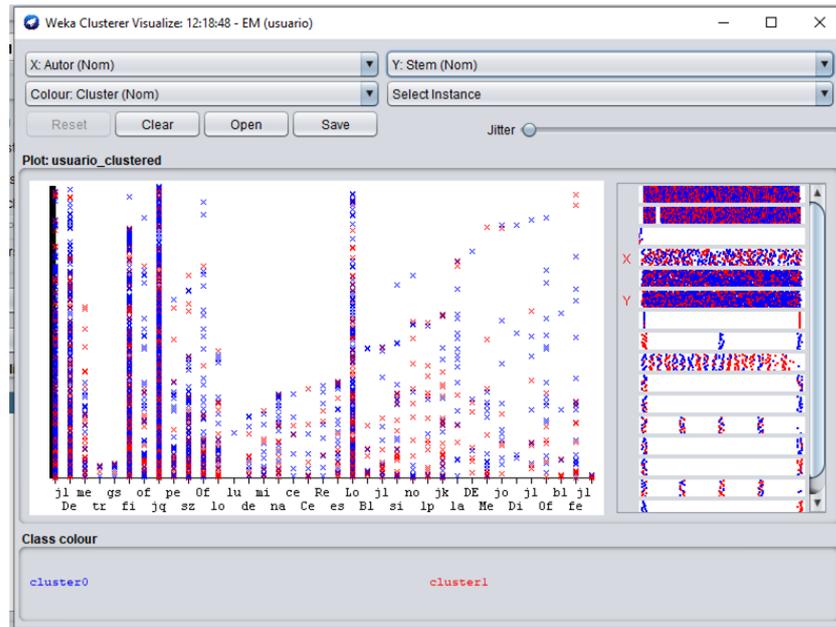


Figura 16.-Asignaciones de clúster por AUTOR y STEM

Nota.- Muestra las asignaciones de clúster en el proceso autor y stem (raíz de la palabra)

En la figura los colores son los se describen en la tabla 10, a continuación:

Agrupación	Color
0	Azul
1	Rojo

Tabla 11.-. Colores de las agrupaciones en la Fig. 16

Significado de las palabras procesadas que corresponden a los comentarios procesados

En la figura 16 se observa: una distribución casi pareja de ambos clusters (Agrupaciones), pero con predominancia de cluster 0 en algunos autores, mientras que en otros la predominancia es de clúster 1. Para evaluar el significado de esta distribución es preciso antes constatar qué significa que un autor emplee mayor cantidad de palabras de uno u otro tipo

En cluster 0: refleja palabras que tienen que ver **con lo que se desea hacer, y sus características**, es decir las palabras serían: (probando, rápido, universitario, experiencias, universidad, evaluación, métodos, odontología estomatología, obstetricia, enfermería, veterinaria, tecnología, nutrición, administración, sociología, exámenes, superarchimegaavanzada, imagínense, doctorado, avanzadas, finalmente, herramienta, orientación, desarrollarla, enseñando, investigando, trabajado, empresa, desperdician, tantos, magísteres, doctores, logrado, llevar, laboral, pasión, formas, futuro, definido, seguir, corriente, pregunta, gustara, perfectamente, macroeconómica, juegos, por supuesto, económico, internacional, pesado, ahorita, llevando, españolas, intercambio, tienen, mejores, primeras, practicas, malditos, europeos, claaaro, interesa, publica, oportunidades, afuera, curricula, llevamos, creeme, bravazooo, encanto, primer, estamos, viendo, bendita, demanda, actual, mercadeo, distintas, netamente, estudiar, electivos, desventajas, tengan, complementan, enfoques, distintos, dependerá, siguiendo, agentes, conjunto, perspectiva).

En cluster 1: refleja **estados de ánimo y apreciaciones, y sus características**, es decir las palabras serían: (estoy, igual, nada, mejor, tema, asumo, nos, gusta, bueno, pobre, ciclo, bien, vale, pena, tema. Ideas, creo, esta, crear, hacer, razón, útil, ósea, buen, libro, clara, paul, fácil, menos, tener, base, vez, ahí, quede, autor, pucp, asumo, todos, bueno, gusta, pobre, pero, bien, vale, pena, temas, ideas, intro, micro, macro, color, crear, menos).

Como consecuencia, se puede decir que los autores con prevalencia en cluster 0 están comunicando **necesidades o peticiones**. En cambio, los que presentan prevalencia en cluster 1, están comentando **apreciaciones personales o estados de ánimo**.

Agrupación	Color
0	Verde

1	Amarillo
---	----------

Tabla 12.-Colores de las agrupaciones a continuación en la Figura 17

Aquí se observa que:

- a) Hay una ruptura de la curva (ver círculo rojo a la izquierda) que correspondería a un **conjunto de IDComentarios (comentarios) que no tienen STEM importantes y por lo tanto no entraron en la curva.**
- b) También se observa un comportamiento anómalo hacia la derecha (el círculo) que indica que los STEM se reúsan respecto a comentarios anteriores.

De (a) se puede decir que hay comentarios que son superfluos y el *Sentiment Analysis* detecta que su contenido no es necesario o interesante para el estudio.

De (b) denota comentarios repetidos o que parafrasean con las mismas palabras.

En cuanto a la relación de los comentarios de los **IDComentario** (eje X) respecto a **STEM** (eje Y), se muestra en la figura 17.

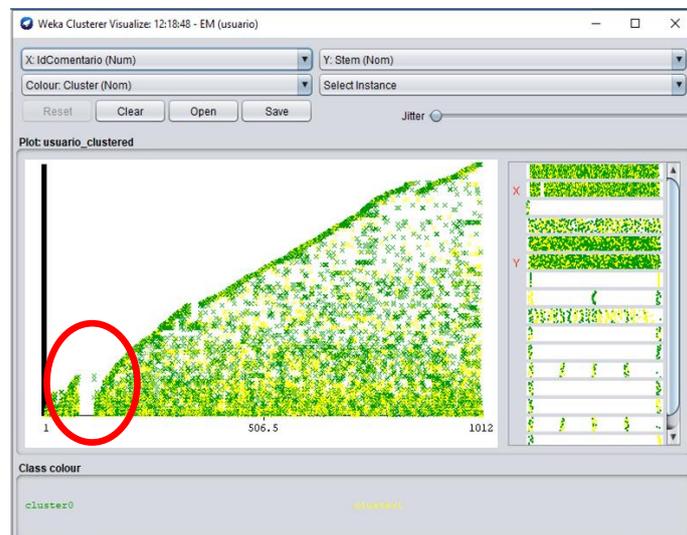


Figura 17.- Asignaciones de clúster por IDComentario y STEM

Nota.- Muestra las asignaciones de cluster por Idcomentario y stem (raíz de la palabra)

IV. CONCLUSIONES

- a) Se ha elaborado un sistema web basado en el modelo de preferencias de usuario en el sitio web Foros Perú; con la implementación del prototipo (ver figura 1).
- b) Esta tesis ha caracterizado epistemológicamente el proceso de lenguaje natural con la minería de opinión y preferencias, mediante la clasificación masiva de datos de manera automática con algoritmo implementado en weka, mostrando los fundamentos y las metodologías para poder realizarlo sobre un foro de comentarios denominado Foros Perú; dicha caracterización está dada por la aplicación en el prototipo implementado de las wavelets morfosintácticas, así como la aplicación del algoritmo Expectation Maximization y su log likelihood (**ver anexo 6 y 7**);
- c) Se ha diagnosticado el estado actual de la dinámica del proceso de lenguaje natural PLN con la minería de opinión y preferencias en el sitio web Foros Perú, basado en el prototipo que emplea técnicas establecidas de campo, como son el algoritmo heurístico Expectation Maximization (EM), y otras novedosas aún, como las Wavelets Morfosintácticas del procesamiento de lenguaje natural.
- d) Se ha obtenido un modelo de preferencias de usuario, para conocer que usuarios son más y menos activos en el sitio web Foros Perú; por lo que de la discusión de resultados se puede afirmar que el prototipo empleado:
 1. -Es capaz de extraer conclusiones de cada archivo de metadatos. (Ver tabla 6)
 2. -Extrae cruces de interés a los efectos de determinar los influencers, seguidores activos y seguidores menos activos
 3. -Los metadatos empleados son suficientes para un análisis automatizado inicial que indique tendencias de opinión dentro de un foro, de manera sistemática y sin sesgos temáticos.
- e) Se ha estructurado un sistema predictivo basado en el modelo de predicción de preferencias de usuario en el sitio web Foros Perú, implementando un prototipo que opera el algoritmo de porter para extraer (stem) las raíces de las palabras y aplicar la investigación (**Ver tabla 1 y tabla 4**), observando que usuarios son más activos.
- f) Se ha validado y corroborado los resultados alcanzados, (**ver tabla 6, anexos 7 y 8**).

V. RECOMENDACIONES

- a) Se recomienda la inclusión de los métodos prácticos de calificación o evaluación fruto de la experiencia cualitativa existente en el proceso predicción de preferencias de usuario.
- b) Se sugiere un recojo o acopio de información exhaustivo a fin de poder alinear el conocimiento histórico de los mensajes contenidos en una página web determinada que se desee analizar.
- c) En el proceso de diagnóstico se recomienda remitir previamente los instrumentos de recojo de la información con los formatos de salida pre establecidos a fin de realizar de la forma más precisa y optima el proceso de información sensible.
- d) El modelo teórico requiere ser integrado de manera continua con la práctica y la adecuación de los procesos referidos a fin de mantener su vigencia y pueda ser mejorado o empleado para integrar procesos futuros sensibles una vez evaluado su funcionamiento.
- e) El modelo práctico requiere de una integración con la tecnología actual a fin de poder generar confianza y permita la integración física de la información con las nuevas técnicas de análisis de la información utilizando inteligencia artificial, como sentimental analysis, minería de opinión, empleando el proceso de lenguaje natural.
- f) El aprendizaje automático del modelo empleando la minería de opinion permite muchas posibilidades dentro del sentimental analysis que logre una buena capacidad para gestionar gran cantidad de información para establecer patrones que apoyen a obtener un conjunto de características de los perfiles de usuarios de los comentarios del sitio web Foros Perú, por lo que se recomienda continuar con la experimentación con el algoritmo Expectation Maximization (EM) de WEKA a fin de poder incrementar aún más los márgenes de confianza para una mejor formación de opiniones usando las waveletes morfosintácticas técnica empleada como la forma moderna de ultima generación de la presente investigación.

VI. REFERENCIAS

- Aguila Obra, A. R., y Padilla Meléndez, A. (2017). *Los modelos de negocio basados en las redesinterorganizacionales*. Obtenido de Universidad de Malaga: https://www.google.com.pe/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwj9s5q3_svsqAhUwHbkGHUMADhcQFjABegQIBhAB&url=https%3A%2F%2Fdialognet.unirioja.es%2Fdescarga%2Farticulo%2F565037.pdf&usg=AOvVaw3WSQOXJcgVAgobCdbkJOeL
- Arango, M. L. (2019). Foros virtuales como estrategia de aprendizaje. *Debates Latinoamericanos*.
- Alfonso Cutro DATA PRIX. (2019). *Minería de Datos* . Obtenido de Evolución de la Minería de Datos: <https://www.dataprix.com/es/mineria-datos-aplicada-encuesta-permanente-hogares/171-evolucion-historia-mineria-datos>
- Agencia de Marketing Digital experto seo . (2017). *Estrategias de Fidelización de Clientes* . Obtenido de Como fidelizar clientes nuevos y antiguos con SEO de Internet: <https://expertoenseo.pe/estrategias-de-fidelizacion-de-clientes/>
- Blanco, E.-J., & Sanz, H. (2016). Algoritmos de clustering y aprendizaje automático aplicados a Twitter. *Universidad Politécnica de Catalunya*.
- Barrueta Meza & Castillo V. (2018) Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos. Universidad Peruana de Ciencias Aplicadas
- Bermudez Q. C. & Bermudez Q. S. (2019) Sistema de predicción de estadísticas de nacimientos y defunciones en Colombia soportadas
- Bordignon, F., & Panessi, W. (2001). Procesamiento de variantes morfológicas en búsquedas de textos en castellano. *Revista Interamericana de Bibliotecología*.
- Bordignon, F. R. A., & Panessi, W. (2001). Processing morphological in searches of Spanish texts | Procesamiento de variantes morfológicas en búsquedas de textos en castellano. *Revista Interamericana de Bibliotecología*.
- Brun, R. E., & Senso, J. A. (2004). Minería textual. (Spanish). *Text Mining. (English)*.
- C, P. L. (2017). *Minería de datos: Técnicas y Herramientas*. Madrid.: Internacional Thomson

- Castaño Gomez (2019) Modelo predictivo para inferir en el proximo presidente de estado a traves de un vocabulario ontologico en twitter
- Cortez Vásquez, A., Vega Huerta, H., y Pariona Quispe, J. (2018). *Procesamiento de lenguaje natural*.
- Cordova Baldeón I.(2015) El Proyecto de Investigación cuantitativa Editorial San Marcos.
- Cordova Baldeón I. (2016) El Informe de Investigación cuantitativa Editorial San Marcos.
- Débora Hisgen y López De Luise (2017). *Dialog Structure Automatic Modeling*.
- Daniela López De Luise, Débora Hisgen, A. C. and M. M. R.-Aig. U. de P. M. B. 1050. B. A. A. (2012). *Modeling dialogs with linguistic wavelets*.
- Débora Hisgen and Daniela López De Luise-AIGroup, Universidad de Palermo, M. B. 1050. (2017). *Dialog Structure Automatic Modeling*.
- El lenguaje no verbal: un proceso cognitivo superior indispensable para el ser humano. (2015). *Revista Comunicación*. <https://doi.org/10.18845/rc.v20i1>.
- Dolores del Castillo & Ignacio Serrano (2013) Modelo computacional cognitivo de toma de decisiones basado en el conocimiento: aplicación en la inferencia de explicaciones.
- Gonzalez Sandoval (2016) Lo electoral: entre la certidumbre de las Viejas practicas y el caos de las nuevas tecnologías, Minería de texto para el analisis de campañas online.
- Guerra Guachamin Michelli (2016) Diseño e implementación de una aplicación de Minería de datos que permita sugerir personas con intereses afines mediante el analisis de sus preferencias en medios sociales.
- Henriquez M. & Salcedo J. (2016) Procesamiento de lenguaje natural
- Hernández, M. B.,y Gómez, J. M. (2013). Aplicaciones de Procesamiento de Lenguaje Natural. *Revista Politécnica*.
- <https://www.forosperu.net/>. (2018). *No Title*. De Un Tiempo a Esta Parte, Foros Perú Se Ha Convertido En El 4Chan Peruano. No Solo Por Su Utilidad Como Foro, Su Manera de Compartir Información o Por Las Alucinantes Historias Que Encuentras Ahí.
- Ignacio Saporiti, J. A. (2017). *Minería de opiniones y visualización de datos aplicables a estudios de mercado*.

