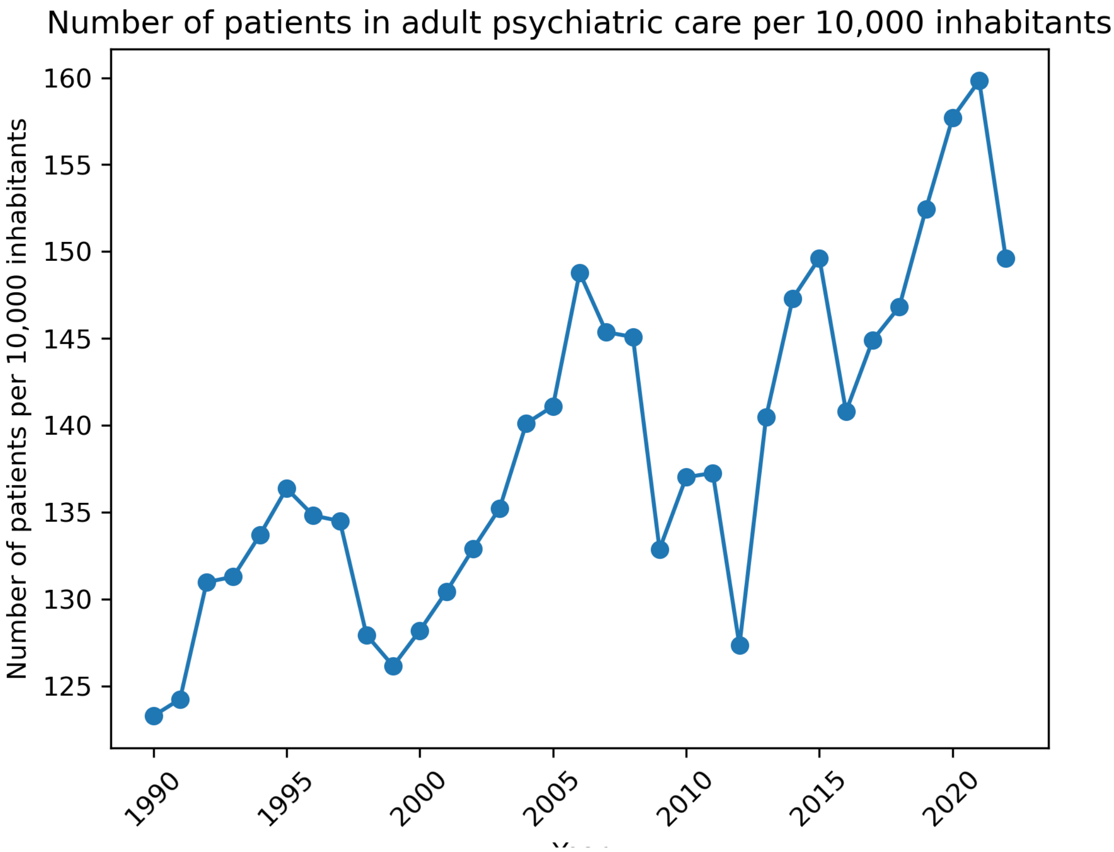
Final Project of Specialized Models: Time Series and Survival Analysis

Describing the Data

The data analyzed originates from the Hungarian Central Statistical Office (KSH), accessible through the link: https://www.ksh.hu/stadat_files/ege/hu/ege0027.html. Specifically, I examine the statistics related to the number of patients in adult psychiatric care per 10,000 inhabitants. This data is crucial as it sheds light on the mental health landscape within the adult population and provides insights into the demand for psychiatric care services. Understanding the prevalence of psychiatric patients per 10,000 inhabitants offers valuable information for policymakers, healthcare professionals, and researchers to gauge the effectiveness of mental health interventions, allocate resources appropriately, and identify potential areas for improvement in mental healthcare services. Moreover, by analyzing this data, we can assess trends over time, detect any significant changes or disparities, and develop strategies to address mental health challenges effectively. Ultimately, prioritizing mental health initiatives and ensuring adequate support for psychiatric care services is paramount for promoting overall well-being and quality of life within society.



