Bike Sharing **Prediction**

Elia Samuel
12/01/2025

Agenda

01
Business Problem

02
Data Understanding

03
Data Preprocessing

04
Modeling

05
Conclusion

06
Recommendation

01

Business **Problem**

Context
Goals
Success Criteria

Understanding the Challenge in **Bike Sharing Demand Prediction**

Context:

- Bike-sharing systems are a new generation of automated rental services, operating in over 500 cities worldwide.
- These systems provide rich data that can act as virtual sensors to monitor urban mobility patterns.

Key Challenges:

- Fluctuating demand due to time, weather, and mobility patterns.
- Ensuring optimal availability of bikes at various stations to meet customer needs efficiently.
- The operational cost of redistributing bikes between stations to meet demand.

Critical Need:

- A **reliable and accurate predictive model** to predict bike rentals under varying conditions.
- Insights into factors affecting demand to optimize operations and improve customer satisfaction.

Proposed Predictive Model for **Bike Sharing**

Approach:

- Build a machine learning model using historical data to predict bike rental demand.

Evaluation Metrics:

- **RMSE (Root Mean Squared Error)**: Measures the average magnitude of prediction errors.
- **MAE (Mean Absolute Error)**: Evaluates the average absolute error between actual and predicted values.
- **MAPE (Mean Absolute Percentage Error)**: Represents prediction error as a percentage of actual values.
- **R2**: indicates how effectively the model explains price variations, supporting more informed and strategic decision-making for the company.

Outcome:

- A model that delivers accurate, actionable insights for bike-sharing companies.

Goals and Business Value

Primary Goals:

- Develop an **accurate prediction model** for bike rentals using historical and external data.
- Identify key factors influencing rental demand (e.g., time of day, weather).
- Optimize bike distribution to ensure customer satisfaction and reduce operational costs.

Expected Outcomes:

- **Enhanced Customer Experience**: Ensure bikes are available at the right time and place.
- **Operational Efficiency**: Reduce costs associated with manual bike redistribution.
- **Business Growth**: Improve customer loyalty and rental volumes, boosting revenue.

Strategic Impact:

- Position the company as a leader in bike-sharing services through data-driven decision-making.

02

Data **Understanding**

Overview **Bike Sharing**

Data Source

Capital Bikeshare System Data.
https://capitalbikeshare.com/system-data

Time Period

January 1, 2011 - December 31, 2012

Data Size

12165 rows, 11 columns

Feature **Bike Sharing**

Feature	Description	Impact to Business
dteday	date	track rental trends over time
season	season(1:winter, 2:spring, 3:summer, 4:fall)	Seasonal demand variation
weathersit	1: [Clear, Few clouds, Partly cloudy, Partly cloudy], 2: [Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist], 3: [Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds], 4: [Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog]	Direct weather impact on demand
holiday	0: Holiday, 1: Not Holiday	Captures impact of holidays on rentals
temp	normalized temperature in Celsius. The values are derived via (t-tmin)/(tmax-tmin), tmin=-8,t_max+39 (only in hourly scale)	Weather influence on rentals

Feature **Bike Sharing**

Feature	Description	Impact to Business
atemp	Normalized feeling temperature in Celsius. The values are derived via (t-tmin)/(tmax-tmin),tmin=-16, t_max=+50 (only in hourly scale)	Comfort level impact on user behavior
hr	Hour (0-23)	Key understanding to peak rental hours
hum	normalized humidity. The values are divided into 100 (max)	Weather conditions affecting rentals
casual	count of casual users	Insights into non-registered user behavior
registered	count of registered users	Insights into loyal customer base
count	count of total rental bikes including both casual and registered	Overall rental business performance

03

Data **Preprocessing**

Data Cleaning

	Method	Result
Missing Value	Drop Missing Value	No Missing Value
Duplicate Data	Drop Duplicate Data	1 rows dropped
Identify Outlier	Check Distribution of each feature	Data is non-normal distribution
New Features	Extract New Feature from Feature ‘dteday’	New features: day, is_weekend, month, year, date

Feature Engineering

	Method	Feature
Data Categorical	One Hot Encoding	Month, hour, day, date, weather, year
	Cyclical Encoding	Month, hour, day, date
Data Numerical	Scaling	Month, hour, day, date, year
	KBinsDiscretizer	Temp, humidity

