BIKE SHARING DEMAND

Aadit Shukla
01-12-2022

Abstract

Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city.

Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Eventually, providing the city with a stable supply of rental bikes becomes a major concern, which will grow the business of bike sharing. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes, so it's a need of the hour to solve this problem. Bike-sharing rental process is highly correlated to the environmental and seasonal settings. For instance, weather conditions, precipitation, day of week, season, hour of the day, etc. can affect the rental behaviors. Therefore, the proposed model will Predict the demand for rental bikes given information about the weather and time of the day.

Aim: Combine historical usage patterns with weather data in order to forecast bike rental demand.

Technologies Used: Python, MS Excel, Tableau.

- I have targeted the problem as a 'Regression Analysis' and used 'Random Forest Regression' Machine Learning Algorithm to build my final model.
- For Data Pre-processing, I checked for missing values, outliers, and looked for Multicollinearity problems. I have used Label Encoder which converted categorical data into numeric form. For Feature Engineering, the Datetime column looks interesting so I extracted month, day, hour, weekday, date and year from it.
- Derives some insights from the data using data visualization.
- Before reaching my final model, I have built 'Random Forest Regressor', 'AdaBoostRegressor', 'Bagging Regressor', 'Kneighbors Regressor', however 'Random Forest Regressor' gave a bit better result as compared.

1. Introduction

A bicycle-sharing system, bike share program, public bicycle scheme or public bike share (PBS) scheme is a shared transport service in which bicycles are made available for shared use to individuals on a short-term basis for a price or free. Many bike share systems allow people to borrow a bike from a "dock" and return it at another dock belonging to the same system. Docks are special bike racks that lock the bike, and only release it by computer control. The user enters payment information, and the computer unlocks a bike. The user returns the bike by placing it in the dock, which locks it in place.

In recent years, an increasing number of cities across the world have started to offer both mechanical bike share and electric bicycle sharing systems.

Below is a list of the people and organizations that we have specifically design our products and services for:

- Potential Couples/Young Adults
- Tourist
- Households / Families
- Schools (High Schools, Colleges and Universities)
- Sport Organizations (For cycling Competitions)
- Event Planners
- Tour Guides

One thing about a Bike sharing company is that you only require less effort to market your service, especially if your dockyard is located in an area that attracts visitors and tourists.

7 best scooter and bike sharing platforms in India:

- Ola pedal
- Yulu
- Zoomcar PEDL
- Letscycle
- Bounce
- Zypp
- Vog

2. Problem Statement

The proposed model will Predict the demand for rental bikes given information about the weather and time of the day. We are provided hourly rental data along with weather data. Here the training set comprises the first 20 days of each month, while the test set is the 21th to the end of the month. Here I have predicted the total count of bikes rented during each hour covered by the test set, using only information available prior to the rental period.
The problem statement is to combine historical usage patterns with weather data in order to forecast bike rental demand.

3. Data Description

A Short description of the Features:

- datetime - hourly date + timestamp
- season - 1 : spring, 2 : summer, 3 : fall, 4 : winter
- holiday - whether the day is considered a holiday
- workingday - whether the day is neither a weekend nor holiday
- weather -
    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 Pellets + Thunderstorm + Mist, Snow + Fog
- temp - temperature in Celsius
- atemp - "feels like" temperature in Celsius
- humidity - relative humidity
- windspeed - wind speed
- casual - number of non-registered user rentals initiated
- registered - number of registered user rentals initiated
- count - number of total rentals

```
Train_data.head()
```

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	3	13	16
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	8	32	40
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	5	27	32
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	3	10	13
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	0	1	1

```
Test_data.head()
```

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed
0	2011-01-20 00:00:00	1	0	1	1	10.66	11.365	56	26.0027
1	2011-01-20 01:00:00	1	0	1	1	10.66	13.635	56	0.0000
2	2011-01-20 02:00:00	1	0	1	1	10.66	13.635	56	0.0000
3	2011-01-20 03:00:00	1	0	1	1	10.66	12.880	56	11.0014
4	2011-01-20 04:00:00	1	0	1	1	10.66	12.880	56	11.0014

4. Market / Customer / Business Need Assessment

The importance of data produced by the bike on a daily basis and the data generated by these systems makes them attractive for researchers because the duration of travel, departure location, arrival location, and time elapsed is explicitly recorded. Bike sharing systems therefore function as a sensor network, which can be used for studying mobility in a city. How this data can be used by machine learning to tell the company about the availability of bikes in a particular place, so that they can arrange bikes to the public at the right time.