- INEI. (2017). Lima – Perú. *Perú. Anuario de Estadísticas Ambientales*.
- J.P., S. (2010). Minería de opinión para la voz del consumidor. *Instituto Tecnológico de Buenos Aires Universidad Privada – Tesis de grado de ingeniería industrial.*, 78.
- Lopez Sanchez , J. I., & Sandulli, F. D. (30 de 08 de 2019). Evolución de los modelos de negocios en internet: situacion actual en españa de la economia digital.
- López De Luise, Hisgen, Cabrera y Marcelo Morales (2012). *Modeling DIALOGS WITH LINGUISTIC WAVELETS*.
- Lozano, V. (2020). *Operación “koda” de la policía nacional Los niños que “jugaban” a difundir porno infantil en Instagram*. <https://www.elperiodico.com/es/sucesos-y-tribunales/20201229/menores-difusion-instagram-porno-infantil-policia-nacional-operacion-koda-11424968>
- Luis Rendeules. (2020). *La Policía advierte que los pedófilos buscan menores de edad en TikTok*. <https://www.elperiodico.com/es/sucesos-y-tribunales/20200414/pedofilos-buscan-menores-tik-tok-7926585>
- Lucas Montenegro Garcia (2014) Analisis de Sentimientos y Predicción de eventos en twitter – Universidad de Chile
- Marin L. Juan & Marcelo L. T. (2020) Analisis de datos para el marketing digital emprendedor: Caso de estudio del parque de innovación empresarial de manizales
- Mariñelaren, Errecalde y Castro (2017). Extracción de conocimiento con técnicas de minería de textos aplicadas a la psicología. *Revista Argentina de Ciencias del Comportamiento*, 13.
- Mineria de Datos y su aplicación en el marketing . (30 de 08 de 2017). Obtenido de Minería de Opinión y Preferencias : <https://blog.ida.cl/analitica-web/mineria-datos-marketing-digital/>
- MD Agencia Ernesto Uranga. (2019). *Evolución de Internet y el mercado hoy en dia* . Obtenido de Mercado empleando el internet : <https://www.mdmarketingdigital.com/blog/la-evolucion-del-internet-y-el-mercado/>
- Ma, H., Qian, W., Xia, F., Wei, J., Yu, C., & Zhou, A. (2013). On benchmarking online social media analytical queries. *1st International Workshop on Graph Data Management Experiences and Systems, GRADES 2013 - Co-Located with SIGMOD/PODS 2013*. <https://doi.org/10.1145/2484425.2484435>

- Madrid Periodico, D. El. (2020). *En redes sociales Iglesias y Montero denuncian amenazas a sus hijos y gritos ante su casa.*
<https://www.elperiodico.com/es/politica/20200824/iglesias-montero-denuncia-escrache-galapagar-vivienda-hijos-8086378>
- Oliva Valdebenito (2014) Minería de opinion y analisis de sentimientos
- Prieto, M. ´. (2017). Estudio y Aplicación de Nuevos M´etodos de comprensión de texto orientada a palabras. *Universidad de Valladolid - Departamento de Informatica - Tesis Doctoral .*
- Pollete, Castro Porras y Hernandez N.(2016) implementación de un modelo predictivo basado en datamining y soportado por SAP predictive analytics en retails.
- PROMETEUS . (2019). *Evolución y definición de Minería de Opinión .* Obtenido de Concepto del Analisis de Sentimiento ó Minería de Opinión :
<https://prometeusgs.com/tecnologia-al-servicio-del-sentiment-analysis/>
- SAS Software y Soluciones de Analitica. (2019). *Procesamiento de Lenguaje Natural .* Obtenido de Evolución del Procesamiento de Lenguaje Natural :
https://www.sas.com/es_pe/insights/analytics/what-is-natural-language-processing-nlp.htm
- Rodriguez C & Miñani O (2017) Desarrollo de una aplicación informatica basada en un modelo de machine learning para mejorar la evaluación de prestamos crediticios
- Saporiti Ignacio & Tibaldo J. A. (2014) Minería de opinions y visualización de datos aplicables a estudios de mercado
- Shahana, P. H., & Omman, B. (2015). Evaluation of features on sentimental analysis. *Procedia Computer Science.* <https://doi.org/10.1016/j.procs.2015.02.088>
- Stacey, Patricia; Hallo, F. (2016). El Marketing Digital. Un análisis del uso de herramientas digitales en el Ecuador. *Valor Agregado.*
- Santillan A. Nilton (2019) Minería de opiniones basado en aprendizaje supervisado en la evaluación de destinos turisticos de la región Puno.
- Segura Vasquez (2019) Evaluación de algoritmos de clasificación para el minado de opinión en twitter

Stocca Juan Pablo (2010) Minería de opinion para la voz del consumidor

Times, L. A. (2021). *Cuando un miembro de Qanon disfrazado de bisonte presidió el Senado de EEUU*.

Tul, Q., Ali, M., Riaz, A., Noureen, A., Kamranz, M., Hayat, B., & Rehman, A. (2017). Sentiment Analysis Using Deep Learning Techniques: A Review. *International Journal of Advanced Computer Science and Applications*.
<https://doi.org/10.14569/ijacsa.2017.080657>

Uribe Saavedra, A. F. (2015). Uso de las Redes Sociales Digitales como herramienta de Marketing. *Tesis universidad autonoma de barcelona*, 201.

Yoo, S. Y., Song, J. I., y Jeong, O. R. (2018). Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications*.
<https://doi.org/10.1016/j.eswa.2018.03.055>

Vadivukarassi, M., Puviarasan, N., y Aruna, P. (2017). Sentimental Analysis of Tweets Using Naive Bayes Algorithm. *World Applied Sciences Journal*.

Witten, I. H., Frank, E., y Hall, M. a. (2017). Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). In *Complementary literature None*.

Cordova Baldeón I.. (2013) El Proyecto de Investigación cuantitativa Editorial San Marcos ISBN 9786123029616.

Anexos

Anexo 01: Matriz de consistencia.

Anexo 02: Operacionalización de las variables.

Anexo 03: Instrumentos.

Anexo 04: Validación de instrumentos por juicio de expertos.

Anexo 05: Validación de los aportes de la investigación.

Anexo 06: Formato Epguss visado por asesor de tesis especialistas.

Anexo 07: Expectation Maximization (Em) – Usuario.Csv.

Anexo 08: Expectation Maximization (Em) – Blog.Csv.

Anexo 09: Constancia de Originalidad de Turnitin

Anexo 10: Declaración Jurada de consentimiento informado

ANEXO 1.- MATRIZ DE CONSISTENCIA

DEFINICIÓN DEL PROBLEMA	OBJETIVOS	TÍTULO	VARIABLES	HIPÓTESIS
<p>Problema General</p> <ul style="list-style-type: none"> ¿Cómo mejorar el procesamiento de lenguaje natural que limita la predicción de preferencias de usuarios en el sitio Web Foros Perú? <p>Problemas Específicos</p> <ul style="list-style-type: none"> ¿Como se caracteriza epistemológicamente el proceso de lenguaje natural y minería de opinión mediante el sentimental analysis? ¿De que manera se diagnostica el estado actual 	<p>OBJETIVO GENERAL</p> <ul style="list-style-type: none"> Evaluar el sistema predictivo basado en el modelo de preferencias para la predicción de preferencias con prototipo de usuarios en el sitio web Foros Perú. <p>OBJETIVOS ESPECÍFICOS</p> <ul style="list-style-type: none"> Comprobar las características epistemológicas del proceso de lenguaje natural y minería de opinión mediante el sentimental analysis. Determinar el diagnóstico del estado actual de la dinámica del proceso de lenguaje natural PLN con la minería de opinión y preferencias en el sitio web Foros Perú. 	<p>“Sistema predictivo basado en la predicción de preferencias de usuario en el sitio web foros Perú”</p>	<p>VARIABLE INDEPENDIENTE</p> <p>“Sistema Predictivo basado en el modelo web que mejora la predicción de preferencias de usuarios”</p>	<p>Hipótesis general</p> <p>H1: Un sistema predictivo basado en el modelo web mejora la predicción de preferencias de usuarios.</p> <p>Hipotesis específicas</p> <p>H2: El sentimental de analysis resalta, caracteriza epistemológicamente</p>

<p>de la dinámica del proceso de lenguaje natural PLN con</p> <p>la minería de opinión y preferencias en el sitio web Foros Perú?</p> <ul style="list-style-type: none"> • ¿Cómo elaborar un modelo de predicción de preferencias de usuario, para mejorar el intercambio de información en el sitio web Foros Perú? • ¿De que manera se propone estructurar un sistema predictivo basado en el modelo de predicción de preferencias de usuario, que verifique que usuarios son más activos en el sitio web Foros Perú? • ¿Cómo se validará y corroborará los resultados alcanzados? 	<ul style="list-style-type: none"> • Verificar el modelo de preferencias de usuario, para mejorar el intercambio de información en el sitio web Foros Perú • Verificar la estructura de un sistema web basado en el modelo de preferencias de usuario en el sitio web Foros Perú. • Demostrar la validación y corroboración de los resultados alcanzados en el estudio. 		<p>VARIABLE DEPENDIENTE:</p> <p>“Predicción de preferencias de usuario en el sitio web foros Perú”</p>	<p>nte en el proceso de lenguaje natural y minería de opinión</p> <p>H3: Si se diagnostica el estado actual de la dinámica del procesamiento de lenguaje natural NLP con la minería de opinión y preferencias en el sitio web Foros Perú, mediante etapas entonces la comprensión y precisión será adecuada para verificar el perfil</p>
---	--	--	---	--

				<p>de los usuarios que interactúan en foros Perú.</p> <p>H4: Si se elabora un modelo de predicción de preferencias de usuario, se mejora significativamente el intercambio de información en el sitio web Foros Perú.</p> <p>H5: Si se estructura un sistema basado en el modelo de predicción de preferencias de</p>
--	--	--	--	---

				<p>usuario, entonces se verificará que los usuarios son más activos en el sitio web Foros Perú.</p> <p>H6: Si se valida y corrobora los resultados alcanzados, empleando herramienta estadística inferencial entonces se demostrará que los usuarios son más activos y/o pueden tener más seguidores en los</p>
--	--	--	--	---

				<p>comentarios que escriben.</p> <p>H4: Si se elabora un modelo de predicción de preferencias de usuario, se mejora significativamente el intercambio de información en el sitio web Foros Perú.</p> <p>H5: Si se estructura un sistema basado en el modelo de predicción de preferencias de usuario, entonces</p>
--	--	--	--	--

				<p>se verificará que los usuarios son más activos en el sitio web Foros Perú.</p> <p>H6: Si se valida y corrobora los resultados alcanzados, empleando herramienta estadística inferencial entonces se demostrará que los usuarios son más activos y/o pueden tener más seguidores en los comentarios que</p>
--	--	--	--	---

				escriben.
--	--	--	--	-----------

OPERACIONALIZACIÓN DE VARIABLES

“Modelo web para la predicción de preferencias de usuario en el sitio web foros Perú”

VARIABLES	DIMENSIONES	INDICADORES	TÉCNICAS O INSTRUMENTOS.	FUENTES DE VERIFICACIÓN
V. DEPENDIENTE “Predicción de preferencias de usuario en el sitio web foros Perú”	Tiempo	Recolección de datos Procesamiento de datos Emisión de Reportes	Observación / guía de observación Análisis y revisión documentario	Expertos, software de captura y extracción de información
	Predicción (significado)	Margen de error en la recolección Margen de error en el procesamiento	Observación / guía de observación Análisis y revisión documentario	Usuarios que opinan y acceden a los foros en internet
	Comprensión (Polaridad)	Cantidad de textos	Observación / guía de observación Análisis y revisión documentario	Usuarios que opinan y acceden a los foros en internet

ANEXO 3.- INSTRUMENTO

INSTRUMENTO: Guía de Observación

Objetivos del Instrumento:

Objetivo General:

Establecer la polaridad y significado de los comentarios, para poder construir el sistema predictivo basado en el modelo web de acuerdo a lo analizado.

Objetivos Específicos:

- Importar datos de varias fuentes (comentarios), tales como archivos web, bases de datos y hojas de cálculo.
- Limpiar los datos mediante la eliminación de los valores (repetidos) atípicos y la combinación de las fuentes de datos.
- Desarrollar un modelo predictivo preciso basado en los datos agregados mediante estadísticas, herramientas de ajuste de curvas o aprendizaje automático.
- Integrar el modelo en un sistema de predicción de la carga en un entorno de producción.

Ítem a observar:

1. Recolección de datos con el método countsizer, probado con aplicativo creado por el investigador para analizar los textos de los comentarios,
2. Existencia de margen de error en la recolección, al separar los caracteres especiales, tildes, etc.
3. La cantidad de textos de los comentarios detalla más el significado de su mensaje e identifica la polaridad.
4. El procesamiento de datos para análisis de lenguaje natural, puede ser manual con hoja de cálculo ó puede ser automatizado con un aplicativo de visual Basic elaborado por el investigador para analizar los palabras y textos.
5. Elaboración del prototipo para representar el modelo predictivo web de preferencias de usuario en internet.

ANEXO 4.- INSTRUMENTO DE VALIDACION NO EXPERIMENTAL POR JUICIO DE EXPERTOS

Encuesta a Expertos.

Estimado Profesor (a):

Ha sido seleccionado en calidad de experto con el objetivo de valorar la pertinencia en la aplicación del modelo web de predicción de preferencias de usuario. Las interrogantes están en función de evaluar la pertinencia científico-metodológica de la aplicación de los aportes teórico y práctico.

Datos del experto: Sofia Irene Delgado Wong.

Años de experiencia docente. 6 años

Categoría Docente. Profesor Titular

Grado Científico. Doctor en Administración

1. Novedad científica del modelo web para la predicción de preferencias de usuario.

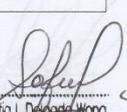
Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
x				

2. Pertinencia de los fundamentos teóricos del modelo web para la predicción de preferencias de usuario.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
	x			

3. Nivel de argumentación de las relaciones fundamentales aportadas en el modelo.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
	x			


 Dra. Sofia I. Delgado Wong
 CLAD 002726

4. Nivel de correspondencia entre el aporte teórico y el aporte práctico de la investigación.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
	x			

5. Claridad en la finalidad de cada una de las acciones de la estrategia didáctica propuesta.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)x	(3)	(2)	(1)

6 Posibilidades de aplicación de la estrategia didáctica propuesta.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
	x			

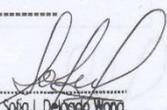
7 Concepción general de la estrategia según sus acciones didácticas desde la perspectiva de los actores del proceso formativo.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
	x			

8 Significación práctica de la estrategia para la pertinente predicción de preferencias de usuarios

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
	x			

Observaciones generales:


 CLAD
 Dra. Sofía I. Delgado Wong
 CLAD 002226

Encuesta a Expertos. *... entre el aporte teórico y el aporte práctico de la investigación.*

Estimado Profesor (a): *Bastante Adecuada* *Adecuada* *Poco Adecuada* *No Adecuada*

Ha sido seleccionado en calidad de experto con el objetivo de valorar la pertinencia en la aplicación del modelo web de predicción de preferencias de usuario. Las interrogantes están en función de evaluar la pertinencia científico-metodológica de la aplicación de los aportes teórico y práctico.

Datos del experto: CARLOS CAMPOS WILSON VEDURO

Años de experiencia docente. 19 años

Categoría Docente. Profesor Titular

Grado Científico. Doctor Ciencias de la Educación

1. Novedad científica del modelo web para la predicción de preferencias de usuario.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
x				

2. Pertinencia de los fundamentos teóricos del modelo web para la predicción de preferencias de usuario.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
	x			

3. Significación práctica de la estrategia para la pertinente predicción de preferencias de usuarios.

3 Nivel de argumentación de las relaciones fundamentales aportadas en el modelo.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
	x			

[Handwritten signature]

[Handwritten signature]

4. Nivel de correspondencia entre el aporte teórico y el aporte práctico de la investigación.

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
	x			

5. Claridad en la finalidad de cada una de las acciones de la estrategia didáctica propuesta.

Muy Adecuada (5)	Bastante Adecuada (4)x	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
---------------------	---------------------------	-----------------	----------------------	--------------------

6 Posibilidades de aplicación de la estrategia didáctica propuesta.

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
	x			

7 Concepción general de la estrategia según sus acciones didácticas desde la perspectiva de los actores del proceso formativo.

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
	x			

8 Significación práctica de la estrategia para la pertinente predicción de preferencias de usuarios

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
	x			

Observaciones generales:

ANEXOS N° 5 VALIDACIÓN DE LOS APORTES DE LA INVESTIGACIÓN

ESTIMADO DOCTOR:

Ha sido seleccionado en calidad de experto con el objetivo de valorar la pertinencia en la aplicación del aporte práctico.

DATOS DEL EXPERTO:

NOMBRE DEL EXPERTO	Karla Oblitas De Los Casas
PROFESION	Ing. ^o y Licenciada en educación
TITULO Y GRADO ACADEMICO	Doctora en Educación
ESPECIALIDAD	Ing. ^o Químico / Lic. en Edu. Mast.
INSTITUCION EN DONDE LABORA	U. C. V.
CARGO	D. T. C.

DATOS DE LA INVESTIGACIÓN:

TITULO DE LA INVESTIGACION	Modelo Predicción de Preferencias Usuario
LINEA DE INVESTIGACION	Inteligencia Artificial
NOMBRE DEL TESISISTA	Mgtr. Luis Santiago García Merino
APORTE PRÁCTICO	Sentimental Analysis - Estadística Inferencial Algoritmo Expectation Maximization.

Novedad científica del aporte teórico.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
X				

Pertinencia de los fundamentos teóricos del aporte teórico.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)
X				

Nivel de argumentación de las relaciones fundamentales aportadas en el desarrollo del aporte teórico.

Muy Adecuada	Bastante Adecuada	Adecuada	Poco Adecuada	No Adecuada
(5)	(4)	(3)	(2)	(1)

X				
---	--	--	--	--

Nivel de correspondencia entre el aporte teórico y el aporte práctico de la investigación.

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
X				

Claridad en la finalidad de cada una de las acciones del aporte práctico propuesto.

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
	X			

Posibilidades de aplicación del aporte práctico.

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
	X			

Concepción general del aporte práctico según sus acciones desde la perspectiva de los actores del proceso en el contexto.

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
	X			

Significación práctica del aporte.

