

Three ideas for capstone project 2

Project Idea 1:

Data set: MovieLens Datasets

<http://grouplens.org/datasets/movielens/latest/>

Data set introduction:

These datasets will change over time, and are not appropriate for reporting research results. We will keep the download links stable for automated downloads. We will not archive or make available previously released versions.

Full: 26,000,000 ratings and 750,000 tag applications applied to 45,000 movies by 270,000 users. Includes tag genome data with 12 million relevance scores across 1,100 tags. Last updated 8/2017.

Story to tell: To build my personal movie recommender system.

Potential client: Recommender systems are necessary for large companies like Google or Facebook, because it is valuable from a perspective of revenue and engagement (Facebook ads, Youtube recommendations).

.

Project Idea 2:

Data set: Personalized Medicine: Redefining Cancer Treatment

<https://www.kaggle.com/c/msk-redefining-cancer-treatment/data>

Story to tell: A lot has been said during the past several years about how precision medicine and, more concretely, how genetic testing is going to disrupt the way diseases like cancer are treated. This project aims to predict the effect of Genetic Variants to enable Personalized Medicine.

Potential client: Biotech Company, pharmaceutical company.

Project Idea 3:

Data set: Twitter data set for Arabic Sentimental Analysis

<http://archive.ics.uci.edu/ml/datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis>

Data set introduction: By using a tweet crawler, 2000 labelled tweets (1000 positive tweets and 1000 negative ones) were collected. On various topics such as: politics and arts. These tweets include opinions written in both Modern Standard Arabic (MSA) and the Jordanian dialect.

Story to tell: The selected tweets convey some kind of feelings (positive or negative) and the objective of our model is to extract valuable information from such tweets in order to determine the sentiment orientation of the inputted text.

Potential client: Social media, mobile companies such as T mobile, Verizon, Sprint.