

## Laboratorio 1

### Descripción y objetivo

El laboratorio está diseñado para familiarizar al estudiante con varios *frameworks* para desarrollo de modelos de recomendación: el *framework surprise* para Python y una exploración inicial de otras implementaciones de filtrado colaborativo.

Los objetivos del laboratorio son:

- Familiarizarse con la estructura típica de archivos que representan la matriz de utilidad usada por los algoritmos de recomendación
- Crear un algoritmo de recomendación simple dentro de la estructura del *framework*
- Utilizar la implementación de un algoritmo de recomendación visto en clase.
- Realizar una validación estadística *offline* de los resultados obtenidos por el recomendador.

### Material

- Entorno de desarrollo Python 3.9, versiones superiores pueden causar problemas de compatibilidad en librerías. Para despliegues locales les recomiendo el uso de anaconda, será de utilidad en los proyectos del curso
- *Dataset* Movielens100k
  - disponible en: <https://files.grouplens.org/datasets/movielens/ml-100k.zip>
- Google colab, Saturn Cloud, Amazon Sagemake o entorno local
  - <https://colab.research.google.com>
  - [Saturn Cloud | Code-First AI Infrastructure | Saturn Cloud](#)
  - [SageMaker Studio Lab](#)
- Notebook Python
  - [https://github.com/valegrajales/MINE4201-2026/blob/main/lab01/MINE4201\\_SR\\_Laboratorio\\_1.ipynb](https://github.com/valegrajales/MINE4201-2026/blob/main/lab01/MINE4201_SR_Laboratorio_1.ipynb)

### Metodología

Realice el laboratorio en los grupos previstos para el trabajo práctico del curso. Se realiza una entrega por grupo.

### Entregable

La entrega de resultados del laboratorio consta de dos partes: El software desarrollado y un informe de laboratorio.

Realice un informe donde documente su desarrollo y avance de cada uno de los puntos del laboratorio. Inicialmente muestre, mediante imágenes de pantalla, su logro de los objetivos de cada punto.

En los puntos en los cuales se espera que usted explore y experimente, realice una pequeña descripción de lo que encuentra y obtiene.

Documente los resultados obtenidos en los desarrollos realizados y haga un breve análisis sobre los resultados. En particular, no olvide resaltar sus hallazgos con respecto al alcance y limitaciones del framework analizado

### Formato y hora de entrega

Realice su entrega del laboratorio de la siguiente forma:

- Archivo en formato **zip**, nombrado de la siguiente forma: **Lab1-NN-login1-login2 (-login3) .zip**, donde NN es el número del grupo y luego se encuentran los login uniandes de los integrantes del grupo.
- Contenido del archivo:
  - Notebook modificado con los resultados obtenidos
  - PDF del Notebook (imprima como **pdf** el notebook desarrollado)

- Documento de informe en formato pdf, nombrado de la siguiente forma: **Lab1-NN-login1-login2 (-login3) .pdf**

Fecha límite de entrega: febrero 18, hasta las 20:50 en la actividad de bloque neón.

## Desarrollo

### 1. Entendimiento de los datos

Para evaluar un sistema de recomendación es necesario un conjunto de datos, en este caso descargue el conjunto de datos (*dataset*) Movielens100k.

El conjunto de datos Movielens100k tiene 100,000 ratings (entre 1-5) de 943 usuarios para 1682 películas, lea el archivo README para obtener más información acerca de este conjunto de datos.

El archivo `u.data` es el archivo principal del conjunto de datos y contiene los *ratings*, uno por línea. Este archivo representa la matriz de utilidad que usan los algoritmos de recomendación basados en filtrado colaborativo.

El archivo `u.item` es el archivo que describe los items. En este se pueden encontrar los nombres de las películas que tiene el dataset, entre otra información.

### 2. Análisis de un dataset y uso de framework Surprise

El *framework* Surprise (<http://surpriselib.com/>) es un proyecto para la evaluación de algoritmos de recomendación desarrollado en Python. Este permite explorar el uso de diferentes sistemas de recomendación. Para su uso vamos a usar la herramienta google colab.

- Ingrese a google colab <https://colab.research.google.com> e importe el notebook disponible en [https://github.com/valegrajales/MINE4201/blob/main/labs/lab01/MINE4201\\_SR\\_Laboratorio\\_1.ipynb](https://github.com/valegrajales/MINE4201/blob/main/labs/lab01/MINE4201_SR_Laboratorio_1.ipynb)
- Siga las instrucciones de la guía y reporte sus hallazgos en el informe

### 3. Otros frameworks para filtrado colaborativo

Surprise no es un framework robusto o flexible a extensiones. Revise alguno de los siguientes frameworks, verifique su documentación e indique de manera simple cómo se crearía en este un sistema de recomendación similar al implementado en este laboratorio. Si escoge uno disponible en Python realice la prueba en el notebook que va a entregar.

- Lenskit: <http://lenskit.org/>
- RankSys: <https://github.com/RankSys/RankSys>
- Apache Mahout (Los algoritmos clásicos están en la versión 0.9), se adjunta proyecto eclipse con desarrollo de prueba realizado para versiones anteriores del curso <https://mahout.apache.org/users/recommender/userbased-5-minutes.html>