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
	X			

Observaciones generales:

Muy Adecuada (5)	Bastante Adecuada (4)	Adecuada (3)	Poco Adecuada (2)	No Adecuada (1)
X				



**DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN Y SISTEMAS
TESIS EXPEDITA PARA VERIFICACIÓN DE SIMILITUDES,
ORIGINALIDAD Y AUTENTICIDAD DE LAS TESIS.**

El Asesor Especialista y Asesor Metodológico de la Tesis:

**"Sistema Predictivo basado en el modelo para la predicción de preferencias
en usuariosdel sitio web Foros Perú"**

Aprobada mediante Resolución N°: **096-2018/EPUSS-USS**

Presentada por:

GARCIA MERINO LUIS SANTIAGO

Apellidos

Nombre(s)

*Dictaminan que la Tesis está **EXPEDITA** para la verificación de similitudes,
originalidad y autenticidad de las tesis, implementada por la USS*

*Asesor
Especialista:*

Dra. MARIA DANIELA LOPEZ DE LUISE

*Asesor
Metodológico:*

Dr. CALLEJAS TORRES JUAN CARLOS

Firma y fecha

Firma y fecha

ANEXO 6

RESULTADO CON EL ALGORITMO
MAXIMIZACION ESPERANZA CON EL
ARCHIVO EXPORTADO
EM -USUARIO.CSV

ANEXO 6

Expectation Maximization (EM) – Usuario.csv

Muestra los valores resultants de aplicar el algoritmo EM

Buffer de Salida de EMusuario.csv

=== Run information ===

=== Run information ===

Scheme: weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv
1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100

Relation: usuario

Instances: 23468

Attributes: 14

Palabra

Stem

Frecuencia

TamPalabra

TermPalabra

Mayuscula

EsTitulo

TiPalAnt

Resaltada

Tipopagina(URL)

Tip.Palabra

Ignored:

IdComentario

Tema

Autor

Test mode: evaluate on training data

=== Clustering model (full training set) ===

EM

==

Number of clusters selected by cross validation: 2

Number of iterations performed: 1

Attribute	Cluster	
	0 (0.58)	1 (0.42)
=====		
Palabra		
hola	1	21
estoy	1	87
creando	10.9991	1.0009
este	1	93
espacio	17	1
funcione	9	1
muera	1	9
rapido	17	1
igual	1	25
nada	1	47
pierde	11	1
probando	11	1
estudiantes	29	1
economia	362.0001	1.9999
egresados	11	1
mejor	1	63
compartan	11	1
experiencias	10.9994	1.0006
cualquier	19	1
universidad	73.9996	2.0004
podamos	6.9996	1.0004
debatir	11	1
diferentes	11	1
cursos	121.9998	2.0002
temas	1	55
libros	40.9997	1.0003
profesores	43	1
nuestra	15	1
amada	1	7
carrera	69	1
porque	112	2
asumo	1	17
todos	1	57
estudiamos	9	1
nos	1	41
gusta	1	29

cierto	47	1
bueno	1	121
estudiante	11	1
pacifico	15	1
motivo	5	1
pobre	1.0001	4.9999
convalidacion	4.9997	1.0003
traslado	4.9999	1.0001
externo	7	1
cursando	4.9996	1.0004
aproximadamente	4.9975	1.0025
mezcla	5	1
ciclo	1	79
pero	1	383
bien	1	125
vale	1	13
pena	1	11
reaprendiendo	4.998	1.002
monton	5	1
tema	1	83
ideas	1	7
buscar	7	1
conflicto	9	1
criticar	5	1
universidades	32.9998	1.0002
espero	13	1
comentarios	10.9991	1.0009
saludos	17	1
futuros	9	1
economistas	47	1
tienes	36.9999	1.0001
razon	1.0001	24.9999
jldark	5	1
creo	1	121
comentamos	5	1
relacionados	14.9998	1.0002
esta	1	165
ciencia	14	2
evitamos	4.9999	1.0001
hablar	7	1
crear	1	7
hacer	1	59
util	1	11

oigan	1	9
nsoy	1	7
necesito	13	1
buen	1	25
libro	1	83
general	45	1
alguna	29	1
recomendacion	6.9998	1.0002
tampoco	21	1
estudio	43	1
recomiendo	17	1
paul	1	1
samuelson	11	1
introduccion	29	1
clara	1	7
precisa	7	1
interesante	51.9997	2.0003
principios	32	2
moderna	7	1
librito	7	1
color	1	7
crema	1	7
menos	1	91
pags	1	7
intereso	7	1
economista	45	1
camuflado	17	1
tim	1	7
algo	1	107
facil	1	35
conseguir	9	1
recontra	22	2
genial	9	1
leerlo	9	1
pdta	1	7
entiendo	15	1
osea	1	15
refieres	10	2
basico	20	2
divaga	6.9999	1.0001
demasiado	17	1
conceptos	11	1
tener	1	57

base	1	25
entenderlo	7	1
capte	1	7
primera	19	1
vez	1	51
situaciones	8.9997	1.0003
presentan	9	1
ahi	1	79
hagan	1	11
comunes	5	1
tendras	5	1
conocimiento	8.9996	1.0004
rico	1	5
ese	1	73
ademas	43	1
teoria	88	2
economica	81	1
seria	1	19
chequeada	5	1
ejemplo	24	2
monopolio	5	1
pensamiento	30.9997	1.0003
ricardo	15	1
solo	1	85
como	1	325
realmente	9	1
quede	1.0001	8.9999
quiere	21	1
transmitir	4.9999	1.0001
autor	1	15
tambien	178	2
pucp	1	31
parece	63	1
buena	1	49
iniciativa	5	1
maravillosa	4.9982	1.0018
alternativa	6.9991	1.0009
todavia	21	1
facultad	17	1
estudios	11	1
generales	12	2
podria	47	1
buscarlo	6.9999	1.0001

aqui	1	43
biblioteca	7	1
otro	1	55
lado	1	13
gustaria	15	1
saber	1	43
algunos	48	2
buenos	24.9994	1.0006
microeconomia	85.9995	2.0005
actualmente	12.9998	1.0002
curso	1	135
walter	7	1
nicholson	29	1
algun	1	25
pueda	1	13
ser	1	87
complementario	6.9989	1.0011
intro	1	19
micro	1	137
macro	1	95
krugman	19	1
parecen	9	1
puede	1	25
parkin	21	1
cuando	72	2
titulo	13	1
pense	1	17
iba	1	19
queja	1	7
volviendo	8.9999	1.0001
mencionaron	13	1
iniciarse	7	1
muy	1	83
explicado	9	1
fueron	9	1
empeze	7	1
algunas	17	1
partes	10.9999	1.0001
mankiw	28	2
use	1	13
aunque	76	2
sino	1	43
previo	7	1

recomendable	10.9989	1.0011
mmm	1	23
alguien	45	1
justamente	7	1
llamado	10.9999	1.0001
elementos	7	1
simplemente	22.9996	1.0004
perfecto	9	1
hablamos	5	1
profundizar	4.9994	1.0006
pasaremos	4.9998	1.0002
varian	15	1
chiang	9	1
depende	17	1
busque	4.9999	1.0001
baca	1	15
ampliaciones	8.9993	1.0007
vaina	1	13
medio	1	27
difícil	47	1
lucho	1	7
contra	9	1
slutsky	11	1
shepard	9	1
roy	1	9
marshall	9	1
hecho	1	21
sabes	1	7
entenderas	4.9999	1.0001
cuanto	47	1
recomendaria	6.9989	1.0011
mankew	5	1
robert	5	1
frank	1	5
ingenieros	7	1
ahora	1	51
les	1	57
gustado	4.9998	1.0002
todo	1	79
han	1	57
llevado	11	1
obviamente	25	1
hay	1	85

llaman	10.9999	1.0001
atencion	23	1
agarramos	3	1
odio	1	5
caso	1	31
contabilidad	14.9995	1.0005
toda	1	5
area	1	15
encima	5	1
recien	19	1
voy	1	27
usando	4.9999	1.0001
macroeconomia	61.9994	2.0006
jose	1	5
gregorio	5	1
recuerdo	17	1
llama	1	31
javier	4.9999	1.0001
silva	1	5
ruete	1	5
era	1	41
politica	28	2
paises	24.9998	1.0002
emergentes	9	1
anteriores	7	1
lei	1	9
dornbusch	5	1
fisher	4.9999	1.0001
ultimo	23	1
for	1	5
dummies	7	1
personalmente	8.9995	1.0005
desagrado	3	1
comienzo	7	1
parecia	5	1
subjetivo	3	1
curvas	3.9998	2.0002
indiferencia	2.998	1.002
ver	1	37
preferencias	6.9991	1.0009
efecto	2	2
sustitucion	2.9998	1.0002
renta	1	7

gustaba	2.9998	1.0002
graduado	3	1
hace	1	37
anos	1	23
estudie	3	1
conocida	3	1
estado	33	1
linda	1	3
experiencia	8.9996	1.0004
quizas	12.9998	1.0002
choque	3	1
poco	1	39
dictan	4.9998	1.0002
numeros	5	1
calculo	19	1
cuantitativa	5.9985	2.0015
incluso	19	1
donde	1	59
viene	1	9
elemental	3	1
nuevos	3	1
repasen	3	1
funciones	19	1
geometria	7	1
analitica	7	1
esto	1	59
siguiente	11	1
pues	1	31
monetaria	17	1
servir	2.9999	1.0001
suerte	13	1
disculpa	5	1
persona	9	1
ejerce	5	1
profesion	5	1
puedes	18.9999	1.0001
decir	1	59
econometria	49.9993	2.0007
quisiera	15	1
informacion	10.9999	1.0001
materia	5	1
tengo	1	57
entendido	13	1

vendria	13	1
necesitas	8	2
matematica	46	2
estadistica	47.9991	2.0009
mostraron	5	1
fundamentos	10.9993	1.0007
unico	1	13
habia	1.0001	10.9999
eran	1	11
letras	7	1
representan	10.9993	1.0007
etc	1	27
numero	7	1
jaja	1	61
caray	1	5
plato	1	5
fuerte	11	1
bastante	41	1
teorico	11	1
cuantitativo	4.9991	1.0009
nutre	1	5
antes	1	21
llegas	7	1
esas	1	29
alturas	5	1
estas	1	25
suficientemente	4.999	1.001
curtido	5	1
afrontarlo	5	1
pasan	1	11
naa	1	5
verdad	27	1
matematicas	26.9999	1.0001
encantan	5	1
quiza	1.0001	6.9999
lenguaje	7	1
filosofia	9	1
hacen	1	27
aburridones	4.9975	1.0025
pasarlos	3.9999	2.0001
inevitable	3	1
conta	1.0002	4.9998
ligada	6	2

espera	7	1	
evaluacion	13		1
privada	9		1
proyectos	21		1
finanzas	46		2
basan	1		5
ciencias	18		2
contables	5		1
conoce	9		1
grupo	1		13
reforzar	5		1
mis	1		15
conocimientos	10.9997	1.0003	
metodos	15		1
cuantitativos	6.9991	1.0009	
mientras	7		1
llevo	1		37
pregrado	4.9999	1.0001	
salir	1		19
gama	1		7
conformarme	4.999	1.001	
enseñan	7		1
digo	1		13
cuenta	31		1
favor	1		11
medicina	13		1
humana	3		1
odontologiaestomatologia	2.9973	1.0027	
enfermeria	3		1
obstetricia	2.9958	1.0042	
veterinaria	2.9958	1.0042	
tecnologia	14		2
medica	3		1
nutricion	5		1
examen	15		1
inevitables	2.9988	1.0012	
necesita	3		1
mundo	1		33
tiene	1		45
miedo	1		3
taaan	1		3
ligado	5		1
cierta	7		1

manera	19	1
financiera	14	2
sigue	1	11
administracion	41.9999	2.0001
tan	1	51
contable	3	1
nicos	1	3
niifs	1	3
mayor	1	29
analisis	27	1
estados	11	1
financieros	6.999	1.001
profunda	5	1
directa	3	1
aplican	3	1
complejas	3	1
sociologia	7	1
parecio	15	1
buenisimo	3	1
toco	1	3
expo	1	3
chicos	5	1
chicas	7	1
bailan	3	1
calles	3	1
sienten	3	1
proyecto	3	1
vania	1	3
masias	3	1
veces	1	13
daba	1	3
entrevistabas	2.998	1.002
lambda	23	1
escuchado	9	1
san	1	23
marcos	15	1
preparan	3	1
examenes	5	1
bcr	1	7
avanzadas	7	1
sunat	1	3
sbs	1	7
otros	1	27

desarrollo	50	2
opinen	3	1
imaginense	2.9999	1.0001
debe	1	27
superarchimegaavanzada	2.998	1.002
doctorado	5	1
verdaderos	3	1
machos	3	1
finalmente	14.9999	1.0001
lograr	5	1
herramienta	2.9991	1.0009
estan	1	35
dando	1	5
orientacion	2.9998	1.0002
empezando	3	1
descuidar	3	1
veo	1	13
instrumento	2.998	1.002
segunda	9	1
lleva	1.0002	14.9998
objetivo	11	1
central	11	1
importante	9	1
hacias	3	1
burla	1.0001	2.9999
punto	1	15
sirve	1	11
demostrar	7	1
teoremas	7	1
estar	1	13
afirmaciones	2.9988	1.0012
vida	1	17
real	1	7
desarrollarla	2.9999	1.0001
final	1.0001	20.9999
quedas	2.9999	1.0001
estancado	3	1
enseniando	5	1
investigando	2.9996	1.0004
haber	1	11
trabajado	7	1
empresa	21	1
solicita	3	1

malo	1	7
desperdician	2.9979	1.0021
tantos	10.9997	1.0003
magisteres	3	1
doctores	3	1
logrado	3	1
casi	1	39
nunca	1	11
llevar	23	1
campo	1	9
laboral	17	1
largo	1	7
plazo	1	11
veria	1	3
pasion	3	1
formas	9	1
futuro	5	1
definido	3	1
seguir	7	1
corriente	3	1
pongo	1	5
otra	1	33
pregunta	16	2
malla	1	3
curricular	3	1
claro	1	45
gustara	2.9999	1.0001
llevarlos	5	1
concuerto	3	1
perfectamente	2.9996	1.0004
contigo	3	1
monetari	3	1
macroeconomica	8.9995	1.0005
equilibrio	3	1
bienestar	5	1
juegos	6.9999	1.0001
porsupuesto	4.9986	1.0014
historia	23	1
economico	51	1
internacional	14.9998	1.0002
pata	1	3
dijo	1	33
pesado	6.9996	1.0004

ahorita	13	1
patas	1	13
llevando	12.9999	1.0001
dicen	1	21
espaniolas	3	1
intercambio	10.9993	1.0007
tienen	46	2
mejores	19.9998	2.0002
notas	1	5
primeras	9	1
practicar	13	1
malditos	3	1
europes	5	1
claaaro	3	1
interesa	17	1
publica	11	1
oportunidades	8.9991	1.0009
afuera	9	1
oye	1	5
curricula	5	1
llevamos	6	2
dos	1	19
creeme	3	1
bravazooo	3	1
encanto	3	1
vimos	1	3
primer	17	1
marx	1	5
estamos	12	2
viendo	7	1
salio	1	15
bendita	3	1
curva	1	19
demanda	20	2
actual	15	1
mercadeo	13	1
jajaja	15	1
pesar	1	11
pocas	1	9
cosas	1	39
distintas	9	1
netamente	8.9999	1.0001
salon	1.0001	8.9999

uds		1	15
estudiar	19		1
univ	1		3
electivos	11		1
cual	1		23
escoger	7		1
ventajas	3		1
desventajas	4.9986	1.0014	
ambos	1		9
tengan	6.9992	1.0008	
complementan	2.9979	1.0021	
enfoques	3		1
distintos	5		1
modo	1		7
dependera	5		1
estes	1		3
siguiendo	5		1
agentes	3		1
conjunto	7		1
perspectiva	6.9989	1.0011	
toma	1		7
cada	1		29
agente	3		1
individual	3		1
separado	7		1
dependiendo	6.9997	1.0003	
especialidad	4.9993	1.0007	
mira	1.0001	4.9999	
interesado	3		1
cuales	6.9998	1.0002	
pros	1		3
contras	3		1
tipo	1		23
sociales	13		1
prodia	3		1
despues	23		1
acabar	3		1
eres	1		9
conviene	13		1
dar	1		11
diferente	3		1
relevantes	5		1
interesantes	20.9998	1.0002	

desconocen	3	1
piensan	17	1
vas	1	19
sabiendo	2.9994	1.0006
mide	1	5
pbi	1	21
inflacion	13	1
estos	1	17
indicadores	8.9996	1.0004
significa	6	2
noticias	5	1
hablan	7	1
aprender	5	1
desempleo	9	1
mercado	43	1
sucede	9	1
impone	3	1
salario	5	1
minimo	29	1
estilo	3	1
sacarle	2.9998	1.0002
jugo	1	3
debes	1	5
verlo	1	5
llegar	7	1
regulacion	26	2
seas	1	5
hagas	1	3
maestria	19	1
complementario	2.998	1.002
cultura	3	1
elegiria	5	1
usualmente	15	1
empieza	3	1
entender	20	2
personas	11	1
empresas	22	2
consumidores	2.9996	1.0004
productores	4.9996	1.0004
interactuan	2.9979	1.0021
mercados	17	1
luego	1	35
pasa	1	23