04

Modelling

Benchmark

Transformed Target

Linear Regression
Ridge
Lasso
DTree Regressor
Random Forest
Gradient Boosting
XGB Regressor

List Model

Transformer
Binning, One Hot , Robust, Drop

Transformer_cyc
Binning, One Hot , Robust, cyclical_enc

RMSE, MAPE, MAE, R2

Benchmark Result

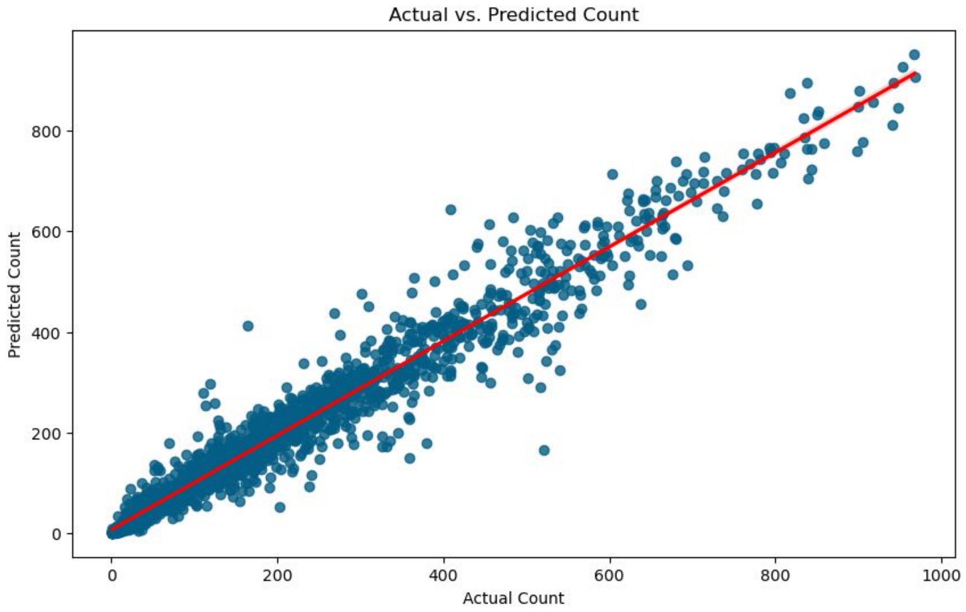
	Model	RMSE	MAE	MAPE	R2
Transformer	XGBRegressor	60	37	0.37	0.88
Transformer_cyc	XGBRegressor	46	27	0.25	0.93

Tuning

	RMSE	MAE	MAPE	R2
Before Tuning	44.713	27.119	0.235	0.945
After Tuning	40.866	25.182	0.218	0.954

```
colsample_bytree': 1.0,
gamma': 0,
learning_rate': 0.05,
max_depth': 6,
n_estimators': 600,
reg_alpha': 0.5,
subsample': 0.8}
```

Actual vs Predicted



Strong Model Fit:
The points are well-aligned with the diagonal red line, indicating that the model predicts values close to the actual values for most instances.

High Accuracy for Lower Values:
The clustering of points near the line for lower values suggests that the model performs very well in predicting small-to-moderate counts, with minimal errors in this range.

SHAP Analysis of Bike Rental Prediction Model

Critical Features:

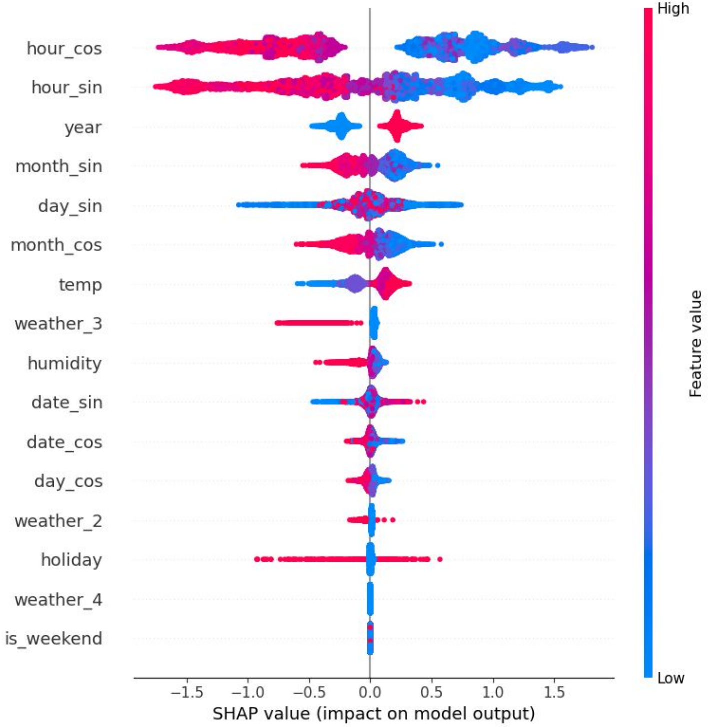
- **hour_cos/hour_sin**: Most impactful features, indicating bike rentals are strongly influenced by time of day.
- **Year**: Suggests an increasing or decreasing trend in rentals over time.
- **Month & Day Cycles**: Monthly and daily patterns play a significant role.