Other benefits of bike sharing schemes include

- Transport flexibility
- Financial savings for individuals
- Health benefits, Fitness
- Reductions to vehicle emissions, Traffic
- Reduced congestion and Fuel consumption
- Environmental benefit (Reduced noise and air pollution through decreased automobile usage)

5. Target Specification and Characterization

5.1 Regression: Prediction of bike rental count hourly or daily based on the environmental and seasonal settings. Bike-sharing rental process is highly correlated to the environmental and seasonal settings. For instance, weather conditions, precipitation, day of week, season, hour of the day, etc. can affect the rental behaviors.

5.2 Event and Anomaly Detection: Count of rented bikes are also correlated to some events in the town which easily are traceable via search engines. Therefore, the data can be used for validation of anomaly or event detection algorithms as well.

6. External Search (Information Sources)

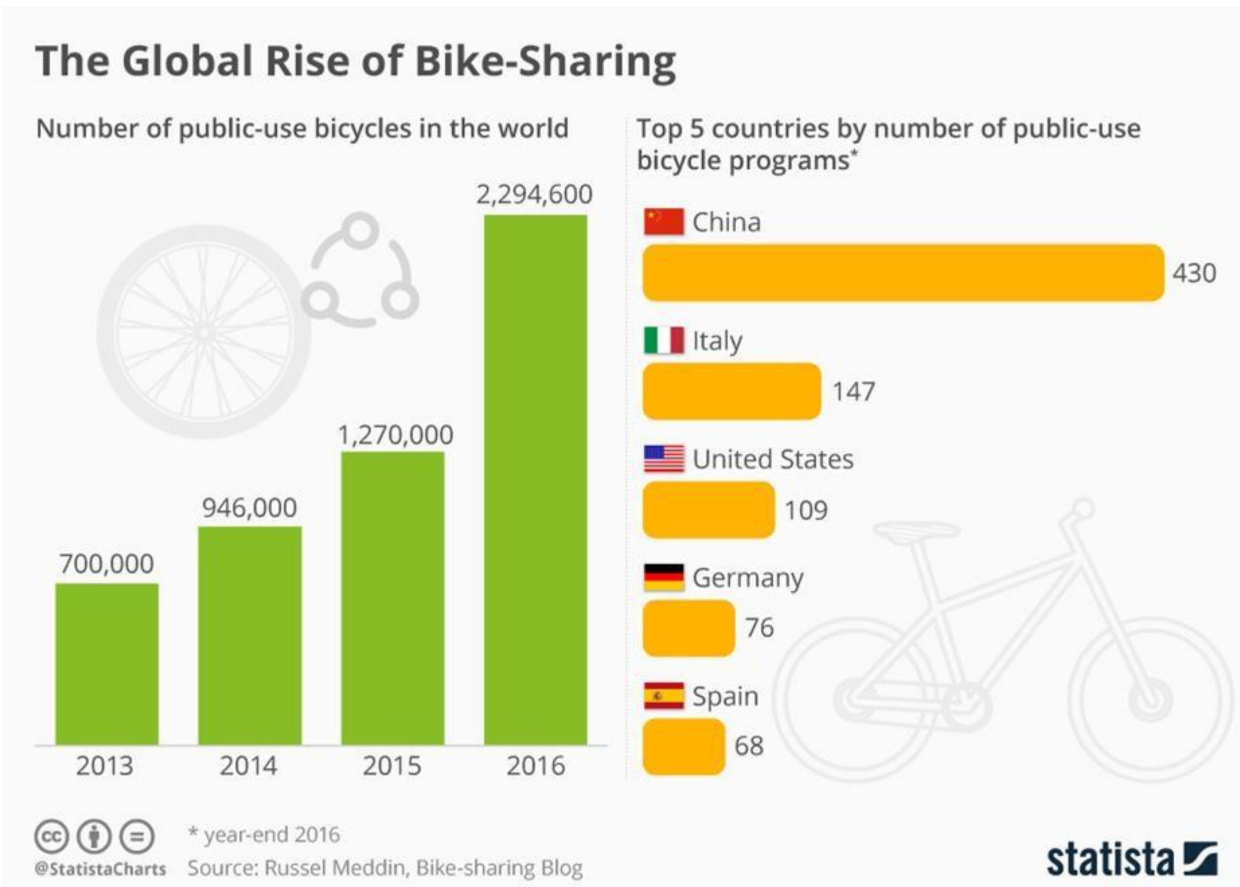
The sources I have used as reference for prediction of the demand for rental bikes given information about the weather and time of the day, have mentioned below:

