SISTEMA RECOMENDADOR BASADO EN FILTRADO COLABORATIVO Y MINERIA DE DATOS

GERMAN PAEZ BELTRAN

FUNDACION UNIVERSITARIA SAN MARTIN
FACULTAD DE INGENIERIA
PROGRAMA DE INGENIERIA SISTEMAS
BOGOTA, D.C.
2010

SISTEMA RECOMENDADOR BASADO EN FILTRADO COLABORATIVO Y MINERIA DE DATOS

GERMÁN PÁEZ BELTRÁN

FUNDACIÓN UNIVERSITARIA SAN MARTÍN
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA SISTEMAS
BOGOTÁ
2010 I

SISTEMA RECOMENDADOR BASADO EN FILTRADO COLABORATIVO Y MINERIA DE DATOS

DIANA MILENA GERMAN PAEZ BELTRAN
COD. 042036
germanp28@hotmail.com

MONOGRAFÍA DE GRADO

ASESOR TÉCNICO
FABIAN ANDRES GIRALDO GIRALDO
CONSTRUCCION DE SOFTWARE

FUNDACIÓN UNIVERSITARIA SAN MARTÍN
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA SISTEMAS
BOGOTÁ
2010 I

Nota de aceptación

_____

_____

_____

_____

_____

_____

_____

Fabián Andrés Giraldo Giraldo
Asesor

_____

Andrés Solarte
Jurado 1

_____

Christian Benavides
Jurado 2

Bogotá, 5 de Junio de 2010

A mi Familia y Amigos,
Que siempre llenan de felicidad mi entorno
Y me inspiran valor para seguir adelante
En la carrera de la vida sin temor a nuevos retos.

AGRADECIMIENTOS

A Dios, por brindarnos la dicha de la salud y bienestar físico y espiritual

A la señora Bertha Lilia y sus hijos quienes me brindaron un gran apoyo, tanto económico como moral y emocional para lograr sacar adelante mi título profesional

A mis padres y hermanos, como agradecimiento a su esfuerzo, amor y apoyo incondicional, durante la formación tanto personal como profesional.

A mi Asesor que siempre estuvo disponible con sus valiosos consejos para que este trabajo se llevara a cabo.

A mis docentes, por brindarnos su guía y sabiduría en el desarrollo de mi formación profesional

A mis compañero de universidad que siguieron el paso de graduación conmigo y fueron de gran ayuda en el crecimiento profesional

CONTENIDO

LISTA DE TABLAS

pág.

LISTA DE FIGURAS

pág.

RESUMEN

El proyecto de grado "Sistema Recomendador basado en filtrado colaborativo y minería de datos" es una aplicación que tienen por objetivo el descubrimiento, análisis y recomendación de información relevante, tomando en consideración el comportamiento y preferencias de los usuarios de una tienda de comercio electrónico.

La aplicación está basada en la utilización de una herramienta ETL llamada Scriptella con la cual se logra la obtención de información desde la tienda de comercio electrónico, la transformación de datos y el posterior almacenaje en un DataMart, con el objeto de aplicar un algoritmo genético para el proceso de búsqueda de artículos recomendados para cada usuario de la tienda de comercio electrónico, además de la exploración de datos con la utilización de la herramienta WEKA y sus clases de asociación y secuenciación, que permiten el descubrimiento de vínculos entre productos y frecuencias de compra.

INTRODUCCIÓN

El comercio electrónico y los mecanismos de filtrado colaborativo y tratamiento de datos para el aprovechamiento de la información y la obtención de ventajas sobre los competidores, está enmarcado en varios escenarios como lo son la educación, la investigación y el comercio electrónico, que es la razón central del desarrollo de este trabajo.

A lo largo de la historia del comercio electrónico y los sistemas informáticos se ha logrado un avance en la investigación de métodos y estrategias que lo ayuden a comprender el comportamiento del usuario en tiendas de comercio electrónico. Al principio se basaban en simples sistemas estadísticos para mantener la información del flujo de venta de productos y los mejores usuarios. Al pasar el tiempo y con el crecimiento de competencia se hizo necesaria la adopción de nuevos métodos y técnicas para persuadir a los usuarios de tiendas electrónicas para la compra de productos y la extracción de información implícita de gustos e inclinaciones sobre los artículos o productos que se presentan en la tienda de comercio electrónico.

Es por esto que en la lucha de mantenerse a la vanguardia es necesario incorporar mecanismos y herramientas que faciliten a las tiendas de comercio electrónico, a obtener el máximo aprovechamiento de los datos contenidos en la tienda, valiéndose de técnicas avanzadas de optimización y exploración de datos como lo son el filtrado colaborativo basado en algoritmos genéticos y exploración de datos basados en minería de datos. Es de esta oportunidad que surge la necesidad de crear un sistema capaz de explorar y analizar la información de los usuarios para la obtención de nuevos datos relevantes a cada usuario registrado en la tienda de comercio electrónico.

Partiendo de este escenario se creó un sistema multiagente que permita encontrar productos recomendados para cada usuario, valiéndose de técnicas de filtrado colaborativo y minería de datos, utilizando la información dejada por cada usuario de las calificaciones otorgadas a cada producto y la información contenida en el carrito de compra.

El presente trabajo se divide en cuatro etapas que ayudan a completar el objetivo general propuesto, donde la primera etapa comienza con la extracción de información de la tienda de comercio electrónico con la ayuda de una herramienta ETL llamada Scriptella para la obtención de la información, aquí se realizan tratamientos sobre los datos para su posterior almacenaje en una bodega de datos. Partiendo de esta información se inicia la segunda etapa en donde se

procede a la construcción de un algoritmo genético capaz de generar recomendaciones para cada usuario de la tienda, apoyado en la información extraída de las votaciones de productos por cada cliente.

Para la tercera etapa se utiliza la información del carrito de compra de los usuarios con el fin de reglas de asociación de productos y realizar un intercambio de recomendaciones de acuerdo a sus preferencias, esto apoyado en la exploración de información en un DataMart para cada usuario.

En la cuarta etapa se genera un proceso de análisis de información obteniendo información de la bodega de datos, con el fin de extraer las secuencias de compra de productos para poder generar reglas de secuenciación para determinar los productos que un usuario puede llegar a comprar en próximas visitas. Se realiza una modificación sobre la tienda de comercio electrónico con el objetivo de presentar a cada usuario los recomendados, cada vez que este visite un producto en la tienda.

Finalmente se busca, con el desarrollo de este trabajo, que los administradores de tiendas de comercio electrónico se acerquen al máximo al aprovechamiento de la información almacenada, para presentar un beneficio positivo en el movimiento de productos en la tienda, logrando así obtener mayores ganancias sobre los productos vendidos en ella.