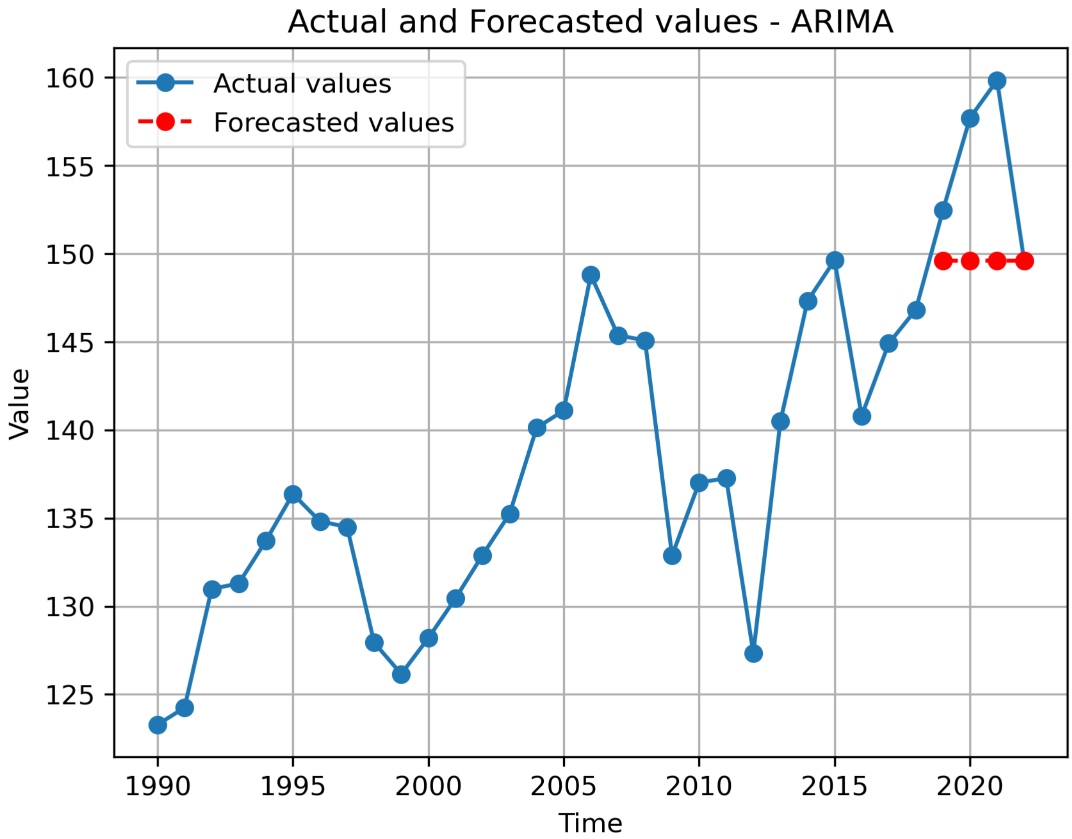
The chart displays annual data depicting the number of patients in adult psychiatric care per 10,000 inhabitants. It clearly illustrates an upward trend over the years, accompanied by some outlier values, indicating that the dataset is non-stationary. The ADF statistic value is -1.91, and the p-value is 0.33. Since the p-value is greater than the commonly accepted threshold of 0.05, we do not have sufficient evidence to reject the null hypothesis, indicating that the time series is non-stationary.

The Main Objectives of this Analysis

The aim of the analysis is to thoroughly examine the time series data regarding the number of patients in adult psychiatric care per 10,000 inhabitants in Hungary. This analysis serves several purposes. Firstly, we seek to understand the frequency and distribution of psychiatric disorders over time, exploring how the number of patients in care varies across different periods. Additionally, we aim to identify any temporal trends or seasonal patterns in the number of patients, which could be leveraged for future forecasting. Furthermore, the analysis aims to determine the most accurate methods for making predictions based on the time series. To achieve this, we compare the methods of ARIMA modeling, Facebook Prophet, and Neural Networks to ascertain which approach provides the best forecasting accuracy and under what time horizons reliable results can be achieved. The ultimate goal is to utilize these forecasts to assist in the more efficient allocation of resources for psychiatric care and in making better-informed decisions to meet the healthcare needs of patients.