- https://www.kaggle.com/c/bike-sharing-demand/data?select=train.csv
- https://www.kaggle.com/c/bike-sharing-demand/data?select=test.csv
- https://en.wikipedia.org/wiki/Bicycle-sharing_system

This dataset was provided by Hadi Fanaee Tork using data from Capital Bikeshare.

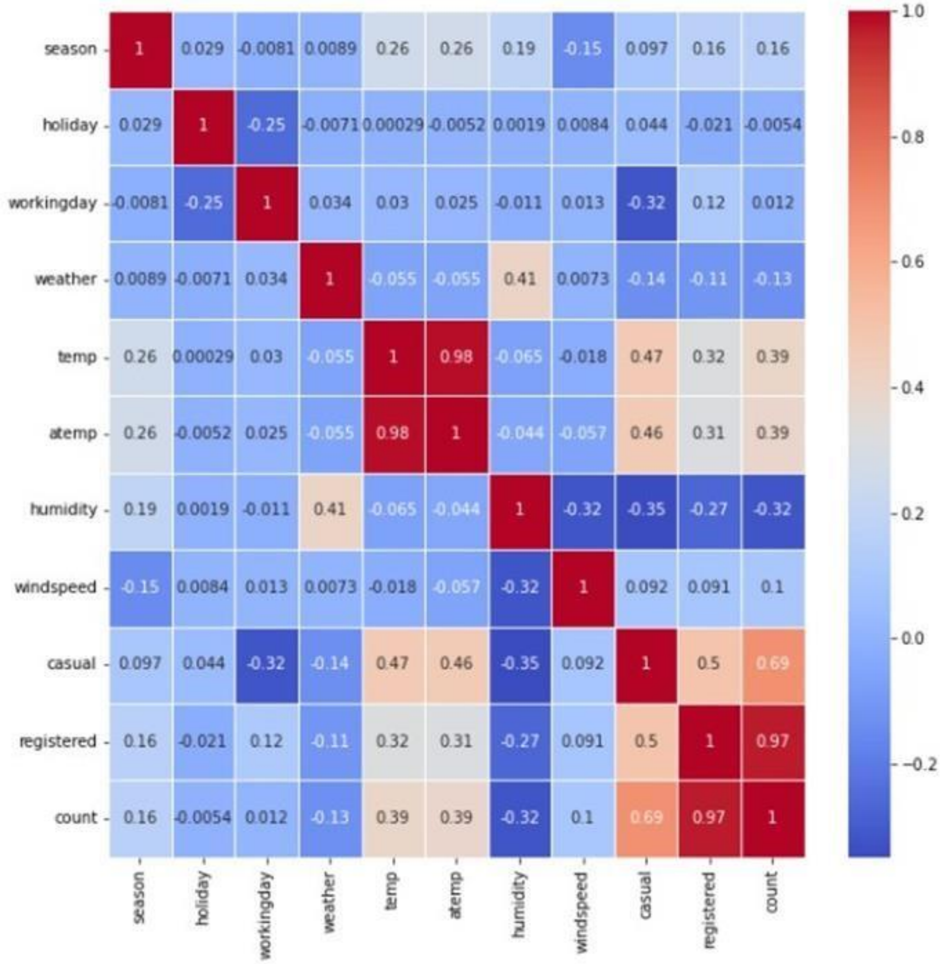
7. Benchmarking

Bike sharing companies like Yulu, Ola pedal, Letscycle have been using this type of model to grow their businesses. This technique would also be beneficial when applied to small businesses.



