

School of Electronic Engineering and Computer Science Queen Mary University of London

MSc PROJECT DEFINITION 2021-22

This project definition must be undertaken in consultation with your supervisor. The feasibility of the project should have been assessed and the project aims should be clearly defined.

Submission of this document implies that you have discussed the specification with your supervisor.

Project Title: Forecasting and upgrading the sales by using the Machine learning

Module: ECS750P/ECS751P - 2021/22 (JANUARY START ONLY)

Student name: Vinoth Murugan

Student e-mail: ec211235@qmul.ac.uk

Student Number : 210958005

Student phone number: 07423297225

Supervisor: Mahesha Samaratunga

Contact email: m.samaratunga@qmul.ac.uk

PROJECT AIMS:

For Business Organizations Needs improvement Clients or users every year. The accuracy of sales forecasts has a big impact on the business. Data mining techniques can be used to extract hidden knowledge from large datasets to improve the accuracy and effectiveness of predictions [1]. With accurate forecasts, the company can increase market expansion and revenue generation. The foundation of a smart budget is the use of data mining tools. Data mining tools are especially effective in transforming large amounts of data into meaningful information for cost and revenue forecasts.

At the organizational level, forecasting is an important input to many decision-making processes in various functional areas such as operations, marketing, sales, manufacturing, and finance [3]. The preparatory dataset used in this task will experiment with three or more AI techniques and evaluate the model exhibit. From the point of view of execution accuracy, the best solution calculation for the expected value is selected. The models are decision trees, gradient boosted trees, and generalized linear models [4]

PROJECT OBJECTIVES:

- This study aims to identify reliable sales trend forecasting methods that use data mining techniques to generate the most revenue possible
- This research will show how an organisation is managed with the work demand, sales, and operating profit
- The most accuracy and reliability intelligent sales trend prediction model available
- The accurate forecasting will lead to improve market and sales
- By forecasting we can upgrade the sales with customer behaviour

METHODOLOGY:

- The models are exhaustive and reliability for sales prediction is by use of data mining
- The dataset is collected from the online store or website like Kaggle for last five consecutive year
- The analysis is done by data collection, understanding the data, evaluating, modelling, and deploying

- The transformation of dataset to modelling finding the correlation between the qualities that have a favorable impact on sales revenue
- The forecasting sales for past years and so we can easily predict the upcoming year sales
- Comparison of classifier and by accuracy finding out the good classifier
- The prediction is done by machine learning techniques by the expected models Decision tree, Gradient boosted trees, Generalized linear model or other best models.

PROJECT MILESTONES:

- ⇒ The data collection, understanding and evaluating
- ⇒ Correlation, modelling and evaluating
- ⇒ Forecasting and visualization
- ⇒ Implementing machine learning techniques
- ⇒ Developing and dissertation draft

REQUIRED KNOWLEDGE/ SKILLS/TOOLS/RESOURCES:

- Python 3.7.7
- NumPy
- Pandas
- Scikit
- matplotlib

The above skills can be done using google colab and Jupyter notebook

- Tableau or Excel or power BI
- Business understanding
- Data management
- Time series skills
- Analyse the machine learning











Literature Review:

1. A. Sayli, I. Ozturk and M. Ustunel, "Brand loyalty analysis system using K-Means algorithm", *Journal of Engineering Technology and Applied Sciences*, vol. 1, no. 3, 2016.
2. Q. Huang and F. Zhou, "Research on retailer data clustering algorithm based on spark", *AIP Conference Proceedings*, vol. 1820, no. 1, pp. 080022, 2017, March.
3. S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), 2018, pp. 53-58, doi: 10.1109/iCCECOME.2018.8659115.

4. S. H. Sastry, P. Babu and M. S. Prasada, *Analysis & Prediction of Sales Data in SAP-ERP System using Clustering Algorithms*, 2013.
5. M.N. Maingi, *A Survey on the Clustering Algorithms in Sales Data Mining*.

TIMEPLAN:

The below timeline is provisional

<input type="checkbox"/>	Task		Start date	End date
<input type="checkbox"/>	Getting the dataset		Jul 15	Jul 20
<input type="checkbox"/>	Understanding		Jul 21	Jul 29
<input type="checkbox"/>	Evaluating		Jul 30	Aug 10
<input type="checkbox"/>	Correlation		Aug 11	Aug 17
<input type="checkbox"/>	Modelling		Aug 18	Aug 21
<input type="checkbox"/>	Deployment		Aug 22	Sep 2
<input type="checkbox"/>	Implementing machine learning		Sep 3	Oct 1
<input type="checkbox"/>	Dissertation rough draft		Oct 2	Oct 15
<input type="checkbox"/>	Dissertation draft		Oct 16	Oct 25
<input type="checkbox"/>	Project submission			Nov 21