

Machine Learning Activity 11

Due Date: 14 September 2024

Instructions:

You may complete this task individually or in a group consisting of no more than six students. This assessment will contribute to your final mark for Assessment 4. **Depending on the quality of your group's submission, this assessment may constitute the entirety of Assessment 4.** However, if the instructor deems the submission unsatisfactory, you will be required to participate in an in-class assessment scheduled for the 17th of September 2024, from 08:00 to 17:00. It is in your best interest to perform well on this assignment to be exempt from the scheduled in-class assessment.

Your submission must demonstrate a comprehensive understanding of the phases involved in a machine learning project. Additionally, you will be required to present your assessment in person to the lecturer on the 17th of September 2024.

One member of the group must submit the GitHub link to the project before midnight on the due date.

Theme: "Empowering Spaza Shops: Time Series Forecasting for Local Businesses in South Africa"

Spaza shops are the backbone of many communities in South Africa, providing essential goods and services in both urban and rural areas. These small, informal businesses are often family-run and play a crucial role in the local economy. However, like many small businesses, spaza shops face challenges such as fluctuating demand, supply chain issues, and competition from larger retailers. Accurate forecasting can help these businesses optimize their inventory, reduce waste, and improve profitability.





You are therefore required to create predictive models that can help spaza shop owners better understand and anticipate their business needs. Participants will explore time series forecasting techniques to predict sales, demand for specific products, and other business metrics crucial for the success of spaza shops.

Objective:


Develop a solution that uses time series forecasting to predict various business metrics for spaza shops. The solution should incorporate three forecasting models: Prophet, ARIMA, and LSTM. You must source out your own dataset, focusing on data relevant to small businesses or retail sales, and apply these models to solve a common problem faced by spaza shop owners.

Possible Datasets:


Participants should find or create datasets that reflect the operations of spaza shops. Examples include:

-  Daily or weekly sales data of spaza shops.
-  Inventory levels of essential goods.
-  Seasonal trends in product demand (e.g., increased sales during holidays or specific seasons).
-  Impact of local events on sales (e.g., festivals, community gatherings).

Tasks:

-  **Data Collection and Preprocessing:** Participants must find a dataset that aligns with the theme. They should clean and preprocess the data, handling missing values, outliers, and seasonality where necessary.

Model Implementation:

-  **Prophet:** Participants will use Prophet to model the trend and seasonality of the data. They should highlight how holidays, weekends, and other factors are accounted for in the model.

- 📊 **ARIMA:** Implement the ARIMA model to forecast the time series data. Participants should discuss the process of making the data stationary and choosing the appropriate parameters for the model.
- 📊 **LSTM:** Develop an LSTM model to capture the complex, long-term dependencies in the data. They should explain how the model is trained and evaluate its performance.

Model Comparison:

Compare the results of the three models, discussing the strengths and weaknesses of each in the context of the data they have chosen.

Insights and Recommendations:

Based on the models' forecasts, you should provide actionable insights for spaza shop owners. This could include advice on when to stock up on certain products, how to prepare for seasonal demand, or how to adjust operations to maximize profits.

Expected Outcomes:

- 📊 A comprehensive report detailing the steps taken in data preparation, model implementation, and analysis.
- 📊 A Python notebook or code repository containing the implementations of Prophet, ARIMA, and LSTM models.
- 📊 A presentation that outlines the findings, compares the models, and provides recommendations for spaza shop owners.