



UNIVERSIDAD
TECNOLÓGICA
METROPOLITANA
del Estado de Chile

Proyecto Minería de datos: “Análisis de recomendaciones de películas basado en perfiles de usuarios”

Profesor: Pablo Figueroa Plaza
Asignatura: Minería de datos
Estudiante: Francisco Sánchez Fuentes
Fecha: 15/09/2017

Indicé

DEFINICIÓN DEL PROBLEMA.....	4
<i>Objetivo general</i>	4
<i>Objetivos específicos</i>	4
SOLUCIÓN PROPUESTA.....	5
<i>Solución mediante metodología KDD</i>	5
DESCRIPCIÓN DE LOS DATOS SELECCIONADOS.....	7
<i>Recomendación de películas</i>	7
DESCRIPCIÓN DE LAS VARIABLES INPUT	7
DESCRIPCIÓN DE LAS VARIABLES OUTPUT	7

Definición del problema

En este proyecto se busca realizar recomendaciones de películas basado en la calificación proporcionada por un usuario.

El problema radica en la identificación de perfiles de usuario basado en gustos por ciertas películas, la calificación de una película se encuentra medido por la variable *Rating* donde el usuario clasifica en un rango definido que tan buena encuentra que es la película, este procedimiento se reitera las veces necesarias por los distintos usuarios que quiera clasificar la misma película u otra.

Los perfiles de usuario y sus recomendaciones por película están sujetos a factores subjetivos ya que cada recomendación depende de una opinión personal del usuario. Otro factor que puede incidir en la calificación son los factores sociales como Revistas, Redes sociales y Blog entre otros.

Objetivo general

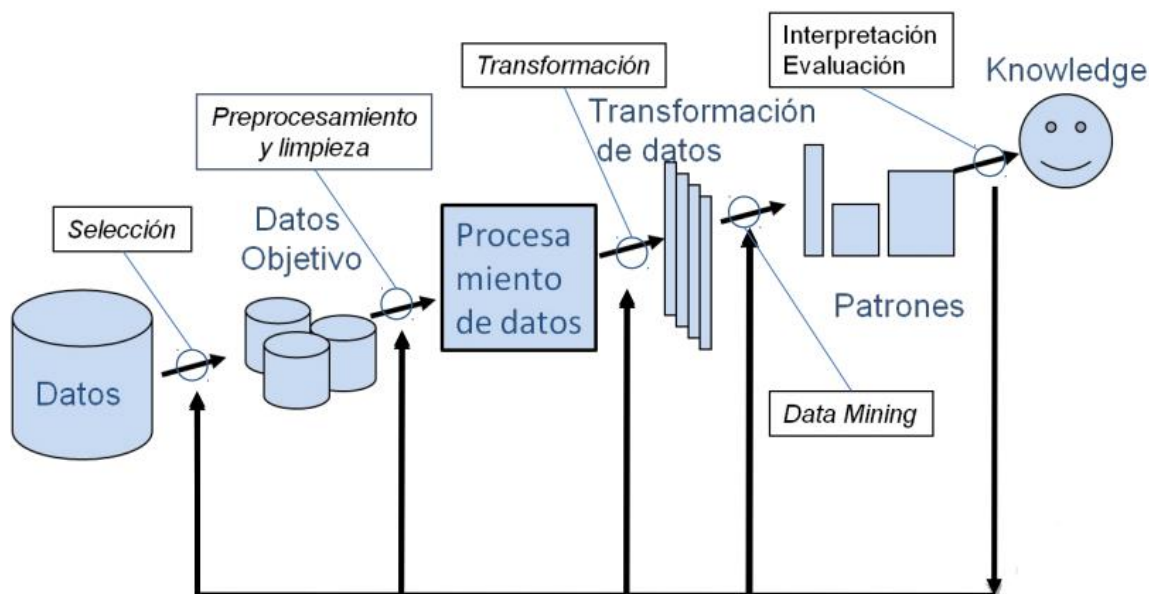
Identificar perfiles de usuario y utilizar un modelo que entregue como respuesta predicción de recomendaciones de pelicular para usuario con un perfil similar.

Objetivos específicos

- Buscar y recopilar conjunto de datos, factores y variables que sirvan para lograr la identificación de perfiles.
- Analizar, discriminar y definir conjunto de datos y variables que se utilizaran.
- Definir el modelo que mejor se adapte a la problemática.

Solución Propuesta

Solución mediante metodología KDD



Etapas:

Datos

Entendimiento de los datos, para que sirva cada variable, que factores inciden sobre la variable y obtener la mayor cantidad de datos siguiendo el modelo de negocio de la aplicación. En esta etapa no se discrimina que datos utilizar en el modelo, se realiza una recolección de estos.

Datos Objetivo

Identificar conjuntos de datos objetivo, en esta etapa corresponde relacionar y agrupar conjuntos de datos, y a su vez agrupar subconjuntos de datos relacionados que puedan existir.

Procesamiento de datos

Limpieza y preprocesamiento de datos. Para esta etapa se busca realizar operaciones básicas de eliminación de ruido, campos vacíos, completar falta de información o descartar. El objetivo de esta etapa es cerciorarse que los datos se encuentren íntegros.

Transformación de datos

Reducción de datos y proyección, en esta etapa se buscan características útiles para la representación de los datos en función del objetivo de la tarea (Reducción de la dimensionalidad).

Minería de datos

Análisis exploratorio, de hipótesis y selección de modelo. Esta etapa corresponde a la elección del algoritmo de minería de datos que se utilizará para la búsqueda de patrones en los datos. Posteriormente se aplica el modelo seleccionado para la búsqueda de patrones de interés.

Patrones

Interpretación de los patrones minados, en esta etapa se busca visualizar los patrones extraídos y visualización de los datos que figuran de los modelos.

Conocimiento

Conocimiento descubierto, en esta etapa se hace uso del conocimiento adquirido implementando a un sistema el resultado (programa, modelo, etc), adoptando de esta forma las medidas o simplemente la documentación y presentación de informes a las partes interesadas.

Se puede llegar a realizar una implementación o puede que solo se necesite ese conocimiento adquirido como parte de una documentación u análisis inteligente de los datos.

Descripción de los datos seleccionados

Recomendación de películas

- **CustomerID:** Número identificador del cliente.
- **MovieID:** Número identificado de una película.
- **Rating:** Clasificación de la película (Rango: 1 - 5).
- **Date:** Fecha cuando fue calificada la película.
- **YearOfrelease:** Fecha de lanzamiento de la película.
- **Title:** Título de la película.

Descripción de las variables Input

- **CustomerID:** Esta variable es el identificador del cliente, es importante que se encuentre dentro de las entradas ya que se utilizara para identificar el perfil de los usuarios.
- **MovieID:** Esta variable al igual que el identificador del cliente será importante para generar un enlace entre perfiles de usuario y películas.
- **Rating:** Esta variable sirve para generar grupos (Cluster) de usuario a los cuales se ofrecerán recomendaciones de películas basados en el parentesco entre perfiles de usuario.

Descripción de las variables output

- **Predicción:** Esta variable de salida corresponde al Rating proporcionado por el modelo que dicta basado en el perfil del usuario si la película le gustara o no. La salida puede estar en el mismo rango que se clasifican las películas y su exactitud es dependiente del modelo.

Predicción: {1,2,3,4,5}