Forecasting Daily Bike Rental Demand Using Time Series Models

Introduction

This report presents the analysis and modeling of daily bike rental demand using time series techniques in R. The goal of the project is to develop forecasting models that can accurately predict future bike rental demand based on historical data and relevant factors such as weather and seasonality.

Data Description

The dataset used for this analysis contains the daily count of rental bike transactions between 2011 and 2012 in the Capital bikeshare system, along with corresponding weather and seasonal information. The data was sourced from the UCI Machine Learning Repository.

Data Exploration

To gain insights into the data, various exploratory analyses were performed:

1. Correlation Analysis: A positive correlation (around 0.63) was observed between temperature variables (temp and atemp) and the bike rental count (cnt), indicating that higher temperatures are associated with increased bike rentals.

2. Seasonal Temperature Analysis: The mean and median temperatures varied across different seasons, with higher temperatures in summer (Season 4) and lower temperatures in winter (Season 1).

3. Rental Type Analysis: Both casual and registered bike rentals showed an increasing trend with higher temperatures, suggesting that temperature is a significant factor influencing bike rental demand.

4. Monthly Analysis: The mean temperature, humidity, wind speed, and total rentals were calculated for each month to identify potential patterns.

Time Series Analysis

Interactive time series plots were created using the timetk package to visualize the bike rental demand over time. These plots revealed clear weekly, monthly, and quarterly patterns, with higher demand observed during warmer months and on weekends.

Data Smoothing and Stationarity

To smooth the time series data and reduce noise, techniques such as exponential smoothing and simple moving averages were employed. The time series was then decomposed into trend, seasonal, and random components, highlighting the presence of non-stationarity.

The Augmented Dickey-Fuller (ADF) test was performed on the deseasonalized data, confirming non-stationarity. To achieve stationarity, differencing was applied to the time series.

ARIMA Modeling and Forecasting

Both manually specified and automatically selected ARIMA (Autoregressive Integrated Moving Average) models were fitted to the bike rental data:

1. Manually Specified ARIMA(1,1,1) Model:
   - The residuals from this model showed some deviation from normality, indicating potential room for improvement.

2. Automatically Selected ARIMA(0,0,1) Model with Non-Zero Mean:
   - The auto.arima function selected this model, which appeared to have a slightly better fit based on the error measures.

Forecasts for future bike rental demand were generated using both ARIMA models, enabling proactive planning and resource allocation.

Findings and Conclusions

The analysis highlighted the following key findings and conclusions:

1. Temperature and seasonality play crucial roles in predicting bike rental demand, with higher demand observed during warmer months and on weekends.

2. Smoothing techniques and time series decomposition helped identify underlying patterns and components in the data.

3. Both manually specified and automatically selected ARIMA models provided forecasting capabilities for future bike rental demand, with the automatically selected ARIMA(0,0,1) model showing a slightly better fit.

4. While the developed models offer valuable insights and forecasting capabilities, considering other relevant factors such as weather conditions, special events, and infrastructure changes could further improve the accuracy of the forecasts.

Recommendations

Based on the findings and conclusions, the following recommendations are made:

1. Incorporate additional relevant variables, such as precipitation, events, and infrastructure changes, into the forecasting models to capture a more comprehensive picture of the factors influencing bike rental demand.

2. Regularly update the models with new data to ensure accurate forecasts and adapt to changing patterns and trends.

3. Explore alternative time series modeling techniques, such as exponential smoothing or neural network-based approaches, and compare their performance with the ARIMA models.

4. Develop a user-friendly interface or dashboard to visualize the forecasts and facilitate decision-making for resource allocation and operational planning in the bike rental business.

By implementing these recommendations, the bike rental company can enhance its forecasting capabilities, optimize resource utilization, and provide a better service to its customers.