TABLE OF CONTENTS

ABSTRACT 3
1 Introduction 3
1.1 Problem Definition 3
1.2 Project Aim and Objectives 4
1.2.1 Project Aim 4
1.2.2 Objectives 4
2 Methodology 5
2.1 Understanding Real-World difficulties 5
2.2 Hypothesis 5
2.3 Proposed AI Model 5
2.4 Selection of Data 6
2.4.1 Dataset Collection 6
2.4.2 Feature Descriptions 6
2.5 Data Pre-processing Techniques 7
2.5.1 Approach for Data Pre-processing 7
2.5.2 Data Information 7
2.5.3 Feature Statistics 7
2.5.4 Encoding Features 8
2.5.5 Outlier Treatment 8
2.6 Selections and Operations of Libraries 9
2.7 Selection of Artificial Intelligence Model 10
2.7.1 Naive Bayes 10
2.7.2 Linear Regression 10
2.7.3 K-Neighbours 10
2.7.4 Random Forest 10
2.8 Selection of Evaluation Methods 10
3 execution 11
3.1 Data Analysis and Visualization 11
3.1.1 Analysis by Features 11
3.1.2 Cross Tabulation Analysis 11
3.1.3 Feature Distribution Analysis 12

3.2 Hypothesis Testing 12
3.3 Bike Price Prediction 13
3.3.1 Application of Regression Models 13
3.3.2 Performance Evaluation 13
3.3.3 Model Validation 14
4 Discussion and Conclusion 15
4.1 Lesson Learned 15
4.2 Hypothesis Test Result 15
4.3 Limitations 15
4.4 Future Recommendations 16
References 16

List of figures:

Figure 1 : Proposed Ai Model and Approach 7
Figure 2 : Selected Data 7
Figure 3 : Data Information 8
Figure 4 : Feature statistics 9
Figure 5 : Feature-Encoded Data 9
Figure 6 : Outlier Detection in features 10
Figure 7 : Normalized Data 10
Figure 8 : Analysis by features 12
Figure 9 : Cross-Tabulation Analysis 13
Figure 10 : Feature Distribution Analysis 13
Figure 11 : P Values of Features 14
Figure 12 : Result of Regression Models 14
Figure 13 : Comparison of R2 14
Figure 14 : Comparison of MSE 15
Figure 15 : Comparison of MAE 15
Figure 16 : Validation Result of Random Forest 15

List Of table:

Table 1: Feature Descriptions…………………………………………………………………......8

Table 2: Methods of Evaluation…………………………………………………………………12

ABSTRACT

Bikes plays a major role in daily routine, which helps for easy transportation. As price of bikes are every high, one cannot afford new bike , as an alternate to this problem people generally look for low cost used bikes. In order to buy a used bike we generally look for brand, age , km driven and owner type. As there are many brands in real world it is difficult to predict the bike price. In the above objective bike price prediction can be done using artificial intelligence and machine learning. As it consists of labeled data(brand,km_driven, age and owner) we can approach regression to get accurate bike price.

1 INTRODUCTION

1.1 PROBLEM DEFINITION

Bike price prediction using regression methods of Artificial Intelligence has gained significant attention from researchers and the industry in recent years (Feng & Wang, 2017). Predicting bike prices is crucial for both bike buyers and sellers, as it allows them to make informed decisions. However, there are several challenges in bike price prediction using AI regression that researchers and practitioners need to overcome (Reddy & Malathi, 2022). The challenges are presented below:

- The availability and quality of data is the primary challenge for bike price prediction. Accurate and perfect data is essential for developing robust prediction models (Gao, et al., 2022). To overcome this problem, several pre-processing techniques need to be applied.
- Another challenge is the determination of the feature's importance which may help to understand the influence of the bike data features for the prediction of price.
- Outlier Detection is another important part of data pre-processing for which overfitting may happen to the models (Mandhare & Idate, 2017). So, a suitable measure needs to be applied to remove outliers.
- Overfitting is a common challenge in bike price prediction using AI regression. Overfitting occurs when a prediction model fits the training data too well, leading to poor generalization performance on new data (Cai, et al., 2019).

1.2 PROJECT AIM AND OBJECTIVES

1.2.1 Project Aim

The project aims to predict the prices of used bikes based on features from historical bike records using artificial intelligence models.