gobierno	32	2
toman	1	3
medidas	5	1
economicas	7	1
produccion	22	2
empleo	9	1
ejemplos	5	1
simples	3	1
doy	1	3
convendria	3	1
ninguna	7	1
peor	1	7
vagamente	3	1
hablando	11	1
actuales	3	1
habla	1.0002	6.9998
nivel	1	25
interese	3	1
movimiento	3	1
podrias	5	1
elegir	15	1
empezar	11	1
social	21	1
siempre	27	1
sido	1	23
gustar	2.9999	1.0001
necesariamente	6.9991	1.0009
trome	1	3
defenderte	3	1
solamente	4.9982	1.0018
mate	1	103
dicho	1	27
momento	9	1
pueden	8.9999	1.0001
parecer	11	1
monotonos	3	1
aburrirte	3	1
curioso	13	1
terminan	9	1
trabajando	5	1
tanto	1	47
lleve	1.0001	10.9999
llevés	7	1

primero	15	1
ves	1	5
superficialmente	2.9991	1.0009
rendimientos	4.9985	1.0015
constantes	7	1
productos	5	1
marginales	2	2
costos	16.9999	1.0001
ingresos	7	1
quieres	7	1
explicacion	2.9998	1.0002
yeah	1	5
quitaste	2.9999	1.0001
palabras	7	1
lovin	1	3
wow	1	5
tenerla	2.9998	1.0002
recuerda	5	1
partir	12.9999	1.0001
modelos	10	2
macroeconomicos	2.9996	1.0004
macroeconometria	2.998	1.002
recomendables	2.9995	1.0005
herramientas	4.999	1.001
informaticas	2.9988	1.0012
paquetes	3	1
spss	1	3
eviews	3	1
minitab	3	1
esos	1	25
estudiaba	3	1
man	1	9
dire	1	5
respecto	21	1
refiere	7	1
pequenos	5	1
agregados	7	1
domestica	3	1
relacion	11	1
hogares	3	1
administrar	2.9997	1.0003
gasto	1	15
familias	3	1

considerar	3	1
precio	18	2
apartir	5	1
grandes	7	1
ingreso	22	2
nacional	7	1
publico	7	1
balanza	3	1
comercial	5	1
inversion	28	2
amplia	3	1
estudia	7	1
deberas	3	1
emplean	5	1
derivadas	16	2
costo	1.0001	24.9999
isocuantas	3	1
hayan	1	5
intersecciones	2.998	1.002
pendientes	3	1
ambas	1	13
relacionadas	4.9999	1.0001
veras	1	3
oferta	33	1
agregada	3	1
ahorro	3	1
ganas	1	7
nomas	1	15
campeon	3	1
desprenderse	2.9924	1.0076
definiciones	4.999	1.001
daria	1	7
definicion	3	1
popular	3	1
ejm	1	5
vista	1	9
vender	10.9999	1.0001
producto	7	1
desarrollar	3	1
competitivo	2.9991	1.0009
minimizar	3	1
existir	3	1
preferencia	2.9993	1.0007

aceptacion	3	1	
maximizar	3		1
clientes	3		1
permitira	3		1
determinar	3		1
rentable	3		1
resumen	7		1
microeconomica	4.9994		1.0006
marketing	5		1
ayuda	1.0002		6.9998
ensene	5		1
urgente	3		1
integras	7		1
deberias	3		1
lanzar	3		1
aca	1		15
tus	1		9
dudas	1		5
gustosos	3		1
contestaremos	2.998		1.002
parte	1.0001		38.9999
mision	3		1
hilo	1		3
semana	11		1
arranco	5		1
parciales	25		1
palta	1.0001		4.9999
toca	1		3
jueves	5		1
arrugue	3		1
fregado	2.9999		1.0001
ponga	1		3
supuestamente	8.9996		1.0004
tendra	5		1
usan	1		3
global	5		1
acuerdo	11		1
compro	3		1
marca	1.0001		2.9999
bambaren	3		1
compre	3		1
bamvarian	3		1
lukas	1		3

asuste		3	1
gente	1		21
sufre	1.0002		2.9998
meten	1		3
derivada	5		1
parcial	23		1
contrario	11		1
empila	3		1
arruga	3		1
toxica	3		1
hablo	1.0002		6.9998
uni	1		11
demostrando	2.9995		1.0005
hicks	1		3
sabe	1		11
diablos	3		1
control	15		1
optimo	15		1
escucho	3		1
ocurre	5		1
refiero	7		1
alfa	1		7
chvr	1		5
complica	3		1
relacionado	10.9998		1.0002
jalar	1.0001		2.9999
varios	21		1
lanzaa	3		1
todas	1		19
preguntaas	3		1
obtener	9		1
respuestas	3		1
personal	7		1
asunto	11		1
grave	1		3
situacion	15		1
explica	9		1
exactamente	2.9988		1.0012
tiempo	23		1
ayudar	9		1
alpha	1		3
digamos	8.9998		1.0002
practico	3		1

sydsaeter	2.9999	1.0001
ptm	1	7
tmb	1	7
empiezo	5	1
maso	1	5
graficas	5	1
equilibrio	9	1
lunes	1	9
maldicion	3	1
parametrica	8.9991	1.0009
somos	1	11
susti	1	3
sencillo	5	1
explicar	3	1
puntos	6.9998	1.0002
acabo	1	7
economistasantes	2.998	1.002
economsita	3	1
xdestoy	3	1
tercer	7	1
quisera	3	1
erudito	3	1
pro	1	5
responda	3	1
interviene	3	1
competencia	14.9995	1.0005
perfecta	7	1
atrazo	3	1
precios	16	2
impuestos	3	1
indirectos	3	1
concretamente	2.9988	1.0012
pone	1	9
tope	1	3
estableciendo	2.998	1.002
maximo	7	1
soles	1	19
siendo	9	1
sentido	3	1
enteinde	3	1
relevante	3	1
irrelevantes	2.998	1.002
curiosidad	3	1

econometriaes		2.9998	1.0002
medicion	3	1	
aplicacion	5	1	
problemas	17	1	
economicos	11	1	
empiezan	3	1	
guardar	3	1	
stocks	3	1	
vende	1.0003	2.9997	
negro	1	7	
superior	19	1	
tbn	1	3	
stoy	1	3	
revisa	3	1	
xvere	1	3	
broer	1	3	
tendre	3	1	
pd	1	13	
sospechaba	3	1	
estrelladefuego	2.9984	1.0016	
fija	1	3	
habra	1	7	
contraera	3	1	
genera	7	1	
exceso	7	1	
perdida	3	1	
eficiencia	5	1	
producira	3	1	
entrara	3	1	
altos	1	5	
cantidad	13	1	
ofreceran	3	1	
disminuira	3	1	
querran	3	1	
stock	1	3	
dibuja	3	1	
ponle	1	3	
sacas	1	5	
maniana	9	1	
lectura	3	1	
idea	1	9	
recomiendan	6.9991	1.0009	
conocer	3	1	

keynesiano	3	1
marxista	3	1
austriaca	3	1
neoclasica	3	1
escuelas	5	1
poderoso	5	1
retrospeccion	4.9999	1.0001
marc	1.0001	4.9999
blaug	1	5
complejito	5	1
logras	5	1
bravazo	11	1
david	1	5
hizo	1	9
llorar	9	1
impotencia	5	1
miembro	5	1
diamanteyomequedoencasael	2.998	1.002
csm	1	5
mantilla	52	2
tomo	1	7
pndjo	1	5
dado	1	13
http	1	65
www	1	31
uylibros	5	1
com	1	11
uyverlibro	5	1
asp	1	5
landreth	5	1
utilizan	9	1
pequenia	5	1
muestra	5	1
https	1	7
docs	1	5
google	9	1
comaue	5	1
edu	1	9
podras	5	1
abrir	1	5
ustedes	19	1
especializarse	6.9983	1.0017
rama	1	11

decision	5	1
tomar	1.0001	6.9999
visto	1	15
econometrica	4.9992	1.0008
facu	1	5
mayoria	27	1
uhm	1	5
decidir	5	1
opciones	9	1
comercio	11	1
cotera	5	1
llevas	5	1
profe	1	39
ahh	1	9
pensando	4.9998	1.0002
dirias	5	1
llevarlo	5	1
cortez	44	2
ensena	7	1
indeciso	5	1
ensenia	13	1
rompen	9	1
abstracto	5	1
matematico	6	2
econometrista	4.9992	1.0008
impresionante	4.9994	1.0006
microeconomista	4.9992	1.0008
pasado	14.9998	1.0002
quedo	1.0001	6.9999
bonifaz	9	1
habian	12.9999	1.0001
realidad	33	1
profesor	19	1
conosco	5	1
gracioso	11	1
jalo	1	7
ensenando	5	1
ojala	1	9
sigas	1	7
ensenaba	5	1
director	5	1
esan	1	39
quedara	4.9999	1.0001

corto	1	11
profes	12.9997	1.0003
araujo	5	1
flaca	1	5
ortiz	1	5
llevan	31.9999	2.0001
mates	1	15
dejo	1	5
cvs	1	3
tantees	2.9999	1.0001
quejado	4.9999	1.0001
llegaba	5	1
tarde	1	7
clases	11	1
laton	1.0001	4.9999
enseniar	11	1
franciskovic	4.9989	1.0011
jorge	1	9
cumpa	1	3
decano	5	1
ph	1	3
ecometria	3	1
universite	3	1
toulouse	2.9999	1.0001
sciences	3	1
francia	3	1
dicta	1.0002	4.9998
consultor	3	1
despacho	3	1
ministerial	2.998	1.002
ministerio	5	1
asesor	3	1
presidencia	8.9995	1.0005
ejecutiva	3	1
gerente	3	1
areas	1	9
operaciones	6.9992	1.0008
planeamiento	2.9984	1.0016
essalud	3	1
directorio	3	1
conasev	3	1
realizado	5	1
consultorias	2.998	1.002

banco	1	9
reserva	5	1
peru	1	83
indecopi	3	1
salud	1.0002	6.9998
profesional	6.9989	1.0011
catedratico	2.9988	1.0012
pacific	3	1
eduardo	3	1
gonzales	7	1
coteramaestria	2.9984	1.0016
investigacion	9.9997	2.0003
vinculados	3	1
servicios	7	1
publicos	7	1
politicas	21	1
publicas	17	1
analista	3	1
docente	13	1
aplicadas	3	1
iria	1	7
utiliza	3	1
sola	1	5
poder	1	7
aplicarla	3	1
papers	3	1
alla	1	11
mmmm	1	3
pucha	1	3
rankeado	3	1
imagino	7	1
buscara	3	1
reemplazo	3	1
pendejito	3	1
provecho	3	1
organizan	3	1
funda	1	3
repaso	3	1
algebra	9	1
full	1	9
convalidaron	2.9998	1.0002
limites	9	1
integrales	16	2

estatica		11		1
dinamica	9		1	
asesina	3		1	
VERDADERO	3		1	
filtro	3		1	
sabia	1.0001		10.9999	
estaban	9		1	
distribuidas	2.9988		1.0012	
llevaban	3		1	
separaron	3		1	
ultima	7		1	
pasaron	5		1	
ultimos	13		1	
apretado	3		1	
horas	1		3	
pusieron	3		1	
temas	5		1	
ecuaciones	16		2	
inecuaciones	2.9984		1.0016	
interes	11		1	
simple	5		1	
compuesto	3		1	
programacion	2.9999		1.0001	
lineal	7		1	
lineales	2		2	
relaciones	11		1	
matricial	3		1	
solucion	7		1	
diferenciales	9.9978		2.0022	
cambio	20		2	
promedio	13		1	
instantanea	2.998		1.002	
optimizacion	8.9999		1.0001	
variable	11		1	
aplicaciones	8.9996		1.0004	
elasticidades	4.9987		1.0013	
anti	1		3	
varias	13		1	
variables	9		1	
restricciones	4.9995		1.0005	
sistemas	13		1	
regresion	3		1	
sumas	1		3	

riemann	3	1
multiples	5	1
densidad	3	1
espacios	3	1
vectoriales	2.9991	1.0009
series	2.9998	1.0002
numericas	3	1
matriciales	2.9991	1.0009
orden	1	3
diferencias	2.9998	1.0002
dinamicos	3	1
multiecuacionales	4.9989	1.0011
diagramas	3	1
fase	1	5
operadores	3	1
rezago	3	1
diferencia	6	2
simultaneas	4.9989	1.0011
ensenian	5	1
misma	1	11
huevd	1	3
misimos	6.9994	1.0006
mismo	1	39
lindo	1	3
vemos	1	11
penultima	3	1
jodida	3	1
saben	1	17
mencionan	3	1
haciamos	2	2
matrices	3	1
ley	1	3
cramer	3	1
verano	11	1
escuela	9	1
peruana	13	1
utopica	3	1
ensenen	5	1
integrados	3	1
propuestas	7	1
completen	3	1
modifiquen	3	1
lista	1	3

carlos	3	1
bolona	3	1
obtenido	5	1
oxford	5	1
hernando	3	1
sotointro	3	1
yamadaeconometria	2.9984	1.0016
cortezteoria	2.9984	1.0016
monetario	3	1
henry	1	3
barclayfinanzas	2.9984	1.0016
kurt	1	3
burneopolitica	2.9984	1.0016
izquierdo	5	1
relativo	5	1
waldo	1	5
mendoza	5	1
dancourt	3	1
sigan	1	3
opinando	3	1
alucina	3	1
tio	1	3
hermano	7	1
mama	1	3
egreso	3	1
crisobal	3	1
huamanga	3	1
ayacucho	3	1
juega	1.0001	2.9999
futbol	5	1
pichangas	3	1
catolica	3	1
dijeron	17	1
felix	1	3
jimenez	9	1
udep	1	23
tenemos	12.9997	1.0003
has	1	5
mencionado	3	1
completo	5	1
mencionaste	2.9999	1.0001
continuidad	2.9997	1.0003
reales	4.9999	1.0001

vectorial	3	1
fernandez	7	1
aburrido	5	1
pondria	3	1
loyola	3	1
gustavo	3	1
yamada	3	1
nombre	9	1
primo	1.0003	2.9997
uniecucionales	2.9984	1.0016
supongo	23	1
haran	1	3
llevaremos	3	1
silabo	3	1
capitulo	9	1
integracion	8.9999	1.0001
duracion	3	1
horasobjetivos	2.998	1.002
especificos	2.9991	1.0009
definir	5	1
integral	9	1
presentar	3	1
interpretacion	6.9999	1.0001
geometrica	3	1
aplicar	3	1
diversos	5	1
muchachos	13	1
limpio	3	1
informativo	2.9988	1.0012
adherir	3	1
aportes	3	1
consultado	3	1
level	1	5
bravo	1	3
dominar	3	1
exagerando	5	1
opinan	5	1
greene	3	1
llego	1.0002	4.9998
referencias	2.9994	1.0006
deru	1	9
estudias	3	1
manya	1	3

serio	1	5
enseñar	3	1
tigres	3	1
inspira	3	1
cato	1	3
diseñas	3	1
fijo	1	13
pagina	5	1
eliges	3	1
confuso	3	1
sistema	3	1
vamos	1	3
cumplimos	3	1
ciertos	3	1
requisitos	7	1
estadísticas	2.9998	1.0002
aparte	7	1
eeggl	1	3
resto	1	15
carreras	37	1
cumplir	3	1
entrar	5	1
sufro	1.0001	2.9999
deberían	5	1
usual	1.0002	4.9998
enemigo	3	1
contador	3	1
americaeconomía	2.9941	1.0059
comnegocios	2.998	1.002
industriasperu	2.9611	1.0389
captara	2.9999	1.0001
inversiones	12.9998	1.0002
hoteleras	3	1
pienso	13	1
suene	1	3
raro	1	13
imposible	3	1
especialización	6.9999	1.0001
turismo	3	1
similar	13	1
sector	15	1
harta	1	3
plata	1	7

pais	1	55
paso	1	19
mente	1	3
informatica	2.9988	1.0012
tuviese	2.9998	1.0002
ciegas	3	1
podriiiiia	3	1
embargo	5	1
enfonca	3	1
dicha	1	3
ulima	1	5
enfoncan	3	1
dichas	2.9993	1.0007
fuese	1	3
amable	3	1
responder	3	1
pfff	1	3
pongan	4.9999	1.0001
xddd	1	3
banca	1	3
opta	1	3
privado	11	1
empresarial	6.9989	1.0011
jpg	1	3
user	1	3
fiicsdiego	13	1
obligatorios	6.9989	1.0011
finalidad	3	1
corrijanme	3	1
equivoco	3	1
especializa	2.9998	1.0002
especializen	2.9999	1.0001
describes	3	1
fuertes	7	1
resumirlo	3	1
dentro	12	2
propios	5	1
dice	1	11
subempleados	2.9991	1.0009
areasector	3	1
mayores	10.9998	1.0002
beneficios	5	1
remuneraciones	2.998	1.002

