



FACULTAD DE INGENIERÍA, ARQUITECTURA Y URBANISMO

**ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA
DE SISTEMAS**

**PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE
SISTEMAS**

**“DISEÑO DE UN MODELO DE CLASIFICACIÓN DE OPINIONES SUBJETIVAS
UTILIZANDO MINERÍA DE TEXTOS, APLICADO EN ANÁLISIS DE REDES
SOCIALES”**

AUTORES:

Tapia Perales, Miluska Rosangela

Ruiz Montalvo, Osmar Niels

ASESOR:

ING. Chirinos Mundaca, Carlos Alberto

PIMENTEL, DICIEMBRE 2013

RESUMEN

El presente proyecto denominado “**Diseño de un modelo de clasificación de opiniones subjetivas utilizando minería de textos, aplicado en análisis de redes sociales**”, trata sobre la aplicación de técnicas de Minería de datos para interpretar la opinión (Una opinión favorable o desfavorable) de los usuarios de las redes sociales que realizan en su actividad cotidiana (publicaciones) con el fin de analizar las tendencias de un determinado producto, servicio o persona que sirva para la aplicación de marketing a fin de generar ventajas competitivas haciendo el uso de la analítica en la web 2.0.

El Modelo propuesto es usando las técnicas de Minería de datos orientado al procesamiento de textos, por lo que se puede denominar como Minería de Textos, específicamente se utiliza las técnicas de clasificación para generar descubrimiento a partir de las predicciones que ejercen los atributos a una clase. En este caso los atributos a evaluar, que serán los predictores de la clase están conformadas por las palabras que se utilizaran para entrenar el modelo. Las palabras están contenidas en los comentarios que emiten los usuarios de las redes sociales, comentarios que se pueden crear a través de un sistema que se ha diseñado exclusivamente para esta investigación. Teniendo como modelo una red social de Microblogging.

En cuanto al algoritmo dentro del modelo de clasificación se ha optado por utilizar redes bayesianas, ya que estas permiten la utilización de todas las variables predictores para definir la clase, es decir no discrimina ninguna, pero las considera según los pesos asignados utilizando una fórmula probabilística con la cual se obtiene la predicción.

Para el desarrollo de este proyecto se han aplicado las metodologías basándose en su uso, teniendo etapas, como es el caso de aplicar Crisp-DM para la etapa del modelo de Minería de datos y XP para lo que será la etapa del desarrollo del prototipo de red social.

Finalmente el modelo y sistema fueron terminados e implementados del cual se puede mencionar que finalmente obtuvieron un grado de eficiencia de 73.66 % validando los comentarios reales con la predicción que calculó el modelo.

ABSTRACT

This project entitled "Design of a classification model using subjective judgments text mining, applied social network analysis," discusses the application of data mining techniques to interpret the subjective opinion (a favorable or unfavorable opinion) of users of social networks that perform in their daily activities (publications) in order to analyze trends in a certain product, service or person to serve for the application of marketing in order to build competitive advantage by using the analytic Web 2.0.

The proposed model is using data mining techniques oriented word processing, so it can be called Text Mining, specifically using classification techniques to generate discovery predictions from exercising the attributes to a class. In this case the attributes to evaluate, to be predictors of the class are made up of the words that are used to train the model. The words are contained in comments broadcast users of social networks, comments that can be created through a system that has been designed exclusively for this research. Taking as a model Microblogging social network.

As for the algorithm in the classification model has been chosen to use Bayesian networks, as these allow the use of all predictor variables to define the class, ie not discriminate any, but are considered as the weights assigned using a probabilistic formulation which is obtained by prediction.

For the development of this project methodologies have been applied based on its use, taking steps, such as applying Crisp-stage DM Data Mining model and XP for what will be the stage of prototype development social network.

Finally the model and were completed and implemented system which can tell you that finally obtained an efficiency of 75.45% real reviews validating the prediction that the model calculation.