**CSE523 Machine Learning**

**Prof. Mehul Raval**

**Time Series Sales Forecasting using Machine Learning Algorithms**

**Week-2 Progress Report**

| **Name** | **Enrolment Number** |
| --- | --- |
| Soham Jagrit | AU2040060 |
| Nitya Panchal | AU2040249 |
| Freya Shah | AU2020039 |
| Gunisha Giri | AU2020210 |

**Research Paper**

**The Reference Paper link:** <https://ieeexplore.ieee.org/document/9071463>

This paper uses different forecasting methodologies to forecast Amazon's future quarterly net sales, based on its historic quarterly data. Three methods are used: Linear regression, Ridge regression and ARIMA. To measure the accuracy of predictions made by a particular model, measures such as MAPE (Mean absolute percentage error) and RMSE (Root mean square error) are used. The proposed models explain various stages in the forecasting process and provide comparative performance of these models. For an online retailer, the special online shopping festival, such as Black Friday and Cyber Monday, makes up a substantial percentage of sales. Therefore, it is useful to forecast future quarterly net sales to prepare for future Black Fridays and Cyber Mondays. The three methods are implemented to forecast Amazon's quarterly sales, which helps Amazon manage its future operations. The data was preprocessed to check for missing values and outliers and further, a time series decomposition is performed to separate components of the data. An ARIMA model was fitted to the data and the model’s performance was evaluated using statistical measures mentioned above. This method can be used in all businesses in order to improve their inventory management.

**Dataset**

We have selected the dataset which was posted in a sales forecasting Kaggle completion. The link for the dataset is: [https://github.com/BekBrace/Sales-For…](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbW5wSEVEcUZ4enBrV050Vk5zUUFUc2JTN0twd3xBQ3Jtc0ttUF9nNEZURWxKbHYyd0ktVlNKSUtEMkNpemgxTHN5SDhzRklsTlJreHphZ1phakN5eV9xbDJ6T2NFUVBGOFhmZTJLWjNjbGV5UkVKcDZ5cldHUExNVEI3clpTZDE2UDlSSWpSVFVMZzVVbGt0QnQzNA&q=https%3A%2F%2Fgithub.com%2FBekBrace%2FSales-Forecast-data-csv&v=20fbgWm5M2o)

We imported the dataset and transformed it in time series format and preprocessed it for missing values and outliers. We have to check whether the data is stationary or not for the ARIMA model.

**Work we did**

In this week we completed the following tasks:

* Read the research paper thoroughly and tried to implement the learnings with respect to our dataset.
* Imported the dataset and checked for missing values.
* Too a close look at the ARIMA model and how to preprocess the data for this particularly and appropriately.

**Tasks for next week**

For the next week,

* We would draft the first report and power point presentation for mid semester presentations
* We would start the implementation of the algorithms mentioned above into a code with respect to the chosen dataset.