Time Series Model

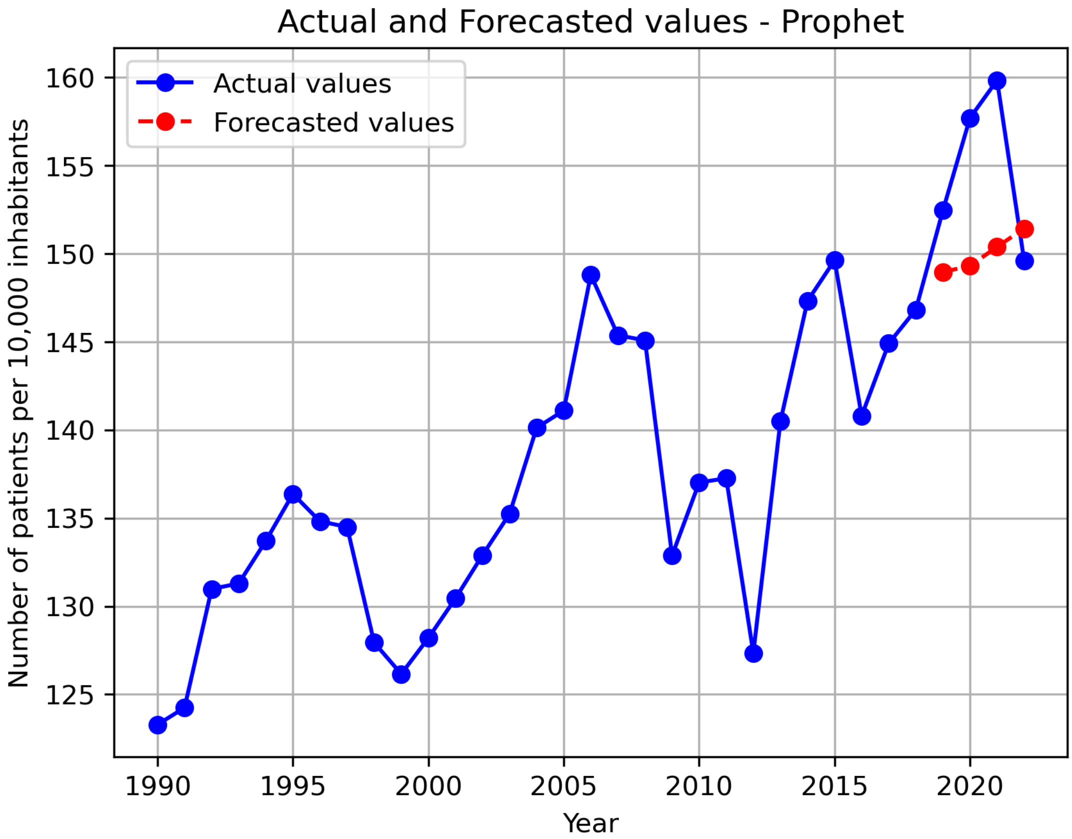
The first time series model fitted was an ARIMA model, which was optimized using the AIC criterion, resulting in an RMSE of 6.68. This indicates that the model's forecasted values deviate from the actual values by approximately 6.68 units on average. Evaluating the model's performance through AIC helped in achieving this level of RMSE.



The plot visualizes the entire time series dataset alongside the forecasted values, where the forecasted values are depicted in red, providing a clear comparison between the actual and predicted trends over time.