Correlation Matrix allows to have a global view of the more or less strong relationship between several variables LINEARLY.

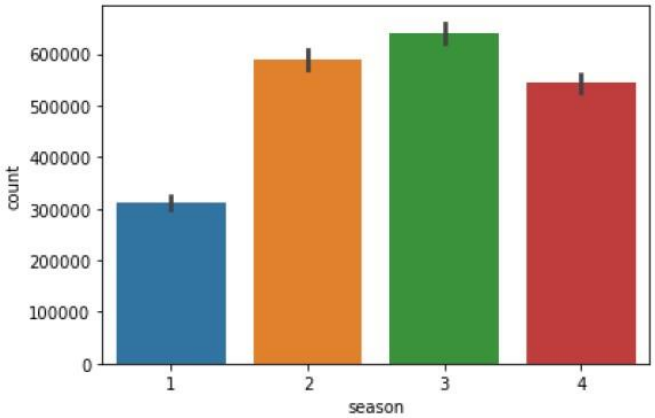
```
plt.figure(figsize=(15,15))
ax = sns.heatmap(Train_data.corr(), cmap = "coolwarm", annot=True, linewidth=1)
bottom, top = ax.get_ylim()
```



Inferences from the above heatmap

- self-relation i.e. of a feature to itself is equal to 1 as expected.
- temp and atemp are highly related as expected.
- humidity is negatively correlated to count as expected as the weather is humid people will not like to travel on a bike.
- Also note that temp (or atemp) affects the count.
- registered/casual and count are highly related which indicates that most of the bikes that are rented are registered.

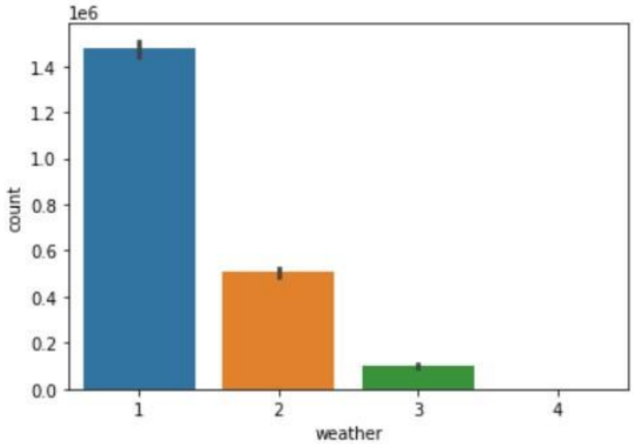
```
sns.barplot(x='season',y='count',data=Train_data,estimator = np.sum)
label=[ "Spring", "Summer", "Fall", "Winter"]
plt.show()
```



- **Season 'Spring' has the lowest demand of bikes out of all the 4 types of Seasons.**

```
sns.barplot(x='weather',y='count',data=Train_data,estimator = np.sum)
label=[" Clear + Few clouds + Partly cloudy + Partly cloudy",
       " Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist ",
       " Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered
       " Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog "]

plt.show()
```



- **Weather 'Clear + Few clouds + Partly cloudy + Partly cloudy' has the highest demand of bikes out of all the 4 types of weather.**
- **Weather " Heavy Rain + Ice Pellets + Thunderstorms + Mist, Snow + Fog" has the lowest demand of bikes out of all the 4 types of weather.**

8. Applicable Regulations

The applicable laws, applicable codes and guidelines can be:

- Bike sharing system license/ Registration.
- Regulations against false advertising. ● Antitrust Regulations.
- Customer’s data protection and privacy regulations. ● Government Regulations for businesses. ● Employment Laws.