1.2.2 Objectives

The objective of this project is to study machine learning model to accurately predict the price of Used Bikes by using these factors:

a. Data Collection: The first objective is to collect suitable data which will contain historical records of bile prices with other necessary features. Additionally, the hypothesis will be prepared which will be tested using the P-test method.

b. Data Pre-processing:

Next, the data will be pre-processed through data cleaning, feature encoding and outlier treatment. Additionally, the P-Test will be applied to determine the hypothesis testing.

c. Data Analysis and Visualization:

The features of the data will be analysed and visualised to get insight into the bike features and prices.

d. Data Preparation:

Then the data will be prepared for prediction by creating the training and test segments using which the artificial intelligence models will be trained and tested accordingly.

e. Model Selection and Bike Price Prediction:

The regression models from artificial intelligence will be selected which will be applied for predicting bike price. To evaluate the performances, the regression metric will be compared, and the best-performing model will be determined which will be validated with the test data.

f. Reflection: Present the overall reflection of the analytics and outcomes.

2 METHODOLOGY

2.1 UNDERSTANDING REAL-WORLD DIFFICULTIES

As the demand for bikes increases, it becomes essential to predict the prices of bikes in the market. With the advancements in Artificial Intelligence, predicting the prices of bikes has become easier than ever before (Sharma, et al., 2022). Regression is a statistical analysis that is commonly used in Artificial Intelligence to predict the prices of products. In the case of bikes, Regression can be used to predict the prices of bikes based on several parameters such as the brand, model, age, mileage, and other features. By analysing these parameters, bike prices can be predicted accurately. Regression models can be used to predict the prices of bikes based on the relationship between the parameters and the price (Chen, 2021). To predict the prices of bikes accurately, the models need to be trained using historical data. This historical data should include information such as the brand, model, age, mileage, and other features of the bikes (Holmgrena, et al., 2017). By analysing this historical data and training the models, the price of bikes can be predicted.

2.2 HYPOTHESIS

For Used Bike price prediction, data consists of features like brand, age, owner type, age and price, it is important to find any significant relation between the features.

Hypothesis is as below.

Null (H_0):

The price of bikes does not have any significant relationship with the owner type and bike age.

Alternative:

H_1: The price of the bike has a significant relationship with the bike owner type.

H_2: The price of bikes is influenced by the age of the bikes.

2.3 PROPOSED AI MODEL

The proposed approach for the prediction of bike price using artificial intelligence is shown below:

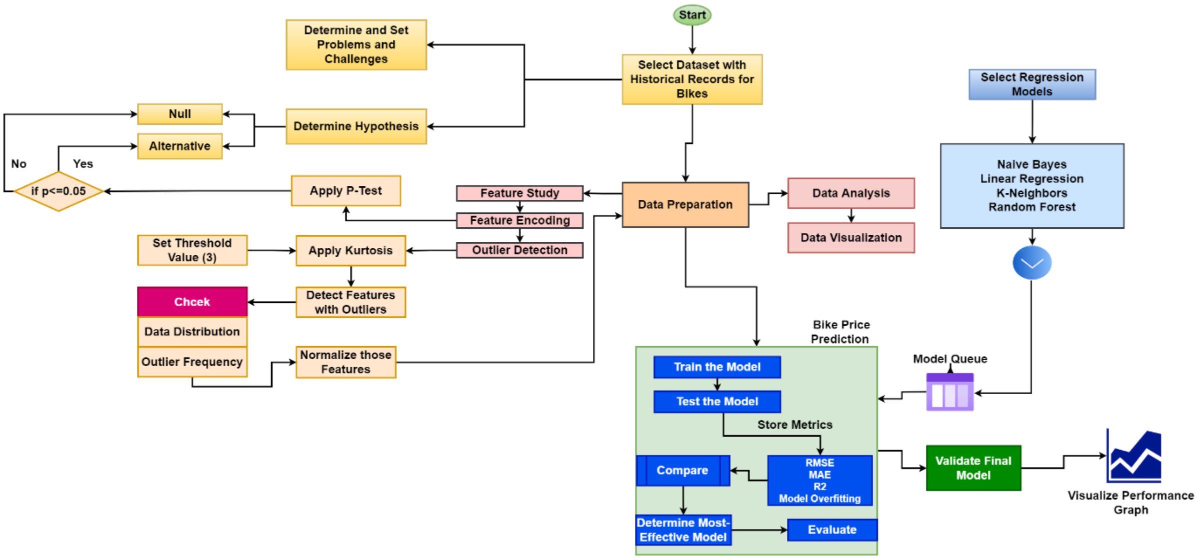

Figure 1 Proposed AI Model and Approach

The presented approach is showing several components of the methodology such as data collection, data preparation & pre-processing and finally the application of regression models to predict bike price.