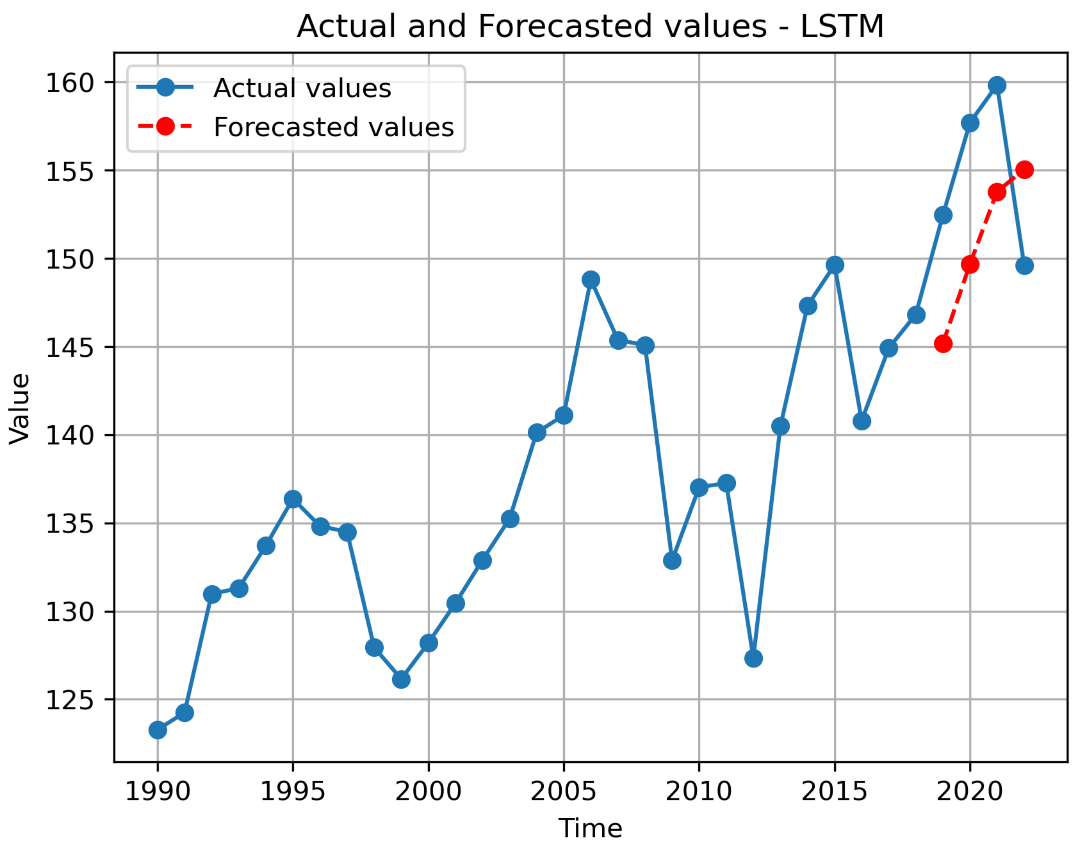
Facebook Prophet

The Prophet model achieved an RMSE of 6.62, which is lower than the RMSE obtained with the ARIMA model. This suggests that the Prophet model provides a better fit to the data and produces more accurate forecasts for the examined time series. Thus, the Prophet model represents a superior alternative for modeling and forecasting the changes in the number of patients in adult psychiatric care.



Neural Network

In the latest approach to time series analysis, an LSTM neural network was utilized, yielding an RMSE of 6.7709917516410485. This demonstrates the model's ability to make reasonably accurate predictions based on historical data. The LSTM network's capability to capture complex temporal patterns contributed to its success in forecasting future values within an acceptable margin of error.



Comparison of Models

Based on the RMSE values obtained, the LSTM model yielded an RMSE of 6.7709917516410485, while the Prophet model achieved an RMSE of 6.61850702951721. Meanwhile, the ARIMA model resulted in an RMSE of 6.676864177958071. Among these models, the LSTM model exhibited the highest RMSE, indicating that it was the least accurate in predicting future values. Conversely, the Prophet model demonstrated the lowest RMSE, suggesting it performed the best in capturing the underlying patterns in the time series data. Therefore, to improve forecasting accuracy, it might be beneficial to focus on refining the LSTM model or exploring further optimization techniques.

Future Work

To enhance and refine the model, several avenues could be explored. Firstly, incorporating additional relevant features into the model could improve its predictive capabilities. These features might include external factors such as economic indicators, weather data, or social events that could influence the time series behavior. Furthermore, fine-tuning the model architecture and hyperparameters, such as adjusting the number of LSTM units, layers, or optimizing the learning rate, could lead to better performance. Additionally, experimenting with different sequence lengths or window sizes for input data could capture more intricate patterns in the time series. Lastly, employing advanced techniques such as attention mechanisms or ensemble learning methods could also contribute to refining the model's accuracy and robustness in forecasting.

GitHub

https://github.com/domokla/coursera_timeseries