decirte	3	1
ingrese	3	1
par	1	7
queremos	3	1
dedicarnos	3	1
ganar	1	3
dejemoslo	3	1
ventajoso	3	1
eco	1	13
extranjero	5	1
trabajar	7	1
aplicas	3	1
decirlo	2.9987	1.0013
menciona	7	1
otras	1	13
psicologia	3	1
derecho	15	1
comunicaciones	2.9974	1.0026
mencionaba	3	1
alta	1	3
tasa	1	7
subempleo	3	1
comillas	3	1
gerentes	5	1
recursos	9	1
humanos	3	1
empresarios	2.998	1.002
politicos	3	1
consideran	13	1
vital	1	9
importancia	4.9997	1.0003
sociedad	9	1
ing	1	5
industrial	13	1
salieron	3	1
premios	9	1
nobel	1	31
noticia	7	1
nadie	1	17
menciono	7	1
elcomercio	17	1
premia	3	1
tres	1	3

expertos	3	1
leido	1	3
practicamente	16.9998	1.0002
premio	15	1
ingleses	3	1
estadounidenses	2.9984	1.0016
generalices	2.998	1.002
formula	7	1
xn	1	5
suma	1	7
consumo	10	2
cobierno	3	1
exportaciones	9.9974	2.0026
netas	1.0001	8.9999
total	1.0001	4.9999
tomados	2.9998	1.0002
ano	1	9
indicador	3	1
nominal	3	1
calcula	5	1
vigentes	3	1
pnb	1	3
producido	3	1
integrantes	4.9996	1.0004
independientemente	4.999	1.001
encuentren	3	1
fenomenos	3	1
afectan	3	1
miden	1	3
nobeles	2.9993	1.0007
departamento	2.9991	1.0009
trabaja	3	1
ganador	5	1
quienes	11	1
foro	1	3
asisten	3	1
conferencias	2.998	1.002
viernes	9	1
ecoomicos	3	1
tratados	3	1
reuniones	3	1
manana	11	1
ido	1	3

vino	1	3
directora	5	1
mundial	9	1
pasada	2.9999	1.0001
harvard	15	1
notificacion	2.9998	1.0002
decia	1.0001	4.9999
venian	3	1
subgerentes	2.998	1.002
bcrp	1	9
homenaje	3	1
noble	1	3
cargo	1.0001	2.9999
castillo	3	1
marco	1.0001	2.9999
vega	1	3
seran	1	7
respectivamente	2.9997	1.0003
quiero	13	1
tratare	3	1
proximo	17	1
antepasada	3	1
chequear	7	1
pag	1.0001	2.9999
web	1	3
quoteaste	3	1
c	1	7
funcion	56	2
disponible	5	1
equivoque	3	1
bruta	1	3
importaciones	4.9997	1.0003
reconocimiento	2.9991	1.0009
llega	1.0002	2.9998
institucion	2.9998	1.0002
universidadempresacentro	2.9984	1.0016
venido	6.9999	1.0001
realizando	3	1
obras	1	5
mvll	1	7
princeton	8	2
entonces	19	1
poner	1	9

gano	1		7
prestigio	3		1
saliendo	3	1	
unidos	7	1	
chicago	13	1	
alumnos	19	1	
investigadores	4.999	1.001	
actualidad	3	1	
stanford	11	1	
quien	1	11	
venida	2.9999	1.0001	
gringos	13	1	
siguen	6.9998	1.0002	
considerando	2.9998	1.0002	
obvio	1.0002	8.9998	
burlon	3	1	
podrian	5	1	
aulas	1	3	
formaron	3	1	
convirtieron	2.9999	1.0001	
exagerado	3	1	
genios	3	1	
universitaria	2.998	1.002	
juego	1.0002	2.9998	
realizar	5	1	
esfuerzo	3	1	
dotes	1	3	
grandeza	3	1	
pequenos	3	1	
encuentran	3	1	
juventud	3	1	
aduldez	3	1	
stand	1	5	
economist	3	1	
yoram	1	3	
bauman	3	1	
risa	1	5	
crees	1	3	
latinoamerica	9.9975	2.0025	
brasil	19	1	
ambito	3	1	
politico	7	1	
vive	1.0001	2.9999	

laborales		3	1
talla	1		3
centros	3		1
apoya	1.0001	8.9999	
nacionales	7		1
particulares	2.998	1.002	
sao	1		5
paulo	1		5
rio	1		3
janeiro	3		1
ciudades	9		1
negocios	15		1
america	5		1
europa	5		1
educacion	23		1
ricos	1		9
pobres	2.9999	1.0001	
publique	3		1
forosperu	2.9987	1.0013	
netshowthread	2.998	1.002	
php	1		5
trolls	3		1
podian	2.9999	1.0001	
opiniones	3		1
enfoque	11		1
cientifica	3		1
sur	1		3
ucsur	1		3
creado	2.9992	1.0008	
bienvenido	5		1
nova	1		3
familiar	5		1
mba	1		11
usp	1		3
gana	1		9
montonal	3		1
banearon	3		1
muelas	3		1
pequeno	5		1
mapa	1		3
distribuido	2.9988	1.0012	
prepare	3		1
sabado	4.9993	1.0007	

termino		5	1
dormir		7	1
roche	1.0001		2.9999
fresh		1	3
filo		1	3
esperando		3	1
nota		1	13
lengua		7	1
mare		1	3
puedo	1.0001		10.9999
creer		1	3
baje		1	3
estudiando		9	1
siento		7	1
gusto		1	5
sentirme		3	1
cuasi		1	5
cachimbo		3	1
nuevo	1.0001		8.9999
ansias		5	1
salido	8.9998		1.0002
asco		1	5
jalaron		5	1
confio		3	1
falta		1	9
unica		1	7
loqueo		3	1
dieron		5	1
utilidad	6.9999		1.0001
cosa		1	15
tenia		1	7
optimizar		3	1
restriccion		11	1
vainasa		3	1
hevdsa		3	1
positiva		3	1
valores	6.9999		1.0001
cumplan		3	1
llene		1	3
hoja		1	3
tire	1.0001		2.9999
minutos		5	1
deje		1	3

faltando	2.9999	1.0001	
min		1.0002	2.9998
acabe		1	5
revisando		3	1
nuevamente		5	1
dato		1	5
homogenea		3	1
grado		1.0001	4.9999
quedaban		2.9999	1.0001
intente		3	1
arreglar		3	1
podia		1.0001	2.9999
entere		5	1
problema		13	1
propuso		5	1
maldito		3	1
hdp		1	3
irme		1	3
bica		1	3
dedicare		3	1
chu		1	3
esperes		3	1
tantas		2.9998	1.0002
complicado		5	1
practica		15	1
justo		1.0001	2.9999
desalentadora		2.9984	1.0016
cumplira		3	1
aporte		3	1
aulafacil		5	1
commicroeconomialeccionestemario		2.998	1.002
htm		1	3
simplicidad		2.998	1.002
trata		1.0001	8.9999
matematicamente		2.9999	1.0001
interpretar		2.9992	1.0008
graficos		3	1
parecieron		5	1
dio		1	5
cambiarme		3	1
broma		1.0001	2.9999
toas		1	5
asociadas		3	1

intresa		3	1
muchisimo	3		1
siganle	2		2
aportando	3		1
jeje	1		3
creas	1		3
tranca	3		1
maticos	6.9999	1.0001	
distinto	7		1
analizar	3		1
entorno	3		1
interno	3		1
plantear	3		1
soluciones	3		1
supuestos	3		1
datos	1		3
profundidad	2.9991	1.0009	
administradores	6.9991	1.0009	
cursitos	3		1
muyyy	1		3
basicos	10		2
camuflados	3		1
dije	1		5
sean	1		9
iguales	4.9996	1.0004	
diria	1.0001	2.9999	
tranka	2		2
considerablemente	2.9998	1.0002	
aportan	3		1
marginal	16		2
beneficio	11		1
producir	5		1
unidad	9		1
adicional	7		1
cuesta	5		1
tranquilo	3		1
egresado	3		1
siente	5		1
tirria	3		1
regresando	3		1
sabian	8.9999	1.0001	
celulares	3		1
caros	1		3

celular		3	1
telecomunicaciones	2.998	1.002	
invertir	11	1	
haberme	2.9999	1.0001	
enterado	3	1	
anios	1	5	
cara	1	7	
sorprendio	5	1	
huevie	3	1	
ocupa	1	3	
incrementos	2.9988	1.0012	
rentabilidad	2.998	1.002	
aumenta	5	1	
incluye	3	1	
fea	1	3	
quemada	2.9999	1.0001	
meti	1	7	
post	1	3	
anterior	13	1	
hoy	1	11	
dia	1	11	
acabaron	3	1	
fin	1	9	
arduos	3	1	
dias	1	9	
paz	1	3	
revente	3	1	
preocupa	3	1	
sobrevivi	3	1	
estuvo	3	1	
anaitico	3	1	
procesos	7	1	
tecnologicos	4.9986	1.0014	
maaaso	3	1	
geografia	5	1	
stuvo	1	3	
xvr	1	3	
contando	3	1	
normal	23	1	
sacandome	2.9997	1.0003	
mugre	1	3	
mejorar	6.9999	1.0001	
ponderado	3	1	

kgon	1.0002	2.9998	
dios		1	7
fpp		1	3
existen		11	1
entiendes		3	1
commicroeconomialeccioneslecc		2.998	1.002
htmclaro		3	1
explican		3	1
sencilla		5	1
amiga		1.0001	4.9999
industrias		3	1
alimentarias		2.998	1.002
unalm		1	3
ensenare		3	1
desesperada		2.998	1.002
lleven		9	1
bab		1	3
pregunto		5	1
manyo		1	3
mujer		1	5
iniciales		3	1
frontera		3	1
posibilidades		5.995	2.005
tobi		1	3
toncs		1	3
varon		1.0002	2.9998
ayer		1	7
errores		3	1
rojas		1	3
paola		1	3
wikipedia		2	2
orgwikifreakonomicshhttp		2.998	1.002
orgwikisuperfreakonomicssel		2.998	1.002
namber		5	1
uan		1	5
worldwaid		3	1
revolucionaron		2.9999	1.0001
libre		1.0001	4.9999
referente		7	1
mit		1	11
importantes		4.9997	1.0003
crisis		10	2
lse		1	11

london		5	1
school	5		1
economics	5		1
detras	3		1
segundas	3		1
cambridge	3		1
maldita	3		1
dx	1		5
perdon	5		1
calculadora	2.9999	1.0001	
casio	1		3
misia	1		3
usar	1		7
editor	3		1
parecido	3		1
arriba	3		1
cociente	3		1
puro	1		3
denominador	4.9979	1.0021	
tratando	3		1
haces	1		11
factorizar	3		1
cubos	1		3
kede	1		3
diciendo	4.9995	1.0005	
ux	1		3
yap	1		7
hallas	3		1
queda	1.0001	10.9999	
reemplazas	15		1
expresion	5		1
reemplazando	2.9997	1.0003	
sta	1		3
factorizas	4		2
cubo	1		3
terminos	6		2
stan	1		3
vuele	1		3
uu	1		3
adentro	3		1
au	1		3
obtienes	3		1
quedaria	3		1

propiedades	2.998	1.002
basicas	3	1
pedirles	5	1
sepan	1	3
industria	6	2
jugos	1	7
sacan	1	5
insumos	3	1
circuito	3	1
quemo	1.0001	2.9999
buscando	3	1
encuentro	5	1
creciendo	5	1
porcentaje	5	1
duda	1	7
ciclos	4.9987	1.0013
acabaste	2.9999	1.0001
estrategia	12	2
ven	1.0001	10.9999
ejercicios	5	1
super	1	5
ayudan	7	1
quedes	2.9999	1.0001
ocurrio	5	1
separar	3	1
sale	1	13
crecimiento	25.9987	2.0013
venden	3	1
juguerias	3	1
envasados	5	1
iiiasi	2.9996	1.0004
decimo	3	1
internacionales	14.9998	1.0002
latina	3	1
considere	5	1
mencione	3	1
ambiental	7	1
naturales	7	1
ongs	1	3
entidades	3	1
ambiente	7	1
uso	1	5
distribucion	4.9999	1.0001

entidadesinstituciones	2.998	1.002
trabajo	13	1
impacto	3	1
referia	3	1
pulpin	3	1
tomas	1	3
electivo	3	1
decisiones	3	1
productivo	5	1
avanzada	5	1
elaboracion	2.9998	1.0002
demografia	3	1
actuarial	3	1
ahm	1	3
figura	3	1
macrodinamica	2.998	1.002
climax	3	1
seminario	3	1
gestionr	3	1
tesis	1	5
referidas	3	1
anda	1	3
planta	3	1
puedan	2.9999	1.0001
d	1	3
cambiando	7	1
gane	1	3
creen	1	7
continuar	7	1
cambiar	5	1
eviten	3	1
mencionar	3	1
nombres	9	1
candidatos	7	1
aclarar	3	1
francamente	2.998	1.002
ingeniero	3	1
dedicado	3	1
mujeres	3	1
ranking	5	1
igualdad	3	1
generos	5	1
crece	1.0002	4.9998

decrece	3	1
gestion	9	1
peimpresanoticiacae	2.998	1.002
dependientes	2.998	1.002
rara	1	3
relacionan	3	1
desigualdad	4.9986	1.0014
curiosamente	2.9988	1.0012
desiguales	3	1
genero	3	1
considero	3	1
lean	1	3
terra	1	3
penoticiasnoticia	2.998	1.002
aspx	1	3
bbm	1	3
panorama	5	1
elecciones	9	1
poniendose	2.9985	1.0015
casos	1	7
lourdes	21	1
susana	15	1
aliar	1.0006	2.9994
alan	1	13
izquierdistas	2.9995	1.0005
saldran	3	1
chongo	5	1
marchas	5	1
protestas	5	1
semanas	7	1
proceso	5	1
inicio	3	1
ciudad	3	1
demorara	3	1
gran	1	7
ventaja	3	1
distritos	5	1
ppc	1	5
jder	1	3
castanieda	5	1
fiscalizarlo	2.998	1.002
resurgir	3	1
partidos	4.9999	1.0001

izquierda		7	1
olvidados	3		1
podran	3		1
presencia	3		1
victoria	3		1
principio	13		1
manejar	7		1
eficiente	3		1
mal	1		11
grafica	3		1
comparativa	2.9988	1.0012	
factores	3		1
capital	5		1
veia	1		3
factor	3		1
usado	1		5
onpe	1		5
villaran	5		1
lima	1		19
forma	1.0002	2.9998	
malogre	3		1
avancemos	3		1
dices	1		3
aliarse	2.9998	1.0002	
serviria	3		1
ganara	2.9997	1.0003	
significar	3		1
alianza	9		1
centro	5		1
iiquierda	3		1
presidenciales	6.9991	1.0009	
dirigentes	3		1
posible	11		1
comun	1		3
aliaanza	3		1
terminara	3		1
aliandose	2.9998	1.0002	
toledo	18		2
tecnicos	3		1
funcionarios	6.9989	1.0011	
enfasis	3		1
redistribucion	6.9999	1.0001	
riqueza	5		1

principal	7	1
tarea	1	7
apra	1	11
salga	1	11
partidaria	3	1
apoyar	3	1
apoyara	3	1
fujimorismo	2.998	1.002
keiko	1	7
hubiera	5	1
alcaldesa	5	1
castaneda	5	1
presidentes	6.9995	1.0005
generar	5	1
expectativas	2.998	1.002
negativas	5	1
inversionistas	4.9998	1.0002
privados	7	1
extranjeros	4.9995	1.0005
afectaria	3	1
locomotora	3	1
suceder	3	1
corte	1	3
flujo	1	3
rato	1	5
quieran	3	1
esperar	5	1
contar	7	1
decian	6.9999	1.0001
pasaria	5	1
lula	1	5
atrae	1.0001	4.9999
alcaldia	5	1
temor	1	3
riesgo	3	1
significara	2.9991	1.0009
renacer	2.9999	1.0001
grupos	4.9998	1.0002
radicales	3	1
coqueteando	2.998	1.002
ayudas	3	1
taylor	3	1
queria	7	1