Weather Influence:

- Temperature positively impacts bike rentals.
- Specific weather conditions (e.g., weather_3, weather_2) reduce demand, reflecting sensitivity to adverse weather.

Seasonality Patterns:

- Features like holiday and is_weekend are less impactful, suggesting demand is less tied to specific days.



Implementation Model

Input Data → ***Model Prediction Process Data*** → ***Prediction Result***

	RMSE	MAE	MAPE	R2
Seen	31.355303	19.346380	0.161999	0.970427
Unseen	23.779914	14.616687	0.111591	0.981855

1. The improved performance on unseen data could be due to regularization techniques or robust feature selection, which prevent overfitting and enhance generalization.
2. The low MAPE on unseen data (11.16%) reflects strong reliability for real-world predictions, especially for forecasting demand trends.
3. This balance between seen and unseen data performance indicates that the model is both accurate and stable.

Implementation Model to Business

Underestimated Rentals

- Model underestimates demand by **23.78 users per day on average.**

Lost Revenue from Underestimation

- **Daily lost revenue:** RMSE × Rental price.
- **Calculation:** 23.78 users × $1/user = **$23.78/day.**

Annual Revenue Loss (Peak Days)

- Assuming 365 peak days per year:
  - **Annual loss:** $23.78/day × 365 days = **$8,681.70/year.**

05

Conclusion

Conclusions on Model Predictions

1. **High Predictive Accuracy**:
   The model achieves excellent performance with R^2 values of **0.970** (seen) and **0.982** (unseen), effectively explaining over **97%** of the variance in bike rental data, indicating strong pattern recognition and generalization.

2. **Reliable Error Metrics**:
   Low RMSE (31.36 seen, 23.78 unseen) and MAPE (16.2% seen, 11.2% unseen) confirm minimal and manageable prediction errors, making the model highly suitable for practical applications like demand forecasting and fleet optimization.

Conclusions (Business)

1. **Optimized Inventory Management**
   With the model's high accuracy and low error rates, the company can forecast bike rental demand more precisely. This enables better inventory planning, ensuring the right number of bikes is available at the right time, reducing overstocking or shortages, and optimizing operational costs.

2. **Enhanced Customer Experience and Revenue Growth**
   Accurate demand predictions allow the company to proactively meet customer needs, particularly during peak seasons or high-demand periods. This improves customer satisfaction by reducing wait times and ensuring availability, which can increase repeat business, attract new customers, and ultimately drive revenue growth.

06

Recommendation

Recommendation

1. **Optimize Resource Allocation**
   - **Leverage accurate demand forecasts ($R^2 > 0.95$)** to optimize bike distribution across stations.
   - **Dynamically adjust bike numbers** to match predicted demand at each location.
   - **Reduce idle bike costs** by minimizing excess supply during off-peak hours.
2. **Enhance Operational Efficiency**
   - **Continuously refine the model** to improve prediction accuracy (reduce RMSE & MAE).
   - **Integrate key factors** like local events and station location to enhance model robustness.
3. **Implement Dynamic Pricing**
   - **Mitigate prediction errors** by implementing a dynamic pricing strategy.
   - **Capitalize on high demand** by adjusting prices based on predicted demand.
   - **Encourage rentals during off-peak hours** through lower prices.

Thank **Ya!!!**

samuelelia0907@gmail.com

https://github.com/SamuelEl09/bike_sharing