9. Applicable Constraints

- Continuous data collection and maintenance.
- The model needs to be updated based on the dataset.
- Taking care of the largest number of bike demand areas.
- Taking care of bike maintenance.

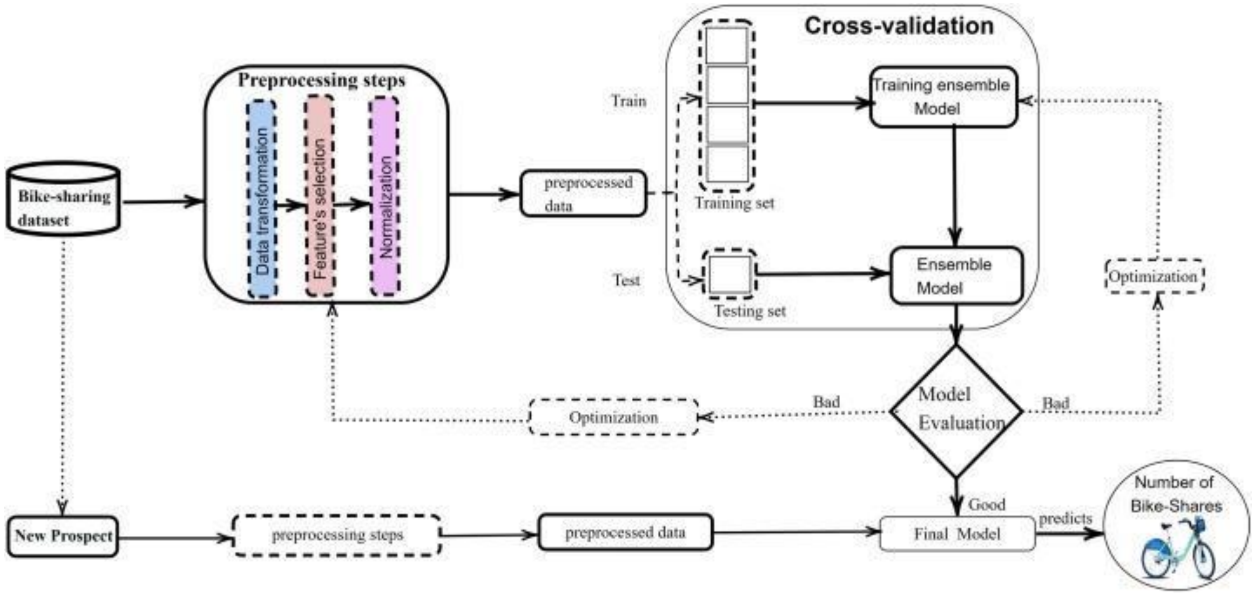
10. Business Opportunity

Business related to ebike service:

- Battery manufacturing
- Bike manufacturing
- Will help delivery service business to grow (like Swiggy and Zomato)
- Travel Data collection which can be used for other business development.
- Small businesses in the area will develop as people can easily travel to nearby shops.
- Supply chain development of ebike parts (wheels, Chassis, Motor, Battery, Wiring harness etc.)

11. Concept Generations and Development

This is a representation of the architecture of the entire process, along with the order of execution. The most amount of time must be spent on preprocessing and visualization.



12. Final Product Prototype (abstract) Model Development

The product takes the following functions to provide a good result.

Back-end

- After transformation, all the variables or features are numeric and the target variable that we have to predict is the count variable. Hence this is a typical example of a regression problem as the count variable is continuous.
- For Data Pre-processing I checked for missing values, outliers, and looked for Multicollinearity problems. I have used Label Encoder which converted categorical data into numeric form. For Feature Engineering, the Datetime column looks interesting so I extracted month, day, hour, weekday, date and year from it.
- Performing EDA to realize the dependent and independent features.
- Derives some insights from the data using data visualization.