conversaciones	2.998	1.002
kouri	1	5
tacharon	3	1
repunto	3	1
hubo	1.0001	2.9999
confusion	3	1
apoyaba	3	1
quedar	7	1
votar	1	3
partido	2.9999	1.0001
facebook	5	1
perfiles	3	1
criticando	3	1
apoyando	3	1
volveria	3	1
loco	1	5
vuelta	3	1
fuesen	2.9999	1.0001
noches	7	1
tuve	1.0001	2.9999
pesadilla	3	1
republica	3	1
correcto	3	1
esperemos	3	1
existio	3	1
nuestros	5	1
gobernantes	6.9989	1.0011
preocupados	2.9988	1.0012
periodo	5	1
nuestras	3	1
puso	1	5
proposicion	2.9998	1.0002
traida	2.9998	1.0002
norteamerica	2.9984	1.0016
asia	1	3
habiendo	2.9999	1.0001
estabilidad	2.998	1.002
baja	1	3
atraccion	5	1
incluyentes	2.998	1.002
reduccion	3	1
pobreza	10	2
corrupcion	9	1

ten	1	7
serie	1	3
fases	1	3
camino	3	1
id	1	5
propia	5	1
llamese	2.9999	1.0001
privadas	3	1
estatales	3	1
pasar	1.0001	8.9999
previa	3	1
consiste	3	1
traigan	3	1
atraer	2.9998	1.0002
extranjera	7	1
haciendo	8.9987	1.0013
momentos	3	1
chile	1	13
crecieron	3	1
gracias	13	1
primeros	13	1
pataleos	3	1
interna	5	1
tentando	3	1
desarrollados	3	1
nuestro	5	1
desorganizacion	2.9998	1.0002
escandalo	3	1
municipales	2.998	1.002
candidata	3	1
debilidad	3	1
practicamos	2.9999	1.0001
continuismo	2.9997	1.0003
turno	1	3
dejarle	2.9999	1.0001
siguientes	5	1
empezo	5	1
busca	1.0003	2.9997
cuota	1	5
popularidad	2.998	1.002
consiguen	3	1
respuesta	7	1
probe	1.0001	2.9999

metodo	4	2
cambios	3	1
necesidad	5	1
fracciones	3	1
espinoza	5	1
resueltos	3	1
propuestos	3	1
usaba	1	5
maynard	3	1
kong	1	3
malearme	2.9995	1.0005
venero	2	2
usa	1	5
preferia	3	1
xd	1	5
contiene	3	1
amartya	3	1
sen	1	3
tomarlo	2.9998	1.0002
guia	1	3
leer	1	5
publicaciones	4.9998	1.0002
exposicion	5	1
xnel	1	3
componentes	2.998	1.002
compras	5	1
g	1	5
pones	1	13
parentesis	3	1
pbn	1	5
bruto	1	3
jale	1	9
cray	1	3
bajas	1	3
retire	3	1
oww	1	3
consumidor	15	1
productor	13	1
destruyo	3	1
perdido	3	1
mateco	3	1
comprendo	3	1
dolor	1	3

sufrir	2.9999	1.0001
acordaba	3	1
enseniaba	3	1
especificamente	2.9991	1.0009
sali	1	3
mediocre	5	1
descuides	3	1
introdutorios	2.998	1.002
vienen	4.9999	1.0001
yale	1	7
articulos	3	1
pertenecen	3	1
devuelto	3	1
exasperando	2.998	1.002
ejecutivo	3	1
traer	1	3
malas	1	3
proteccionistas	2.998	1.002
devolveria	3	1
construya	3	1
museo	1	11
pedian	3	1
puesto	3	1
demandaba	3	1
quedarse	2.9997	1.0003
puse	1	5
exigian	3	1
mostrarse	3	1
ostentoso	3	1
querian	3	1
museos	2.9998	1.0002
invertiria	3	1
duros	1	3
quieren	3	1
colocar	3	1
nacion	5	1
devueltas	3	1
cuestiones	3	1
eticas	2.9999	1.0001
etica	1.0001	2.9999
gringo	3	1
bush	1	3
aceptaba	3	1

condicion	3	1
devolvian	3	1
piezas	3	1
totalidad	3	1
pendavis	3	1
cabeza	5	1
lamentablemente	6.9989	1.0011
reflejarlo	3	1
estudiado	3	1
bloquee	3	1
planteandoles	2.9993	1.0007
problemillas	2.998	1.002
salen	1	3
entretenerse	2.9923	1.0077
resolviendolos	2.9993	1.0007
indefinida	3	1
definidas	3	1
impropias	3	1
regreso	5	1
anima	1	3
ayudarme	3	1
yg	1	5
dada	1	3
hallar	7	1
extremo	3	1
yf	1	3
rectatangente	2.998	1.002
recta	1	5
toque	1	5
integre	3	1
algunaya	3	1
avanzar	3	1
pormas	3	1
datos	1	17
utiles	2.9999	1.0001
tangente	3	1
extremos	3	1
coordenadas	2.998	1.002
usarlos	3	1
modelo	9	1
macroeconomico	2.9996	1.0004
respectivamenteconsumo	2.998	1.002
instante	3	1

cumple	3	1
cir	1	3
dcdt	1	3
positivas	5	1
deducir	3	1
ecuacion	2	2
diferencial	2.999	1.001
drdt	1	3
kar	1	3
bkab	1	3
resolver	7	1
correspondiente	2.998	1.002
expresada	3	1
constantesa	2.9975	1.0025
laca	1	3
multiplicas	2.9991	1.0009
dejas	1	3
fraccion	3	1
pasas	1	3
derecha	3	1
lados	1	5
supone	11	1
dr	1	3
integrar	3	1
logaritmo	3	1
agregas	3	1
constante	3	1
nom	1	3
divide	3	1
poblacion	5	1
per	1	9
capita	3	1
alli	1	9
estuve	7	1
leyendo	3	1
especifico	3	1
consecuencias	2.998	1.002
macroeconomicas	2.9996	1.0004
dolarizacion	4.9999	1.0001
inferencia	3	1
habria	3	1
reajuste	7	1
salarios	3	1

perderia		3	1
jerarquia	3		1
papel	1		3
billetes	3		1
moneda	11		1
generaria	3		1
millonarias	2.998		1.002
perdidas	5		1
eeuu	1		3
haria	1		3
maneras	3		1
subiria	3		1
bancos	2.9998		1.0002
estarian	3		1
locos	1		3
regularizar	2.9992		1.0008
cuentas	7		1
serian	2.9997		1.0003
estables	3		1
regirian	3		1
casas	1		3
cambistas	3		1
desaparecerian	2.9983		1.0017
pierden	3		1
empleos	3		1
demoraria	3		1
confianza	3		1
dolarizada	3		1
vendrian	3		1
doble	1		5
medida	3		1
variaba	3		1
sentirian	3		1
seguros	3		1
estable	3		1
concreto	3		1
ecuador	3		1
guiar	1		3
dolarizacionecuador	2.9914		1.0086
comindex	3		1
optioncom	2.999		1.001
indice	3		1
gini	1		5

menor	1	3	
dirian	3		1
desigual	3		1
enterados	3		1
coreec	3		1
congreso	5		1
regional	3		1
huancayo	3		1
evento	3		1
conee	1		3
realizo	3		1
piura	1		3
plantearan	3		1
frente	3		1
aquejan	3		1
posibles	3		1
alternativas	2.9993		1.0007
tendencia	5		1
mantenga	3		1
biggrin	5		1
demuestra	3		1
reducir	3		1
basta	1		9
dejado	2.9999		1.0001
gastos	2.9997		1.0003
hechos	2.9995		1.0005
calientita	3		1
admitidos	5		1
gob	1		5
pedocsproyecc	4.9986		1.0014
extension	7		1
seleccionados	4.9989		1.0011
pdf	1		5
ufff	1		3
alma	1		3
mater	1		3
recupero	5		1
aumento	9		1
propone	3		1
monto	1		7
blogs	1		3
semanaeconomica	2.9941		1.0059
comblogsapuntes	2.998		1.002

laboralespostshay	2.998	1.002
aumentar	5	1
aumentan	3	1
p	1	3
rould	1	3
alumno	3	1
disminuido	5	1
miremos	2.9999	1.0001
sueldo	17	1
deberia	8	2
productividad	8.9997	1.0003
debajo	3	1
aumente	3	1
ganan	1	7
aprobaron	3	1
piteado	3	1
tendran	5	1
genere	3	1
anio	1	15
comentando	5	1
sobremanera	2.9984	1.0016
postulando	3	1
jalando	2.9999	1.0001
mandan	5	1
atan	1	3
alto	1	5
trae	1	3
carpita	3	1
feria	1	3
contradictorio	2.998	1.002
implicancias	2.9988	1.0012
termina	3	1
desequilibrando	2.998	1.002
pudiendo	2.9999	1.0001
tasas	1	3
trabajador	3	1
peruano	7	1
ajustando	3	1
necesidades	2.9991	1.0009
eleva	3	1
perciben	3	1
coeficiente	2.9984	1.0016
limenos	3	1

haya	1	7
beneficiado	4.9997	1.0003
familia	5	1
seguiran	3	1
promociones	2.998	1.002
stalin	3	1
mariategui	3	1
mao	1	3
tse	1	3
tung	1	3
weber	1	3
ensenio	3	1
proposito	3	1
anunciado	3	1
incremento	3	1
comparacion	8.9999	1.0001
region	5	1
argentina	19	1
pesos	1	23
dolares	10	2
aprox	1	11
mil	1	3
colombia	9	1
subsidio	3	1
transporte	7	1
uruguay	3	1
uruguayos	3	1
mexico	11	1
bolivia	3	1
bolivianos	3	1
costa	1	3
rica	1	3
nicaragua	5	1
varia	1.001	2.999
colones	3	1
cordobas	2	2
oamerica	3	1
revientan	3	1
canasta	5	1
zonas	1	5
cuyos	1	3
monetarios	3	1
nulos	1	3

practican	3		1
precaria	3		1
subsistencia	2.998		1.002
lejanas	3		1
sierra	3		1
percapita	3		1
complejo	3		1
alegrarnos	3		1
saque	1		3
volvi	1		3
crecientes	3		1
clo	1		3
aviso	1		3
sorpresa	3		1
esperaba	3		1
misa	1		3
domingo	3		1
caballero	3		1
metido	2.9999		1.0001
responsabilidad	2.9984		1.0016
metieron	3		1
mandaron	3		1
chorrillos	3		1
fondo	1		7
acentamientos	2.998		1.002
traen	1		3
proximoc	3		1
meto	1		5
techo	1		3
organizar	3		1
liquido	3		1
peso	1		7
cansado	2.9999		1.0001
duro	1		3
emocionante	2.998		1.002
acaba	1		5
faltan	2.9999		1.0001
asimetrica	3		1
criterios	3		1
frego	1.0001		2.9999
tranqui	3		1
aplicada	3		1
matador	3		1

meter	1	3
prox	1	3
creditos	9	1
morir	1	3
es	1	7
paperblog	2.9999	1.0001
comla	1	3
hora	1	7
daran	1	3
acaso	1	5
basicamente	2.9999	1.0001
preservar	3	1
cifras	3	1
repercutan	3	1
restantes	3	1
seremos	2.9967	1.0033
rol	1	3
mano	1	5
quedado	2.9999	1.0001
demostrado	3	1
beneficia	3	1
crecimos	3	1
traduzca	3	1
mefor	1	3
infraestructura	6.9991	1.0009
mantener	2.9999	1.0001
solidez	3	1
soluciona	3	1
radica	3	1
claves	3	1
envio	1	3
pequena	3	1
resena	3	1
michael	6	2
porter	26	2
cade	1	11
guru	1	7
calienta	3	1
urubamba	3	1
cusco	1	3
reclama	3	1
impiden	3	1
prosperidad	8.9991	1.0009

desarrolloel	2.998	1.002	
segundo	5		1
presente	3		1
dirija	3		1
senaló	3		1
auditorio	3		1
centenares	3		1
ejecutivos	3		1
competitiva	2.9991	1.0009	
invitado	3		1
hemos	1		3
caído	1		3
optimismo	5		1
afirmo	3		1
combinan	3		1
sentimientos	2.998	1.002	
piensa	3		1
reformas	3		1
facilitar	3		1
sombras	3		1
preocupacion	2.9998	1.0002	
sostener	2.9999	1.0001	
socialmente	2.9998	1.0002	
ministro	5		1
carranza	9		1
criterio	3		1
echo	1		3
balde	1		3
agua	1		3
fria	1		3
titularon	3		1
sesion	3		1
tocaba	2.9998	1.0002	
participar	3		1
competir	3		1
automatico	5		1
repetio	3		1
pidiendo	3		1
amigo	1.0001	2.9999	
gore	1.0001	2.9999	
visito	3		1
contagio	3		1
entusiasmo	3		1

llegando	3	1
ruta	1	3
llegabamos	3	1
critica	3	1
logren	3	1
produzcan	3	1
ascenso	3	1
asistentes	3	1
remover	2.9999	1.0001
cotarro	3	1
seguramente	2.9991	1.0009
esperaban	3	1
frases	3	1
viviendo	5	1
heterogeneo	2.998	1.002
aprecia	3	1
distinta	3	1
segmentos	3	1
espaldas	3	1
reconocen	3	1
desarrolla	3	1
heterogeneidad	2.998	1.002
ay	1	3
registra	3	1
niveles	6.9997	1.0003
intolerables	2.998	1.002
ningun	2.9999	1.0001
llegado	3	1
rangos	3	1
exportador	3	1
crecer	2.9999	1.0001
debilmente	5	1
diversificado	2.998	1.002
ventas	3	1
primarios	3	1
comodities	3	1
vulnerabilidad	2.998	1.002
expuestas	3	1
sectores	8	2
productivos	2.9998	1.0002
insuficiente	2.998	1.002
indica	3	1
atractivo	3	1

existe	3	1
escasa	3	1
capacidad	7	1
innovacion	5	1
traba	1	3
registran	3	1
peores	2.9998	1.0002
revierte	3	1
robaran	3	1
aumentando	3	1
inseguridad	2.9988	1.0012
insegura	3	1
desarrollarse	2.9997	1.0003
rotundo	3	1
reconociendo	2.9993	1.0007
habido	2.9999	1.0001
principalmente	2.9993	1.0007
fiscal	3	1
funcionado	3	1
concluye	3	1
suficientes	2.9991	1.0009
especialistas	2.9995	1.0005
participacion	2.9998	1.0002
insistencia	2.998	1.002
acrecentar	3	1
burbuja	3	1
encerrado	3	1
elites	3	1
enfaticar	3	1
objetivos	3	1
alcanzado	3	1
metas	1	3
progresivamente	2.998	1.002
propuestaestrategia	2.998	1.002
sostiene	3	1
foco	1	3
prioridades	4.9986	1.0014
actuar	5	1
implica	3	1
agenda	5	1
preguntarse	2.9991	1.0009
parezca	3	1
acerque	3	1

consenso	3	1
prototipo	3	1
intervencion	8.9999	1.0001
apoyados	2.9999	1.0001
burlarse	2.9997	1.0003
cocineros	5	1
mineros	3	1
petroleros	3	1
riquezas	3	1
precisando	3	1
propuesta	3	1
enfaticado	3	1
dotaciones	3	1
tierra	3	1
deben	1	5
usarse	2.9999	1.0001
productivamente	2.9998	1.0002
convertirnos	2.998	1.002
meros	1	3
extractores	2.998	1.002
cuyo	1	3
exito	1.0001	2.9999
durar	1	3
tome	1	3
agotarlos	3	1
mantengan	3	1
asimismo	3	1
apunto	3	1
desarrollan	2.9999	1.0001
articulando	2.9995	1.0005
incipientes	2.998	1.002
mostro	3	1
cuadro	3	1
desarrollada	2.9999	1.0001
crean	1	3
tupidas	3	1
redes	1	3
aislados	3	1
articulados	2.9995	1.0005
complementacion	2.9998	1.0002
subrayo	3	1
cree	1	5
ponerse	2.9985	1.0015

prioridad	5	1
inmediata	3	1
seguridad	3	1
calificacion	2.9998	1.0002
obra	1	3
dificilmente	2.9999	1.0001
oponerse	2.9999	1.0001
orientaciones	2.998	1.002
presenta	3	1
serias	2.9998.	1.0002
debilidades	2.9993	1.0007
campos	2.9998	1.0002
cerrar	3	1
discurso	3	1
convoco	3	1
permiso	3	1
consejo	3	1
competitividad	2.9993	1.0007
raul	1	3
wienerenviado	2.9984	1.0016
especialhttp	2.9984	1.0016
diariolaprimeraperu	2.8899	1.1101
den	1	5
html	1	11
escucha	3	1
incluyendo	3	1
rumbo	1	3
tomaran	2.9997	1.0003
obsoleto	3	1
articulo	5	1
tendencioso	4.9989	1.0011
juzgar	3	1
fuentes	3	1
periodista	3	1
redactor	3	1
manda	1.0002	2.9998
indirectamente	2.9988	1.0012
chiquitas	3	1
parecidas	5	1
opinion	5	1
sesgado	3	1
mes	1	3
idiota	3	1

