Wolt Internship

Prashant Shrestha

January 2024

1 Introduction

For the Data Science Summer Trainee assignment in 2024, my focus revolves around the orders_autumn dataset. The primary objective is to delve into the intricacies of predicting the order flow, specifically for the next day and week. This exploration aims to unravel valuable insights and pave the way for effective forecasting strategies.

2 Data Preprocessing and Exploration

The dataset comprises 18,706 observations across 13 columns. The dataset presents a challenge with 277 missing values in the CLOUD COVERAGE, TEMPERATURE, and WIND SPEED columns. The gaps are imputed with the mean values of their respective columns, ensuring a representative imputation.

The TIMESTAMP column is a important feature, leveraged for extracting temporal insights. This column is used to extract day, month, week progression, time of day and day of the week information, unlocking the potential to reveal intricate patterns within the data.

This strategic approach serves a dual purpose – resolving missing data concerns and enabling exploration of temporal patterns. By using mean imputation, we maintain the integrity of the dataset, while the TIMESTAMP transformations offer a foundation for comprehensive temporal analysis. This strategy aims to enhance the dataset's quality and harness its full potential in revealing meaningful patterns for subsequent predictive modeling.

3 Modelling Approach

In this project I have used ARIMA method for predicting the orders. It stands for Autoregressive Integrated Moving Average, is a powerful form of regression analysis designed for time series forecasting. The primary goal of the ARIMA model is to predict future movements in a time series by examining the differences between values in the series rather than the actual values.

I have chosen ARIMA model because its effectiveness, especially when dealing with a relatively small dataset. ARIMA is a univariate predictive model, meaning it does not require explanatory variables from external sources. This characteristic makes it particularly suitable for the given context, where the goal is to predict order counts based on past patterns within the dataset. Additionally, it also helps to capture seasonal patterns inherent in time series data. The model incorporates components for autoregressive (AR), differencing (I), and moving average (MA), making it well-suited for handling both short-term fluctuations and long-term trends.

The initial steps involved creating a dataset representing the mean daily order counts. A stationary test was conducted, revealing a p-value of 0.67, indicating the need for differencing in the model. Additionally, the data exhibited seasonality. The dataset was split into training and test sets. To determine the appropriate order for the model, the Minimum AIC score metric was used. The model was then trained on the training data. For evaluation, the model's performance was assessed using the testing data, employing the Root Mean Squared Error (RMSE) metric. Figures 1 and 2 depict the actual and predicted values of daily and weekly orders in the test data. Furthermore, they show that the model captures the trend in the dataset.

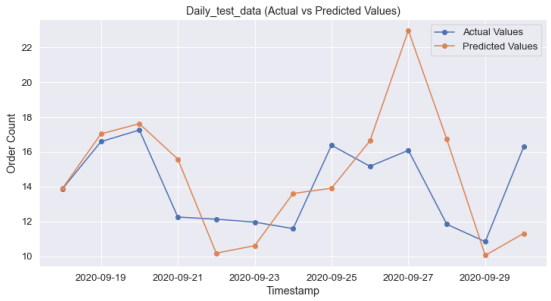

Figure 1: Actual vs predicted mean values of daily orders

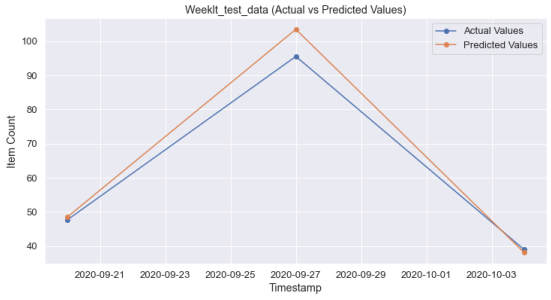

Figure 2: Actual vs predicted mean values of weekly orders

4 Results

A trend emerges in the dataset, indicating that weekends consistently have higher order volumes compared to weekdays. Friday and Wednesday has a more order activity compared to other weekdays. This pattern depicts the significance of considering both daily and weekly variations in order.

The high order counts on Fridays and Wednesday points to potential correlations with specific consumer behaviors or market trends on these particular days. Understanding and leveraging these insights can not only enhance operational planning but also provide a strategic advantage in optimizing resources and marketing efforts. By integrating this understanding of weekly patterns, the predictive model can be fine-tuned to capture and capitalize on the distinct dynamics related with different weekdays. Figure 3 depicts the distribution of order in different week days.

