

Seoul Bike Rental Prediction: A Machine Learning Approach

1. Introduction

This project aims to predict bike rental demand in Seoul using machine learning to optimize bike-sharing operations and improve user experience. We use a public dataset with weather, time, and holiday data.

2. Methodology

- Data Preprocessing: Handling missing values, encoding categorical variables, scaling features, and addressing outliers.
- EDA: Analyzing data distributions, correlations, and trends.
- Feature Engineering: Creating new features like time categories, weekend/weekday, and interaction terms.
- Model Selection: Evaluating Linear Regression, Random Forest, XGBoost, and SVR using R-squared and RMSE.
- Hyperparameter Tuning: Optimizing the best model using GridSearchCV.

3. Results and Recommendations

- XGBoost was the most accurate model (R-squared: 0.866, RMSE: 58,223).
- Random Forest also performed well (R-squared: 0.852, RMSE: 61,834).
- Important features: temperature, hour, season, humidity, solar radiation.

4. Recommendations for Bike-Sharing Operators

- i. Focus on areas with high predicted rentals during peak hours and warmer months.
- ii. Adjust pricing based on real-time demand and weather.
- iii. Proactively manage fleet size.

5. Conclusion

This project demonstrates the potential of machine learning for predicting bike rental demand, helping to optimize bike-sharing programs and improve urban mobility.