duele	1.	3
perfil	3	1
volveri	3	1
loc	1	7
vuelt	3	1
noch	7	1
pesadill	3	1
republ	3	1
correct	3	1
existi	3	1
govern	6.9989	1.0011
period	7	1
proposicion	2.9998	1.0002
tra	2.9999	7.0001
norteamer	2.9984	1.0016
asi	1	3
estabil	2.998	1.002
atraccion	5	1
incluyent	2.998	1.002
reduccion	3	1
pobrez	10	2
corrupcion	9	1
camin	3	1
ID	1	5
llames	2.9999	1.0001
estatal	3	1
cons	3	1
traig	3	1
chil	1	13
graci	13	1
patale	3	1
tent	3	1
desorganizacion	2.9998	1.0002
escandal	3	1
municipal	2.998	1.002
debil	8.9993	1.0007
turn	1	3
cuot	1	5
popular	2.998	1.002
consig	3	1
neces	6.9991	1.0009
fraccion	5	1

espinoz	5	1
resuelt	3	1
usab	1	5
maynard	3	1
kong	1	3
vener	2	2
usa	1	5
preferi	3	1
XD	1	3
contien	3	1
amarty	3	1
sen	1	3
gui	1	5
leer	1	5
exposicion	5	1
xnel	1	3
component	2.998	1.002
G	1	3
parentesis	3	1
pbn	1	5
cray	1	3
retir	3	1
oww	1	3
destru	3	1
matec	3	1
comprend	3	1
dolor	1	3
acord	5	1
mediocr	5	1
introductori	2.998	1.002
yal	1	7
articul	10.9989	1.0011
pertenec	3	1
devuelt	5	1
exasper	2.998	1.002
proteccion	2.998	1.002
devolveri	3	1
constru	3	1
muse	2.9998	11.0002
pedi	3	1
puest	3	1
exigi	3	1
ostent	3	1

invertiri	3	1
dur	1	7
coloc	7	1
nacion	5	1
cuestion	3	1
etic	3	3
bush	1	3
accept	3	1
condicion	3	1
devolvi	3	1
piez	3	1
pendavis	3	1
cabez	5	1
lament	6.9989	1.0011
reflej	3	1
bloque	3	1
problemill	2.998	1.002
entreten	2.9923	1.0077
resolv	8.9993	1.0007
indefin	3	1
impropi	3	1
anim	1	3
yg	1	5
g	1	3
extrem	5	1
yf	1	3
rectatangent	2.998	1.002
rect	1	5
toqu	1	5
algunay	3	1
porm	3	1
tangent	3	1
coorden	2.998	1.002
usarl	3	1
respectivamenteconsum	2.998	1.002
instant	3	1
cir	1	3
dcdt	1	3
deduc	3	1
drdt	1	3
kar	1	3
bkab	1	3
correspondient	2.998	1.002

expres	5	1
constantes	2.9975	1.0025
lac	1	3
supon	11	1
dR	1	3
logaritm	3	1
divid	9	1
poblacion	5	1
capit	3	1
alli	1	9
leyend	5	1
consecuent	2.998	1.002
dolarizacion	4.9999	1.0001
inferent	3	1
habri	3	1
reajust	7	1
perderi	3	1
jerarqui	3	1
papel	1	3
billet	3	1
moned	11	1
generari	3	1
millonari	2.998	1.002
eeuu	1	3
hari	1	3
subiri	3	1
estari	7	1
regulariz	2.9992	1.0008
establ	5	1
regiri	3	1
cambist	3	1
desapareceri	2.9983	1.0017
demorari	3	1
confianz	3	1
dolariz	3	1
dobl	1	5
sentiri	3	1
segur	6.9991	1.0009
ecuador	3	1
dolarizacionecu	2.9914	1.0086
comindex	3	1
optioncom	2.999	1.001
gini	1	5

menor	1	3
coreec	3	1
congres	5	1
regional	3	1
huancay	3	1
event	3	1
cone	1	3
piur	1	3
frent	3	1
aquej	3	1
tendenci	12.9989	1.0011
manteng	5	1
biggrin	5	1
demuestr	3	1
reduc	3	1
bast	1	9
calientit	3	1
admit	5	1
gob	1	5
pedocsproyecc	4.9986	1.0014
extension	7	1
seleccion	4.9989	1.0011
pdf	1	5
ufff	1	3
alma	1	3
recuper	6.9967	1.0033
propon	3	1
mont	1	7
blogs	1	3
semanaeconom	2.9941	1.0059
comblogsapunt	2.998	1.002
laboralespostshay	2.998	1.002
p	1	3
rould	1	3
disminu	5	1
sued	17	1
debaj	3	1
aprob	3	1
pit	3	1
sobremaner	2.9984	1.0016
postul	5	1
mand	9.0001	2.9999
atan	1	3

alto	1	5
carpit	3	1
feri	1	3
contradictori	2.998	1.002
implic	4.9988	1.0012
desequilibr	2.998	1.002
pod	2.9999	3.0001
ajust	3	1
elev	7	1
percib	3	1
coeficient	2.9984	1.0016
limen	3	1
seguir	3	1
promocion	2.998	1.002
stalin	3	1
mariategui	3	1
mao	1	3
tse	1	3
tung	1	3
proposit	3	1
anunci	5	1
comparacion	8.9999	1.0001
region	5	1
argentin	25	1
dolar	10	2
aprox	1	11
mil	1	3
colombi	9	1
subsidi	3	1
transport	7	1
uruguay	5	1
mexic	11	1
bolivi	3	1
bolivian	3	1
nicaragu	5	1
colon	3	1
cordob	2	2
oamer	3	1
revient	3	1
canast	5	1
zon	1	5
cuy	1	5
nul	1	3

precari	3	1
subsistent	2.998	1.002
lejan	3	1
sierr	3	1
percapit	3	1
alegr	3	1
saqu	1	3
volvi	1	3
crecient	3	1
clo	1	3
avis	1	3
sorpres	3	1
doming	3	1
caballer	3	1
respons	2.9984	1.0016
chorrill	3	1
fond	1	7
acent	2.998	1.002
proximoc	3	1
tech	1	3
liqu	3	1
cans	2.9999	3.0001
emocion	2.998	1.002
asimetr	3	1
criteri	5	1
tranqui	3	1
matador	3	1
prox	1	3
credit	9	1
mor	1	3
es	1	5
paperblog	2.9999	1.0001
coml	1	3
acas	1	5
preserv	3	1
cifr	3	1
repercut	3	1
restant	3	1
rol	1	3
traduzc	3	1
mefor	1	3
infraestructur	6.9991	1.0009
manten	2.9999	1.0001

solidez	3	1
radic	3	1
clav	3	1
envi	1	3
resen	3	1
michael	6	2
port	26	2
guru	1	7
calient	3	1
urubamb	3	1
cusc	1	3
reclam	3	1
impid	3	1
prosper	8.9991	1.0009
desarrolloel	2.998	1.002
dirij	3	1
senal	3	1
auditori	3	1
centenar	3	1
invit	5	1
hem	1	3
caid	1	3
combin	3	1
sentimient	2.998	1.002
reform	3	1
facilit	3	1
sombr	3	1
preocupacion	2.9998	1.0002
sosten	2.9999	1.0001
ministr	5	1
carranz	9	1
echo	1	3
bald	1	3
agu	1	3
fri	1	3
sesion	3	1
particip	3	1
compet	3	1
automat	5	1
repiti	3	1
pid	3	1
gor	1.0001	2.9999
visit	3	1

contagi	3	1
entusiasm	3	1
rut	1	3
llegab	3	1
produzc	3	1
ascens	3	1
asistent	3	1
remov	2.9999	1.0001
cotarr	3	1
fras	3	1
heterogene	2.998	1.002
apreci	3	1
segment	3	1
espald	7	1
heterogen	2.998	1.002
Ay	1	3
registr	5	1
intoler	2.998	1.002
rang	3	1
diversific	2.998	1.002
vent	3	1
primari	3	1
comoditi	3	1
vulner	2.998	1.002
expuest	3	1
insuficient	2.998	1.002
atract	3	1
escas	3	1
capac	11	1
innovacion	5	1
trab	1	3
reviert	3	1
rob	3	1
insegur	4.9988	1.0012
Es	1	3
rotund	3	1
fiscal	3	1
conclu	3	1
participacion	2.9998	1.0002
insistent	2.998	1.002
acrecent	3	1
burbuj	3	1
encerr	3	1

elit	3	1
enfatic	5	1
alcanz	3	1
progres	2.998	1.002
propuestaestrategi	2.998	1.002
sostien	3	1
foc	1	3
prioridad	4.9986	1.0014
actu	5	1
agend	5	1
parezc	3	1
acerqu	3	1
consens	3	1
prototip	3	1
intervencion	8.9999	1.0001
cociner	5	1
miner	3	1
petroler	3	1
dotacion	3	1
tierr	3	1
usars	2.9999	1.0001
convert	2.998	1.002
mer	1	3
extractor	2.998	1.002
exit	3	3
agot	3	1
asim	3	1
apunt	3	1
incipient	2.998	1.002
cuadr	3	1
tup	3	1
red	1	3
aisl	3	1
complementacion	2.9998	1.0002
subray	3	1
priorid	5	1
inmediat	3	1
calificacion	2.9998	1.0002
obra	1	3
opon	2.9999	1.0001
orient	2.998	1.002
cerr	8.0002	5.9998
discurs	3	1

convoc	3	1
permis	3	1
consej	3	1
raul	1	3
wienerenvi	2.9984	1.0016
especialhttp	2.9984	1.0016
diariolaprimeraperu	2.8899	1.1101
den	1	5
html	1	11
rumb	1	3
obsolet	3	1
juzg	3	1
fuent	3	1
redactor	3	1
chiquit	3	1
sesg	3	1
mes	1	5
idiot	3	1
duel	1	3
flique	3	1
material	3	1
cocin	3	1
hue	1	3
nevermind	3	1
encontr	3	1
variacion	5	1
link	1	11
sap	3	1
ibm	1	3
comsoftwaredat	2.998	1.002
alizacionesgast	2.998	1.002
iberoamerican	2.9984	1.0016
pecongres	3	1
economicaalgui	2.9982	1.0018
entrad	3	1
aconsej	6.9991	1.0009
cimient	5	1
parking	3	1
pindyck	3	1
brod	2.9999	1.0001
acerc	3	1
tembl	3	1
cerrefilosofi	2.998	1.002

cerreeconomi	2.998	1.002
asimil	3	1
relaj	3	1
sangr	3	1
cerradofil	2.998	1.002
hist	1	3
ire	1	3
toy	1	3
inviert	2	2
cit	3	1
parraf	3	1
latinoamerican	2.998	1.002
impuls	3	1
recuperacion	2.9998	1.0002
diner	7	1
cerc	1	3
bon	1	3
corpor	2.998	1.002
gubernamental	2.998	1.002
activ	3	1
prefer	5	1
banquer	3	1
agenci	3	1
reuters	3	1
cint	1	3
metric	3	1
alrededor	5	1
abdom	3	1
porcentual	3	1
experiment	2.998	1.002
ipc	1	3
seccion	3	1
acog	3	1
thre	3	1
cmg	1	7
acuerdat	5	1
cvme	1	5
dispuest	5	1
ofrec	9	1
mensaj	3	1
anad	3	1
fb	1	3
publicacion	4.9999	1.0001

febrer	5	1
asesinatoenelmarg	8.9977	1.0023
blogspot	9	1
spref	1	3
complicacion	4.9999	1.0001
lagrangean	4.9989	1.0011
langrangean	4.9989	1.0011
despej	5	1
rK	1	5
awl	1	5
cult	5	1
intelectual	4.9986	1.0014
pertenecient	4.9989	1.0011
cotiz	5	1
distingu	4.9986	1.0014
gerenci	5	1
cerebr	5	1
innov	6.999	1.001
anticip	4.9945	1.0055
asegur	5	1
vocabulari	4.9986	1.0014
noticier	5	1
bols	5	1
sabr	7	1
crack	1	5
aspect	9	1
alent	5	1
justific	4.9986	1.0014
lagrang	9	1
multiplic	8.9991	1.0009
lamb	1	5
busquenl	3	1
macroeconomiasector	2.9971	1.0029
duplic	5	1
blog	1	5
leo	1	5
academ	11	1
sudamer	5	1
haganl	5	1
barat	5	1
xD	1	3
way	1	5
cidac	1	5

joven		5	1
computacion		6.9999	1.0001
mensual		9	1
carreral	5		1
portal	5		1
secretari	5		1
prevision	5		1
stps	1		5
eno	1		5
inegi	1		5
ingenieri	16		2
aeronaut	4.9989	1.0011	
naval			5
pilot	5		1
aviador	5		1
extract	5		1
metalurg	4.9989	1.0011	
energet	5		1
bioquim	5		1
topograf	4.9989	1.0011	
hidrograf	4.9989	1.0011	
geolog	5		1
geodest	5		1
terapi	5		1
optometri	5.		1
electr	7		1
electron	5.9961	2.0039	
civil	1		5
construccion	4.9999	1.0001	
quimic	4		2
aliment	5		1
cnnextension	4.9996	1.0004	
exigent	9		1
conozc	7		1
bvl	1		5
plan	1		11
lic	1		5
record	5		1
liebr	5		1
tortug	5		1
lugar	5		1
ojo	1		5
intros	5		1

intermedi	9	1
finaliz	4.9986	1.0014
spreffb	3.	1
xdest	3	1
contaduri	3	1
arquitectur	2.9988	1.0012
mecan	3	1
comunicacion	2.9999	1.0001
emres	3	1
cualqu	3	1
unmsm	1	3
upc	1	3
enfoc	3	1
line	1	3
peg	3	1
postgr	3	1
suponi	3	1
iban	1	3
rector	3	1
junt	3	1
escrib	3	1
exced	3	1
marz	1	5
biqu	3	1
similar	3	1
vuelv	3	1
repet	3	1
acces	3	1
jdi	1	3
taaant	3	1
sumatori	3	1
aritmet	3	1
usas	1	3
mcgrawhill	3	1
azul	1	3
gradient	3	1
hessian	3	1
Ln	1	3
xe	1	3
gxx	1	7
gyy	1	7
gxygyx	3	1
des	3	1

hessi	3	1
l	1	3
gyx	1	5
ll	1	3
gxy	1	5
culmin	3	1
sint	2.998	1.002
vacacion	3	1
abriendom	3	1
enyuc	3	1
mechit	3	1
araoz	1	3
candidatur	10.9994	1.0006
caviar	3	1
ultra	1	3
fich	1	3
probabl	3	1
personaj	3	1
omg	1	3
reaccion	3	1
lug	1	5
aparec	3	1
metropolis	3	1
glob	1	3
metr	1	3
monitor	3	1
gmm	1	3
elabor	3	1
brookings	3	1
institution	2.9999	1.0001
londr	3	1
sigl	3	1
examin	3	1
desempen	3	1
comp	3	1
aquel	1	3
etap	1	3
recesion	3	1
mundiall	3	1
ocurrir	3	1
eluniversal	2.9941	1.0059
uams	1	3
sed	1	3

cid	1	3
itam	1	5
itesm	1	5
operacion	3	1
adi	1	5
yeahhh	3	1
charl	3	1
mech	1	3
beltr	3	1
urrunag	3	1
respald	2.9979	1.0021
abiert	4.9986	1.0014
daniel	3	1
cordov	3	1
siu	1	3
aprist	3	1
esten	1	3
desacuerd	2	2
broo	1	3
leend	3	1
chevr	3	1
esqu	1	3
didact	3	1
books	1	3
gratis	3	1
unam	1	5
lej	1	5
phd	1	5
andres	7	1
torcuat	5	1
tell	1	7
chau	1	3
envergadur	2.9984	1.0016
polariz	3	1
cuerp	3	1
discrep	2.998	1.002
ador	3	1
teoriasestudi	2.998	1.002
resalt	2.998	1.002
belgran	3	1
uba	1	11
optari	3	1
culombi	3	1

uniand		3		1
rosari		3		1
mejorcit		3		1
conveni	7		1	
estadi	3		1	
averigu	2.998		1.002	
cem	1		5	
afin	3		1	
politiz	3		1	
huelg	3		1	
utdt	1		3	
udes	1		3	
atras	1		3	
air	1		5	
nib	1		3	
pim	1		3	
discrimin	2.9978		1.0022	
segreg	5		1	
chamb	3		1	
bogot	5		1	
bsas	1		3	
onda	1		3	
sec	1		3	
hablador	3		1	
pil	1		3	
bonit	3		1	
cuerpon	3		1	
laur	1		3	
boz	1		3	
enan	3		1	
poller	3		1	
dient	3		1	
huev	1		3	
columbi	3		1	
oblig	3		1	
papersarticul	2.998		1.002	
revist	3		1	
academicoslaboral	2.998		1.002	
obligacion	3		1	
publicaspriv	2.998		1.002	
chever	3		1	
pasanti	3		1	
dirig	3		1	

extrani	3	1
irse	1	3
tomari	3	1
aparent	3	1
encuest	3	1
[total]	16017.2768	12082.7232
Frecuencia		
mean	1	1.0212
std. dev.	0.0001	0.144
TamPalabra		
Corta	1.0232	9662.9768
Media	11593.9213	90.0787
Larga	2109.3323	16.6677
[total]	13704.2768	9769.7232
TermPalabra		
ra	328.9906	183.0094
or	138.986	143.014
an	438.9806	333.0194
er	174.9988	353.0012
ir	108.9993	85.0007
re	113.0007	34.9993
on	732.9951	49.0049
ro	10579.5241	7531.4759
en	425.9945	270.0055
ar	354.9985	97.0015
sa	70.9955	61.0045
so	143.9958	208.0042
m	2.999	91.001
se	58.9687	157.0313
un	2.9999	27.0001
in	29.0002	12.9998
si	2.9996	43.0004
ri	3	5
ru	6.8497	97.1503
ur	1	5
al	2	1
[total]	13722.2768	9787.7232
Mayuscula		
SI	2484.9226	1439.0774
NO	11218.3542	8329.6458
[total]	13703.2768	9768.7232
EsTitulo		

