Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

*Team Member’s Name, Email and Contribution:*

*1) Pritam Gangwar*

*E-mail:* [*pritamgangwar786@gmail.com*](mailto:pritamgangwar786@gmail.com)

*- Data sorting.*

*- Data analysis.*

*- Approach towards plan.*

*- Frame work of project.*

*- Graphical representation.*

*- Bar plot and Heat map.*

*- Model presentation.*

*- Model selection and implementation.*

*- Cross validation on model.*

*- Implementation of grid parameters.*

*2) Piyush Mishra*

*E-mail:*

*- Data visualization.*

*- Sorting of values.*

*- Approach towards multiline graph.*

*- Pi-plot and Heat map.*

*- Various model implement.*

*- Histogram plot.*

*- Suggestion of grid parameters.*

*- Project summery template.*

*- Analyzing results of model.*

Problem definition:

Predicting sales performance is one of the key challenges every business face. It is important for firms to predict customer demands to offer the right product at the right time and at the right place. The importance of this issue is underlined by the fact that figuratively a bazillion consulting firms are on the market trying to offer sales forecasting services to businesses of all sizes. Some of these firms rely on advanced data analytics techniques, the kind of which we will be covering in CS-109 classes.

EDA on given Data set:

Rossmann is the largest drugstore in Germany. Moreover, it operates over 3,000 drug stores in 7 European countries. In 2015, Rossmann store managers are tasked with predicting their daily sales for up to six weeks in advance. Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. With thousands of individual managers predicting sales based on their unique circumstances, the accuracy of results can be quite varied.

In their first Kaggle competition, Rossmann challenged Kagglers to predict 6 weeks of daily sales for 1,115 stores located across Germany. Reliable sales forecasts enable store managers to create effective staff schedules that increase productivity and motivation.

Milestones:

1. Project Selection: Form teams of 2 or 3 and select a project from the provided list.

2. Literature Study: By running through the data science process you will be able to answer the following research questions necessary for this project:

- To what extend is sales performance influenced by factors like: promotions,

competition, school and state holidays, seasonality, and locality.

- What is an appropriate model to predict sales?

Go through the following resources for background on the project and write a 1 page summary for all three of them:

- "Schlecker drugstores to close for good" (Deutsche Welle, June 4th, 2012)

"Attempts to rescue the bankrupt drugstore chain Schlecker, once Europe's

largest, have failed. The remaining 3200 stores will close, and the last 13,200

employees will lose their jobs.”, URL:

<http://www.dw.com/en/schleckerdrugstores-to-close-for-good/a-15996229>

Datasets contain historical sales data for 1,115 Rossmann stores. The task is to forecast the "Sales" column for the test set. Note that some stores in the dataset were temporarily closed for refurbishment! In order to deal with this problem, check the Filling Gaps in the Training Set article.

3. Exploratory Data Analysis: Exploring the given data, will help in:

- Reviewing the raw data

- Exploring relationships

- Dealing with NA/missing values.

4. Propose a Model: Propose methodologies and ideas to be implemented, tested and

interpreted for your final project. Pay a specific attention to:

- Correlation between time and sales

- Seasonality

- Autocorrelation

Apply different statistical models and compare them in order to choose the best performing one. You can start with Decision Tree Regression or Linear regression, and then use some of the Ensemble methods. This part of the project is all up to you!

Data Summary:-

Id - an Id that represents a (Store, Date) duple within the set

● Store - a unique Id for each store

● Sales - the turnover for any given day (Dependent Variable)

● Customers - the number of customers on a given day

● Open - an indicator for whether the store was open: 0 = closed, 1 = open

● StateHoliday - indicates a state holiday. Normally all stores, with few exceptions, are closed on state holidays. Note that all schools are closed on public holidays and weekends. a = public holiday, b = Easter holiday, c = Christmas, 0 = None

● SchoolHoliday - indicates if the (Store, Date) was affected by the closure of public schools

● StoreType - differentiates between 4 different store models: a, b, c, d

● Assortment - describes an assortment level: a = basic, b = extra, c = extended. An assortment strategy in retailing involves the number and type of products that stores display for purchase by consumers.

● CompetitionDistance - distance in meters to the nearest competitor store

● CompetitionOpenSince[Month/Year] - gives the approximate year and month of the time the nearest competitor was opened

● Promo - indicates whether a store is running a promo on that day

● Promo2 - Promo2 is a continuing and consecutive promotion for some stores: 0 = store is not participating, 1 = store is participating

● Promo2Since[Year/Week] - describes the year and calendar week when the store started participating in Promo2

● PromoInterval - describes the consecutive intervals Promo2 is started, naming the months the promotion is started anew. E.g. "Feb,May,Aug,Nov" means each round starts in February, May, August, November of any given year for that store

Conclusion:-

- Decision tree should be used in the analysis as it outperforms other models and has

a r-squared score of 0.97.

- Store type B has the maximum number of average sales.

- Maximum number of stores are closed on Sunday. So there is increase in average

sales on Monday.

- Assortment type B has the maximum number of average sales.

- 17.9% of data suggest that closure of public-school affect the sales.

- Average competition distance is 5.4.

- Sales in 2013, 2014 and 2015 are stagnant, there is a hardly increase in sales.

- As the competition distance increase the sales decrease. This shows that customers

are churning to Rossmann Stores, so we should open more store where competition

is available.

- Store type B should be increased and assortment B should be added to it.

- When there is closure of public school more promotion should be done to increase

the sales.

- Stores should be open on Sundays as there is demands on Sundays also, some

people might have gone to the competition stores for the products.

- If there was use of promo the sales increased, use of promo should be in increased

in the stores specially during holidays.

*Please paste the GitHub Repo link.*

*Github Link:- https://github.com/pritamgangwar/Retail-Sales-Prediction*