Hema Saxena

Capstone Two: Project Ideas

Idea 1

Problem:

Can you identify the total sales for every listed product and store in the next month?

Details:

The task is to forecast the total amount of products sold in every shop for the test set. The list of shops and products slightly changes every month. Creating a robust model that can handle such situations is part of the challenge.

I anticipate building an algorithm that will act firstly as a benchmark. Stores can use it to ensure that business goes as usual level they are going to achieve if nothing changes in their strategy. Moreover, they can calculate the incremental value of their new actions on top of this benchmark.

Second, it can be utilized for planning. They can plan their demand and supply actions by looking at the forecasts. It can help to see where to invest more.

Also, it can be used as an excellent guide for planning budgets and targets.

Dataset:

Dataset is available at Kaggle <https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data>

---------------------------------------------------------------------------------------------------------------------------------------

Idea 2

Which 5 movies can be recommended to a user that he will enjoy watching, based on the choice of a movie he liked previously?

Details

Recommendation systems have the potential to change the way websites communicate with users and to allow companies to maximize their ROI based on the information they can gather on each customer's preferences and purchases

Dataset

1. TMDB Dataset(Kaggle) about 5000 Movies data(43.62MB)

(<https://www.kaggle.com/tmdb/tmdb-movie-metadata>)

1. 45,000 movies listed in the Full MovieLens Dataset available on
2. Kaggle for data until July 2017)

<https://www.kaggle.com/rounakbanik/the-movies-dataset>

1. Latest data available on grouplens <https://grouplens.org/datasets/movielens/latest/>

---------------------------------------------------------------------------------------------------------------------------------------

Idea 3

Problem:

Given past credit card transactions with the knowledge of the ones that turned out to be fraud, which new transaction can be marked as fraudulent?.

Details:

Fraud is a major problem for the whole credit card industry that grows bigger with the increasing popularity of electronic money transfers. To effectively prevent the criminal actions that lead to the leakage of bank account information leak, skimming, counterfeit credit cards, the theft of billions of dollars annually, and the loss of reputation and customer loyalty, credit card issuers should consider the implementation of advanced Credit Card Fraud Prevention and Detection methods. Machine Learning-based methods can continuously improve the accuracy of fraud prevention based on information about each cardholder’s behaviour.

I anticipate building a Fraud Detection model that can be trained continuously whenever new data arrives, so new fraud schemas/patterns can be learned, and fraudulent data detected as early as possible.

Dataset:

Dataset is available at Kaggle (<https://www.kaggle.com/mlg-ulb/creditcardfraud>)

---------------------------------------------------------------------------------------------------------------------------------------

Idea 4

Problem:

Can you build a machine learning classifier that accurately predicts which of the 2 groups (if any) will turn out to be more profitable?

Details:

The cost of marketing can be very high, meaning that the decision about which customer group to target is of great financial importance.

Any modern business which invests considerably in marketing can utilize this model. After identifying profitable segments, they can try to target their marketing investment towards those groups of customers who have the potential to bring back the highest Return on Investment for the company.

Dataset:

Dataset is available at Kaggle (<https://www.kaggle.com/tsiaras/predicting-profitable-customer-segments>)