NO	13311.2767	9766.7233
SI	392.0001	1.9999
[total]	13703.2768	9768.7232
TipPalAnt		
sustantivo	1404.9774	1006.0226
verbo	787.9583	568.0417
adjetivo	204.9862	229.0138
otros	11305.3548	7967.6452
adverbio	3	1
[total]	13706.2768	9771.7232
Resaltada		
NO	13577.2794	9730.7206
SI	125.9974	38.0026
[total]	13703.2768	9768.7232
TipoPagina (URL)		
Contenido	13631.4174	9618.5826
Indice	71.8594	150.1406
[total]	13703.2768	9768.7232
Tip.Palabra		
sustantivo	1291.9923	1120.0077
verbo	367.9978	988.0022
adjetivo	209.9823	224.0177
otros	11833.3044	7438.6956
adverbio	3	1
[total]	13706.2768	9771.7232

Time taken to build model (full training data) : 44.76 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 13702 (58%)
1 9766 (42%)

Log likelihood: -11.67969

ANEXO 7

RESULTADO CON EL ALGORITMO
MAXIMIZACION ESPERANZA DEL
ARCHIVO EXPORTADO

EM – BLOG.CSV

ANEXO 7
Expectation Maximization (EM) – blog.csv

=== Run information ===

**Scheme: weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv
1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100**

Relation: blog_DLL3

Instances: 936

Attributes: 7

Autor

TamComenConEspacio

TamComeSinEspacio

CantPalConRep

CantPalSinRep

Ignored:

idComentario

Comentario

Test mode: evaluate on training data

=== Clustering model (full training set) ===

EM

==

Number of clusters selected by cross validation: 10

Number of iterations performed: 7

Atributos	0	1	2	3	4	5	6	7	8	9
Autor	0.15	019	0.04	0.03	0.12	0.24	0.07	0.03	0.04	0.07
Jldark89	10.3315	21.4243	2.042	5.5897	13.5498	40.722	8.1048	2.4674	2.7558	24.012
Deru	12.4814	11.4384	8.3437	3	11.9609	16.2483	15.8791	3.1648	4.8573	9.6263
metalerojc	2.0377	3.9985	1.0179	1	1.0015	4.9623	1.9821	1	1	1
Tribolon	1	3.0488	1	1.9512	1	1	1	1	1	1
gsuz	1	3.0032	2	2.996	1.0178	1.0149	1	1.8329	1	1.1352
Fiescdiego	15.3373	32.4918	2.0003	7.1659	4.0065	27.2492	2.414	11.2724	9.9946	11.0682
Ofiblion	2.0096	1.5414	1.0013	3.4606	6.0907	4.7072	1.9992	2.0023	1.1661	2.0217
Jq13	77.4991	93.1864	14.2126	13.2457	61.697	116.5591	15.5719	7.8367	10.8352	21.3562
Pet924	1.4664	3.0007	1.0002	1	1.9611	2.895	2.5379	1.0752	2.9885	1.075
szf	5.8187	3.9315	3.0132	2	6.4869	3.2072	1.9965	1.9153	2.0234	2.6074
Ofiblion	2.0258	1.0012	1.0037	1	3.8955	2.6813	2.9706	2.1119	1.0361	2.2739
lovin	3.045	2.0405	1	1.9605	3.9849	1.9537	2.0013	1.0137	1	1.0003
Luchex2710	1.9953	1	1	1	1	1.0047	1	1	1	1
deru	1	1.9622	3	1	1.0378	1	1	1	1	1
mickly	1.8147	1	1	1	1.5381	1.1853	1	1.0018	1	1.4601
nauttico	1	2	1	1	1.2295	2.9954	1	1.7558	1	1.0193
cera	1.9823	1	1	1	1	1	1.0177	1	1	1
Cera	1.0154	1	1	1	1	1	1.9846	1	1	1
Redo	1	1.9189	2	1	1	1	1	1.0811	1	1
estrelladefuego	1.0192	1.9743	3.0094	1.0259	1.8677	1.9948	2.9714	1.1315	1	1.0059
Lovin	15.4884	5.2408	8.0113	1.0286	7.2213	7.588	10.7047	4.3841	2.2775	2.0548
Blake	1	1	1	1	1.0253	1.0005	1	1.0074	1	1.9669
Jldrak89	1.0001	1.9985	1.0055	1	1.0015	1	1.9945	1	1	1
sir.kane	1.0126	2	2.9995	1	1	1.9878	1.0005	1.006	1.9905	1.0031
Nova00	1.1863	1	1	1	1	1.0579	1.8137	1.0157	1.9263	1
lp1	1	1	1	1	1.0225	2	1	1.0145	1	1.963
jklm92	1.9955	1	1	1	1	3.0138	1.0047	1.0184	1.9654	1.0022
la unica	1	2.905	1	1	1.095	1	1	1	1	1
Deru	1.0086	1	1	1	1	1	1.9914	1	1	1
Metalerojc	1	1.0002	1	1	1.9983	2	1	1.0002	1	1.0013
Jorge	1	1.9987	1	1.0013	1	1	1	1	1	1
DiegoD	1	1	3	1	1	1	1	1	1	1
jldark	1	1	1	1	1	1.9591	1	1.0141	1	1.0268
Ofiblion	1	1.0001	1	1	1.5601	1.9263	1	1.43452	1	1.0784
blake	1	1	1.006	1	1	1	1.994	1	1	1
fernando123	1	1.0008	1	1	2.9321	1.0032	1	1.0083	1.969	1.0065
jldark90	2.9983	1	1.0017	1	1	1	1	1	1	1
Total	178.4903 178.4903	219.1856	78.6664	69.4271	153.1819	264.9168	99.0158	66.5652	69.7856	106.7651

TamComenConEspacio

mean	102.5398	193.6982	42.2743	230.0522	164.3224	133.3426	74.5366	149.4691	139.8305	147.4368
std. dev	10.8832	13.0788	15.772	9.9907	8.1029	14.1755	8.914	6.6767	4.7014	4.7781

TamComesinEspacio

mean	86.1485	161.46	33.4936	191.9742	137.3509	111.2617	61.7757	124.2151	115.7678	121.8557
std. dev	9.503	11.4006	13.6491	8.8413	7.8553	11.9349	7.8497	6.471	3.694	4.0641

CantPalConRep

mean	8.8674	17.8301	3.7121	21.6312	15.397	11.8833	7.1913	13.4642	13.9911	14.7472
std. dev	1.3287	1.8332	1.5432	1.8167	1.3052	1.083	1.2098	0.6214	0.0968	1.3828

CantPalSinRep

mean	8.6711	16.7725	3.7716	19.8752	14.7661	11.4651	7.0028	13.2449	13.7947	14.1611
std. dev	1.3776	2.4011	2.0216	1.8985	1.4077	1.233	1.1635	0.6014	0.4041	1.5422

Time taken to build model (full training data) : 737.57 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 146 (16%)
1 170 (18%)
2 39 (4%)
3 39 (4%)
4 115 (12%)
5 222 (24%)
6 65 (7%)
7 34 (4%)
8 38 (4%)
9 68 (7%)

Log likelihood: -14.53056

ANEXO 8

Acta de Originalidad de la Tesis Doctoral EPGUSS

ACTA DE ORIGINALIDAD DE LA TESIS DOCTORAL



ACTA DE ORIGINALIDAD DE INFORME DE INVESTIGACIÓN

Yo, JUAN CARLOS CALLEJAS TORRES, docente de la Escuela de Posgrado - USS y revisor de la investigación aprobada mediante Resolución N°:96-2018/EPUSS-USS de la estudiante, Mg. **GARCIA MERINO LUIS SANTIAGO**, titulada "Sistema Predictivo basado en el modelo de preferencias en usuarios del sitio web Foros Perú"

Puedo constar que la misma tiene un índice de similitud del 13%, verificable en el reporte final del análisis de originalidad mediante el software de similitud.

Por lo que concluyo que cada una de las coincidencias detectadas no constituyen plagio y cumple con lo establecido en la Directiva de Similitud aprobada mediante Resolución de Directorio N° 176-2019/ PD-USS de la Universidad Señor de Sipán.

Pimentel, 23.12 del 2020

Código de registro consecutivo: AO-2020-HT-000-104

Dr. Juan Carlos Callejas Torres

Grado. Apellidos y Nombres del revisor de las coincidencias de similitud

CE 001170407

REPORTE DEL PROGRAMA ANTIPLAGIO TURNITIN AUTORIZADO EN EPGUSS

Sistema Predictivo basado en el modelo de preferencias en usuarios del sitio web Foros Perú

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	repositorio.uss.edu.pe <small>Fuente de Internet</small>	5%
2	Submitted to Universidad Cesar Vallejo <small>Trabajo del estudiante</small>	1%
3	blog.ida.cl <small>Fuente de Internet</small>	1%
4	Submitted to tec <small>Trabajo del estudiante</small>	1%
5	www.unlu.edu.ar <small>Fuente de Internet</small>	1%
6	docplayer.es <small>Fuente de Internet</small>	1%
7	Submitted to Universidad Pontificia Bolivariana <small>Trabajo del estudiante</small>	1%
8	www.scribd.com <small>Fuente de Internet</small>	1%

9	Submitted to Universidad Señor de Sipan <small>Trabajo del estudiante</small>	<1%
10	repositorio.ucv.edu.pe <small>Fuente de Internet</small>	<1%
11	core.ac.uk <small>Fuente de Internet</small>	<1%
12	www.iadisportal.org <small>Fuente de Internet</small>	<1%
13	M. D. López Luise. "Automatically Modeling Linguistic Categories in Spanish", Innovations in Computing Sciences and Software Engineering, 2010 <small>Publicación</small>	<1%
14	www.ccee.edu.uy <small>Fuente de Internet</small>	<1%
15	worldwidescience.org <small>Fuente de Internet</small>	<1%
16	revestomatologia.sld.cu <small>Fuente de Internet</small>	<1%
17	www.slideshare.net <small>Fuente de Internet</small>	<1%
18	www.uncp.edu.pe <small>Fuente de Internet</small>	<1%

FORMATO DE CONSENTIMIENTO INFORMADO DE LA INVESTIGACION



ANEXOS N° 9 DECLARACION JURADA DE CONSENTIMIENTO INFORMADO EN LA PAGINA WEB FOROS PERU www.forosperu.net

Investigador: Luis Santiago Garcia Merino – Universidad Señor de Sipán

Título: "SISTEMA PREDICTIVO BASADO EN EL MODELO WEB PARA LA PREDICCIÓN DE
PREFERENCIAS DE USUARIO EN EL SITIO WEB FOROS PERÚ"

Yo, **CARLOS PAUL REYES AYALA**, identificado con DNI 16779254, **DECLARO BAJO JURAMENTO:**

Haber sido informado de forma clara, precisa y suficiente sobre los fines y objetivos que busca la presente investigación así como en qué consiste mi participación de acuerdo a la ley de transparencia de la información y acceso público Nro. 27806 y las normas de privacidad de la pagina foros Peru <https://www.forosperu.net/temas/politica-de-privacidad.1171463/>

Estos datos que yo otorgue serán tratados y custodiados con respeto a mi intimidad, manteniendo el anonimato de la información y la protección de datos desde los principios éticos de la investigación científica. Sobre estos datos me asisten los derechos de acceso, rectificación o cancelación que podré ejercitar mediante solicitud ante el investigador responsable. Al término de la investigación, seré informado de los resultados que se obtengan.

Por lo expuesto otorgo MI CONSENTIMIENTO para que se realice la investigación que permita contribuir con los objetivos según el siguiente detalle:

- Comprobar las características epistemológicas del procesamiento de lenguaje natural y minería de opinión mediante clasificación masiva de datos de manera automática.
- Determinar el diagnostico del estado actual de la dinámica del procesamiento de lenguaje natural NLP con la minería de opinión y predicción de preferencias en el sitio web Foros Perú.
- Verificar el modelo de preferencias de usuario, para mejorar el intercambio de información en el sitio web Foros Perú
- Verificar la estructura de un sistema web basado en el modelo de predicción de preferencias de usuario en el sitio web Foros Perú.
- Demostrar la validación y corroboración de los resultados alcanzados en el estudio.

Chiclayo, 27 de Mayo del 2021

Carlos Paul Reyes Ayala
Especialista en diseño y analítica web
CIP 126242-
DNI 16779254



DICTAMEN DE TESIS EXPEDITA PARA SUSTENTACION DE TESIS



DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN Y SISTEMAS

DICTAMEN DE TESIS EXPEDITA PARA SUSTENTACIÓN

El Presidente y los miembros del Jurado Evaluador de la Tesis:

“SISTEMA PREDICTIVO BASADO EN EL MODELO WEB PARA LA PREDICCIÓN DE PREFERENCIAS DE USUARIO EN EL SITIO WEB FOROS PERÚ”.

Presentada por:

GARCÍA MERINO LUIS SANTIAGO

Apellidos

Nombre(s)

*Dictaminan que la Tesis está **EXPEDITA** para su sustentación:*

Chiclayo, 20 de mayo de 2021

Presidente: **Dr. GUTIERREZ GUTIERREZ JORGE LUIS**

Firma

Secretario: **Dr. DIOS CASTILLO CHRISTIAN**

Firma

Vocal: **Dra. LÓPEZ DE LUISE MARÍA DANIELA**

Firma

ACTA DE SUSTENTACION DE TESIS DOCTORAL



ACTA DE SUSTENTACIÓN DE TESIS N° 293

En la Escuela de Posgrado de la Universidad Señor de Sipán, mediante link de ingreso <https://zoom.us/j/99625299600>, de la Plataforma **Zoom** siendo las 4:00:00 PM del día 26 de mayo de 2021, se reunieron los Miembros del Jurado Evaluador, designados según Resolución N° **095-2021/EPGUSS-USS**, de la Tesis Aprobada con Resolución N° **096-2018/EPUSS-USS** denominada:

“SISTEMA PREDICTIVO BASADO EN EL MODELO WEB PARA LA PREDICCIÓN DE PREFERENCIAS DE USUARIO EN EL SITIO WEB FOROS PERÚ”

Que ha sustentado el Tesista:

GARCÍA MERINO LUIS SANTIAGO

Apellidos	Nombre(s)
-----------	-----------

Del **DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN Y SISTEMAS**

Concluida la sustentación el jurado realizó las preguntas al sustentante. Luego el jurado procedió a deliberar libre y reservadamente, por lo cual **ACUERDA:**

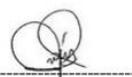
APROBAR POR MAYORÍA

Siendo las 4:55:00 PM del mismo día, se dio por concluida la sustentación, firmando la conformidad los Docentes del Jurado Evaluador y quedando como evidencia el link de grabación:

https://zoom.us/rec/play/qHNNNo5iAkn9E1cJTY_63GE2p1QTwg14a4Qevny38Srp_tsHrviVezmXU9Z1rb3XkkqZc0bdis_FbY-Fd.xIT-3mqVFhBbCfQy

Chiclayo, 26 de mayo de 2021

Presidente: **Dr. GUTIERREZ GUTIERREZ JORGE LUIS**



Firma

Secretario: **Dr. DIOS CASTILLO CHRISTIAN**



Firma

Vocal: **Dra. LÓPEZ DE LUISE MARÍA DANIELA**



Firma