BONSAI - Analytica

Derric Lyns

Van Hack

Prediction system

# Introduction

This software implements a recommender system like that of Netflix which is one of the famous applications in data science and machine learning. A recommender system uses statistical algorithm that seeks to predict user intent. The intuition is that similar types of users are likely to have similar intent.

In this project I’am using py-analytica, sci-learn and correlation algorithms to demonstrate recommender system.

# Training

The training data is taken from UserEvents.csv which has about 1,41,000 records of transactions of user. There are about 1000 user records and 3000 books. The time taken for machine learning is 1 minute per record totaling 3000x1x1000 processing per unit time. It will take atleast 3000 minutes which is 50 hours making it a straight 2 days for the system. Development time is about 7 days. Totaling 12-15 working days to make up to the challenge.

# Requirements

Install python 3.6 above

Install sci-learn, this requires a high performance system with server capabilities with at least 4 cores of processor and 16 GB RAM

# Types of recommender systems

1. Content based filtering: This is based on similarities between different data sets. This is based on statistics. Correlation is used for this model. Please refer **analytica.py**
2. Collaborative Filtering: This leverages the power of crowd. If a user A likes products X and Y, and if another user likes product X, then the user may also like Y. please refer **naïve-analytica.py**

# Sources

The system has been implemented using both the recommender systems

# Test cases

The test data is not processed in this application. This is first check-in. results will be submitted later