- Before reaching my final model, I have built 'Random Forest Regressor', 'AdaBoostRegressor', 'Bagging Regressor', 'Kneighbors Regressor', however 'Random Forest Regressor' gave a bit better result as compared.

Model Building

```
models=[RandomForestRegressor(),AdaBoostRegressor(),BaggingRegressor(),
        KNeighborsRegressor()]
model_names=['RandomForestRegressor','AdaBoostRegressor','BaggingRegressor',
             'KNeighborsRegressor']
r2score=[]
d={}
for model in range (len(models)):
    clf=models[model]
    clf.fit(X_train,Y_train)
    test_pred=clf.predict(X_test)
    r2score.append(np.sqrt(r2_score(test_pred,Y_test)))
d={'Modelling Algo':model_names,'R2_error':r2score}

rmsle_frame=pd.DataFrame(d)
rmsle_frame
```

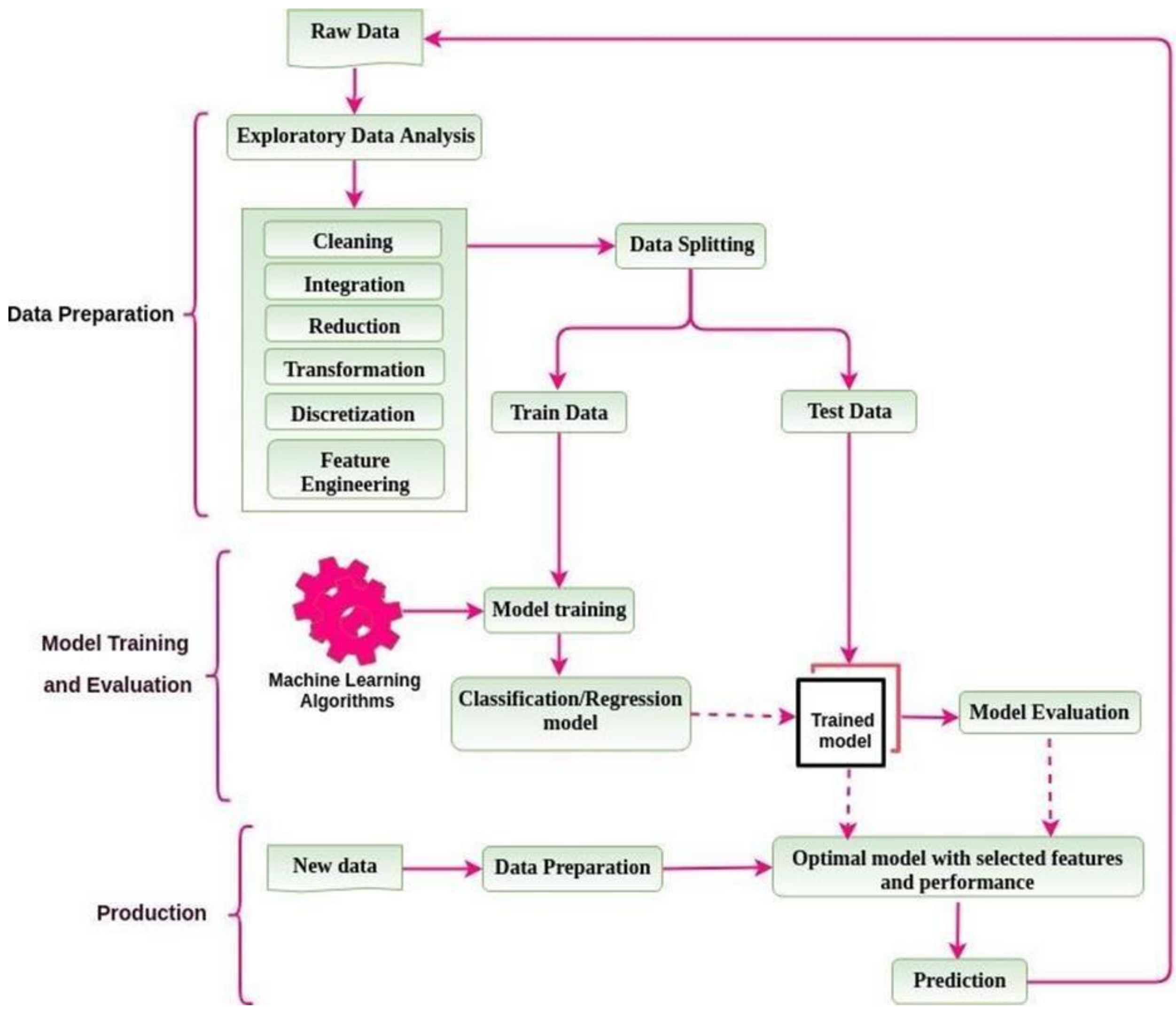
	Modelling Algo	R2_error
0	RandomForestRegressor	0.959059
1	AdaBoostRegressor	0.673915
2	BaggingRegressor	0.954721
3	KNeighborsRegressor	0.472488

