Time Series Forecasting of Hospital Admissions

Chandana Bomminayuni
21ME31018

1. Introduction

Accurate forecasting of hospital admissions is crucial for effective resource allocation, staff scheduling, and inventory management. Unforeseen surges in patient numbers can strain hospital capacity and compromise the quality of care. Conversely, over-resourcing can lead to unnecessary costs. This project addresses this challenge by developing and comparing several time series forecasting models to predict hospital admission rates, providing a valuable tool for strategic planning.

2. Objectives

The primary objectives of this self-project were to:

- **Clean and prepare** a real-world dataset of hospital admission records for time series analysis.
- **Develop and implement** multiple forecasting models, including an Auto-Regressive Integrated Moving Average (ARIMA) model, a Prophet model, and a Hybrid model.
- **Evaluate the performance** of each model using standard metrics, specifically Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
- **Visualize the results** on a dashboard to provide a clear comparison of the actual versus forecasted admissions and the performance metrics of each model.

3. Scope of Works

This project focused on analyzing a dataset of over 15,000 hospital admission records. The scope of work included:

- **Data Preprocessing**: Cleaning the dataset by handling missing values, standardizing data types, and removing duplicate or invalid records. This involved normalizing date formats and addressing extreme outliers in DURATION_OF_STAY.
- **Time Series Creation**: Aggregating the raw admission data into daily and weekly admission counts to create time series datasets suitable for forecasting.
- **Model Development**: Building three distinct time series models:
    - **ARIMA**: An established statistical method for time series forecasting.
    - **Prophet**: A robust, open-source forecasting library developed by Facebook that excels at handling seasonality and holidays.
    - **Hybrid Model**: A simple ensemble approach that averages the forecasts from the ARIMA and Prophet models.

- **Model Evaluation**: Dividing the time series data into a training set (80%) and a testing set (20%) and evaluating each model's predictive accuracy on the unseen test data using MAE and RMSE.
- **Reporting**: Creating a concise summary and dashboard to present the results and a side-byside comparison of the model performance.

4. Methods Used

The project followed a standard data science workflow. The first step was to load the raw data, which contained 15,757 records and 56 columns.

Data Cleaning and Preprocessing:

- The D.O.A (Date of Admission) column was parsed to a standardized datetime format.
- A new DURATION_OF_STAY column was calculated from D.O.A and D.O.D (Date of Discharge).
- Invalid records, such as those with negative or excessively long stay durations, were removed.
- Columns with categorical data, such as GENDER, RURAL, and TYPE OF ADMISSION, were standardized.
- Missing numerical values were imputed using the median, and missing categorical values were filled with the mode.
- Duplicate entries were removed, resulting in a final dataset of 7,588 unique records.

Time Series Preparation:

- The cleaned data was aggregated by D.O.A to create a time series of daily hospital admissions.
- This daily series was then resampled to a weekly frequency, which smoothed out day-to-day noise and was used for the main forecasting task.
- The weekly time series was split into an 80% training set and a 20% testing set.

Model Implementation:

- **ARIMA Model**: An automated ARIMA model (pm.auto_arima) was used to automatically select the optimal parameters (p, d, q) for the training data.
- **Prophet Model**: The Prophet model was configured and trained on the same training data. Prophet is specifically designed for business time series and automatically detects trends and seasonality.
- **Hybrid Model**: A simple hybrid approach was created by taking the average of the weekly forecasts generated by the ARIMA and Prophet models.

5. Results

The forecasting models were evaluated on the test dataset using MAE and RMSE. A comparison of the performance is summarized in the table below:

Model MAE RMSE

ARIMA 19.29 20.46 **Prophet**

14.43 17.36

Hybrid 15.45 17.36

As shown in the table, the **Prophet model achieved the best performance** with the lowest MAE of 14.43 and an RMSE of 17.36

The dashboard visualization clearly illustrates these results, showing how the different models' forecast lines track the actual admissions data. The dashboard also includes a visualization of the error metrics for each model over time.

6. Discussion

The superior performance of the Prophet model suggests that the hospital admissions data likely contains a strong underlying trend and seasonal patterns. Prophet's architecture, which is built to explicitly model these components, gave it a significant advantage over the traditional ARIMA model, which relies on a more statistical approach to capture these patterns.

The interactive Power BI dashboard effectively serves as the final output, providing a user-friendly interface for hospital administrators to view and compare the forecasts. The ability to see the actual versus forecasted values, along with the error metrics, allows for informed decision-making regarding resource planning.