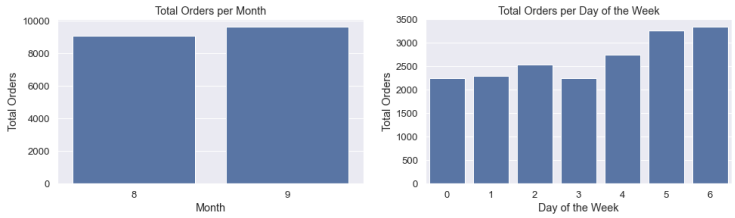

Figure 3: Distribution of order in Week

The temporal distribution of orders exhibits a distinctive pattern, with the highest order volumes concentrated in the afternoon, particularly at 3 and 2 pm. Morning and evening periods also demonstrate substantial order activity, albeit to a slightly lesser extent. In contrast, the night experiences the lowest number of orders.

This temporal analysis not only highlights the peak order hours but also provides a understanding of daily fluctuations. By pinpointing specific time frames with heightened order activity, this insight can inform strategic decision-making and resource allocation to optimize operational efficiency during peak periods. The identified order trends contribute valuable context for refining predictive models and aligning business strategies with the observed temporal dynamics. Figure 4 depicts the distribution of order in different hour and time of the day.

5 Conclusion

Our dataset uncovered valuable insights into the temporal patterns of order counts, highlighting distinct variations across hours, days of the week, and week progression. Moreover, afternoons and weekdays emerged as peak times for order activity.

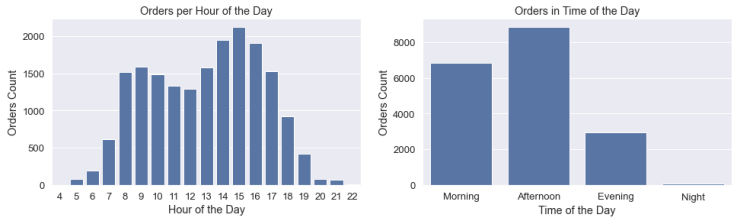

Figure 4: Distribution of order hours

The implemented ARIMA model was effective in capturing the patterns of order counts in the dataset. The model's performance was evaluated using Root Mean Squared Error (RMSE), resulting in satisfactory metrics of 3.11 for daily predictions and 4.62 for weekly predictions.

In summary, the ARIMA model serves as a robust tool for predicting order counts at different time intervals. This project provides a solid foundation for future enhancements and exploration of advanced forecasting techniques. The ultimate goal is to improve predictive accuracy and facilitate informed decision-making in order flow management at Wolt.

6 My Background and Wolt

Through my academic journey and practical work, I've developed a keen interest in leveraging data analytics and natural language processing (NLP) to derive meaningful insights and solve real-world problems. The challenges I find most intriguing revolve around the intersection of technology and language, particularly in the field of NLP.

I am currently in the process of identifying a thesis topic, and I am excited about the prospect of aligning it with the needs and goals of forward-thinking companies like Wolt. While I don't have specific prior coursework directly related to the challenges faced by Wolt, my academic background equips me with a solid foundation in analytics, which I believe is transferrable and adaptable to various domains.

Having followed Wolt's growth and impact, I am drawn to the challenges in optimizing and enhancing user experiences, particularly in the context of food delivery and logistics. I am eager to apply my skills in data analysis, predictive modeling, and NLP to contribute to the optimization of Wolt's operations and customer interactions.

In terms of relevant minors and side projects, I have actively engaged in coursework and projects that involve NLP, sentiment analysis, and the application of machine learning algorithms to textual data. These experiences have not only enhanced my technical skills but have also fostered a deep appreciation for the practical implications of these technologies. Moreover, my GitHub repository helps to show my competence, showcasing not only my hands-on experience

but also my commitment to continuous learning and effective problem-solving capabilities. My projects include a Bike Rental project and a Recipe Selection project. These endeavors showcase my ability to tackle real-world challenges at Wolt, demonstrating a combination of technical proficiency and an understanding of practical applications.

In summary, my ambition is to bring a fresh perspective and innovative solutions to challenges at the intersection of data analytics, NLP, and user experience. I am excited about the opportunity to apply my knowledge and skills in a dynamic and impactful environment like Wolt, contributing to the ongoing success and innovation of the company.