- R2 score: the proportion of the variance in the dependent variable that is predictable from the independent variable(s).” Another definition is “(total variance explained by model) / total variance.” So, if it is 100%, the two variables are perfectly correlated, i.e., with no variance at all.
- Used ‘Random Forest Regression’ Machine Learning Algorithm to build my final model.

Front-end

- Different user interface: The user must be given many options to choose form in terms of parameters. This can only be optimized after a lot of testing and analysis of all the edge cases.

- Interactive visualization of the data extracted from the trained models will return raw and inscrutable data. This must be present in an aesthetic and an "easy to read" style.
- Feedback system: A valuable feedback system must be developed to understand the customer's needs that have not been met. This will help us train the models constantly.



13. Code Implementation/Validation on Small Scale

GitHub Link:
https://github.com/aaditshukla98710/Machine-Learning--Projects/tree/main/Bike%20Sharing%20Demand

14. Conclusion

As an innovative mobility strategy, public bike-sharing has grown dramatically worldwide. Though providing convenient, low-cost and environmentally-friendly transportation, the unique features of bike-sharing systems give rise to problems to both users and operators. The primary issue among these problems is the uneven distribution of bicycles caused by the ever-changing usage and (available) supply. This bicycle imbalance issue necessitates efficient bike re- balancing strategies, which depends highly on bicycle mobility modeling and prediction. Bike-sharing rental process is highly correlated to the environmental and seasonal settings. For instance, weather conditions, precipitation, day of week, season, hour of the day, etc. can affect the rental behaviors. In this report, I have proposed a model which will Predict the demand for rental bikes given information about the weather and time of the day. We extensively evaluated the performance of our model through a two-year dataset from the Kaggle bike-sharing system (BSS). Evaluation results show 95% as an R2 score, which is good enough. We believe this new mobility modeling and prediction approach can advance the bike re-balancing algorithm design.

Apart from this I have visualized the data using python and tableau. So here are a few insights from the data:

- The highest demand is in hours from 7-10 and from 15-19. This is because in most of the cities this is the peak office time and so more people would be renting bikes. This is just one of the plausible reasons.
- Season 'Spring' has the lowest demand of bikes out of all the 4 types of seasons {"Spring", "Summer", "Fall", "Winter"}.
- Weather 'Clear + Few clouds + Partly cloudy + Partly cloudy' has highest demand for bikes.
- Weather " Heavy Rain + Ice Pellets + Thunderstorms + Mist, Snow + Fog" has the lowest demand of bikes.
- Humidity is negatively correlated to count as expected as the weather is humid people will not like to travel on a bike.
- Registered/Casual and count are positively correlated which indicates that most of the bikes that are rented are registered.
- Also note that temp (or atemp) affects the count.

As we know that a business amounts to nothing without its customer, so we need to take care of the customer's problem to grow our business. Since the above technique has only been used by large companies, this can be extended for small businesses. Using this model and insights, the company can make the rental bike available and accessible to the public at the right time as it lessens the waiting time.