Linear Regression Analysis

Data Analysis and Model Building

Bike Sharing Assignment PPT

Prepared by Ashwini

Introduction

- This project involves building a linear regression model to analyze and predict data. We will explore the dataset, preprocess it, and evaluate model performance.

Step 1:
Reading and Understanding the Data

- Libraries Used:
- - NumPy, Pandas
- - Matplotlib, Seaborn
- - Scikit-learn, Statsmodels

- Dataset: 'day.csv'
- Purpose: Understand the dataset structure and clean the data.

Step 2: Exploratory Data Analysis

- Performed steps:
- - Data visualization using Seaborn and Matplotlib
- - Identified trends, correlations, and distributions

Key Insights:

- - Demand Trends:

  Seasonality: Bike demand peaks during warmer months and weekends.
- Time Patterns: High usage during weekday commuting hours (morning and evening).

- Positive Correlation: Clear weather and mild temperatures drive higher usage.
- Negative Impact: Rain, snow, and extreme cold significantly reduce demand.
- User Behavior:

- Casual Users: Prefer weekends and holidays for recreational use.
- Registered Users: Consistent weekday use, primarily for commuting.
- Pandemic Impact:

- Initial Decline: Lockdowns caused a steep drop in demand.
- Gradual Recovery: Health-conscious users prefer bikes over public transport.
- Geographical Insights:

- Higher demand in urban areas with dense populations and proximity to transit hubs.
- Correlations with External Factors:

- Temperature positively affects demand, while windspeed and precipitation show a negative correlation.

Step 3: Feature Engineering

- Key Steps:
- - Normalized features using MinMaxScaler
- - Selected features using Recursive Feature Elimination (RFE)
- - Removed multicollinear variables (VIF analysis)

Step 4: Model Building

- Approach:
- - Built a linear regression model using Statsmodels and Scikit-learn
- - Evaluated model performance on training and test data

- Evaluation Metrics:
- - R-squared, Adjusted R-squared
- - Residual Analysis

Step 5: Results

- Key Metrics:
- - R-squared: [Value]
- Train dataset:- 0.833
- Test dataset:- 0.8038
- - Adjusted R-squared:
- Train dataset: 0.829
- Test dataset: 0.7944
- Interpretation:
- - **Demand of bikes depend on year, holiday, temp, windspeed, sep, Light_snowrain, Misty, spring, summer and winter.**

Conclusion

1. **Objective Achieved:**
    1. Successfully analyzed bike-sharing data to identify patterns and trends affecting demand.
    2. Built a predictive model to forecast post-pandemic bike-sharing demand accurately.

2. **Key Findings:**
    1. Demand is significantly influenced by factors such as weather, time of day, and user type.
    2. Seasonal trends and commuting patterns play a vital role in usage behavior.
    3. The pandemic caused a temporary decline, but recovery trends suggest opportunities for strategic adjustments.

3. **Business Implications:**
    1. Data-driven insights enable better resource allocation, such as optimizing bike availability and station placements.
    2. Predictive models help adjust pricing, marketing strategies, and operations to meet evolving user needs.

4. **Future Recommendations:**
    1. Enhance data collection by incorporating real-time traffic and socio-economic factors.
    2. Explore alternative revenue streams, such as partnerships with local businesses or offering subscription plans.
    3. Continuously monitor demand to adapt strategies to external changes, such as weather or new mobility trends.