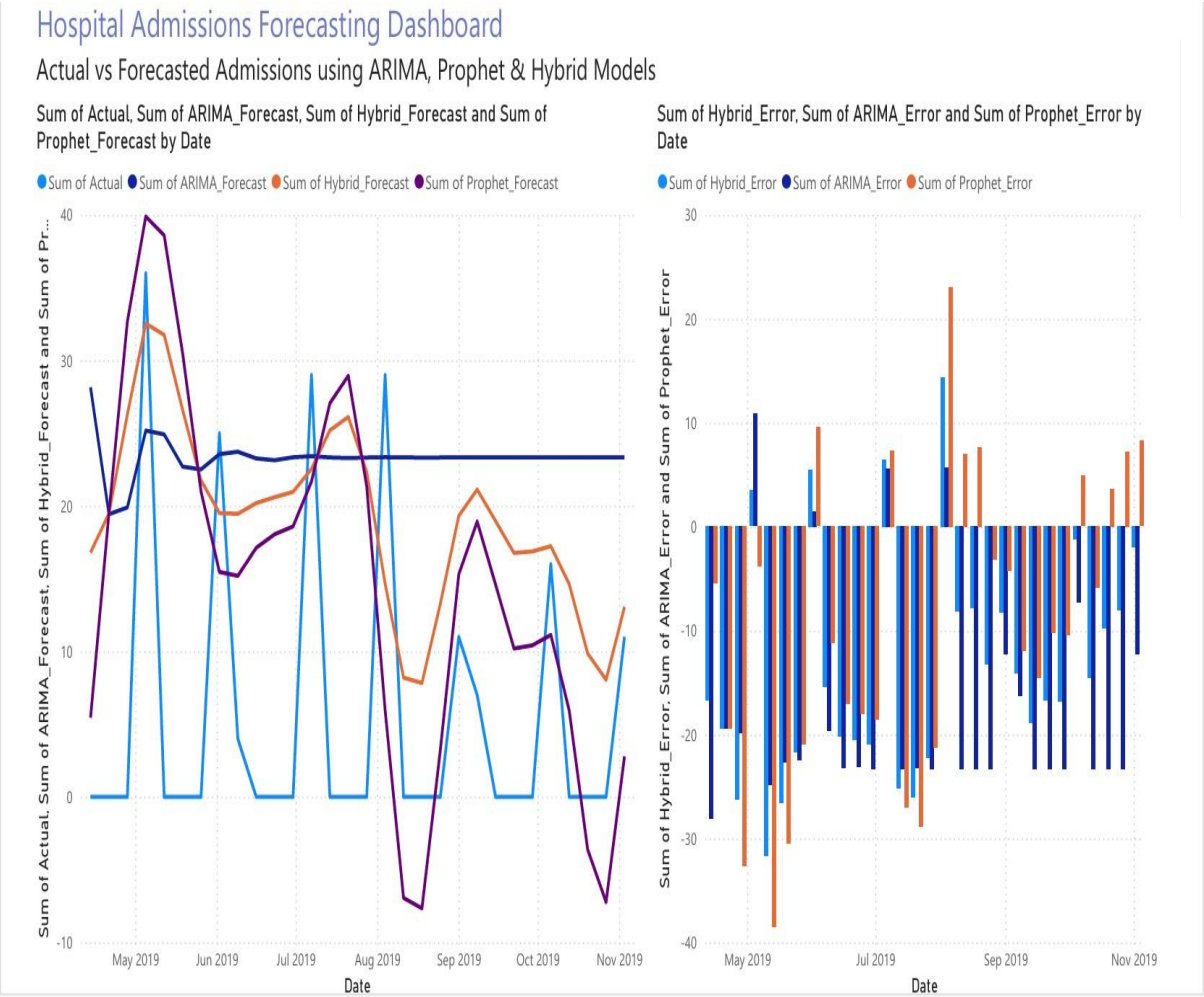
7. Conclusion

This project successfully developed a time series forecasting solution for hospital admissions. By applying data cleaning techniques and comparing three different models, the Prophet model was identified as the most accurate predictor of weekly admissions, achieving the lowest MAE and RMSE. The findings confirm that machine learning-based forecasting can be a powerful tool for healthcare management and can contribute to more efficient and cost-effective operations.

8. References

Time-series forecasting of seasonal items sales using machine learning – A comparative analysis

Link: https://www.sciencedirect.com/science/article/pii/S2667096822000027



Power BI Report

Declaration by the Student

I, K Nikhila, hereby declare that the work presented in this report is an authentic record of my own self-initiated project. This work has been completed in accordance with the guidelines provided for a self-study project. The material has not been submitted to any other university or institution for any academic purpose.

Date: K Nikhila

Place: IIT Kharagpur 21CE31007