2.4 SELECTION OF DATA

2.4.1 Dataset Collection

The dataset for historical bike records has been collected from Kaggle (Domala, 2021). The outlook has been presented below:

bike_name	price	city	kms_driven	owner	age	power	brand
TVS Star City Plus Dual Tone 110cc	35000	Ahmedabad	17654	First Owner	3	110	TVS
Royal Enfield Classic 350cc	119900	Delhi	11000	First Owner	4	350	Royal Enfield
Triumph Daytona 675R	600000	Delhi	110	First Owner	8	675	Triumph
TVS Apache RTR 180cc	65000	Bangalore	16329	First Owner	4	180	TVS
Yamaha FZ S V 2.0 150cc-Ltd. Edition	80000	Bangalore	10000	First Owner	3	150	Yamaha
Yamaha FZs 150cc	53499	Delhi	25000	First Owner	6	150	Yamaha
Honda CB Hornet 160R ABS DLX	85000	Delhi	8200	First Owner	3	160	Honda
Hero Splendor Plus Self Alloy 100cc	45000	Delhi	12645	First Owner	3	100	Hero
Royal Enfield Thunderbird X 350cc	145000	Bangalore	9190	First Owner	3	350	Royal Enfield
Royal Enfield Classic Desert Storm 500cc	88000	Delhi	19000	Second Owner	7	500	Royal Enfield
Yamaha YZF-R15 2.0 150cc	72000	Bangalore	20000	First Owner	7	150	Yamaha
Yamaha FZ25 250cc	95000	Bangalore	9665	First Owner	4	250	Yamaha
Bajaj Pulsar NS200	78000	Bangalore	9900	First Owner	4	200	Bajaj
Bajaj Discover 100M	29499	Delhi	20000	First Owner	8	100	Bajaj
Bajaj Discover 125M	29900	Delhi	20000	First Owner	7	125	Bajaj
Bajaj Pulsar NS200 ABS	90000	Bangalore	11574	First Owner	3	200	Bajaj
Bajaj Pulsar RS200 ABS	120000	Bangalore	23000	First Owner	3	200	Bajaj
Suzuki Gixxer SF 150cc	48000	Mumbai	24725	First Owner	5	150	Suzuki
Benelli 302R 300CC	240000	Mumbai	15025	Second Owner	3	302	Benelli

Figure 2 Selected Data

2.4.2 Feature Descriptions

The dataset contains 32648 historical bike records with 23 different brands from 443 Different cities with a total of 471 bike types. The dataset contains 8 features and the descriptions are shown below in the table:

Table 1 Feature Descriptions

Feature Name	Feature Type	Description
bike_name	Predictor	Name of the enlisted bike
price	Target	Price of the enlisted bike
City	Predictor	City from where the bike has been enlisted
kms_driven	Predictor	Total kilometres are driven by the bike
owner	Predictor	Type of Bike Owning
age	Predictor	Age of the Bike
brand	Predictor	Brand name of the bike
power	Predictor	Bike Driving power

2.5 Data Pre-processing Techniques

2.5.1 Approach for Data Pre-processing

Data pre-processing has been applied to prepare the day compatible with the regression model so that the prediction of bike prices can be done. The techniques which have been applied and their outcomes are presented below.

2.5.2 Data Information

The information on data features and statistics is shown below:

```
RangeIndex: 32648 entries, 0 to 32647
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   bike_name    32648 non-null   object
 1   price        32648 non-null   float64
 2   city         32648 non-null   object
 3   kms_driven   32648 non-null   float64
 4   owner        32648 non-null   object
 5   age          32648 non-null   float64
 6   power        32648 non-null   float64
 7   brand        32648 non-null   object
dtypes: float64(4), object(4)
```

```
Data Statistics

              price     kms_driven           age         power
count  3.264800e+04   32648.000000  32648.000000  32648.000000
mean   6.829542e+04   26344.625184      8.048211    213.511302
std    9.071860e+04   22208.527695      4.031700    134.428868
min    4.400000e+03       1.000000      1.000000    100.000000
25%    2.500000e+04   12000.000000      5.000000    150.000000
50%    4.300000e+04   20373.000000      7.000000    150.000000
75%    8.000000e+04   35000.000000     10.000000    220.000000
max    1.900000e+06  750000.000000     63.000000   1800.000000
```

(a) Feature Information **(b) Data Statistics**

Figure 3 Data Information

From the feature information, it can be seen that the data does not have any missing values and there are four numerical features with four categorical features.

2.5.3 Feature Statistics

Kurtosis and Skewness have been employed on the data features to understand the data distribution and the outcomes are shown below:

```
Data Distribuion around Mean (Kurtosis)

price           55.514449
kms_driven     148.902301
age              3.905490
power            9.713065
```

(a) Data Distribution around Mean

```
Assymetric Data Distribuion (Skewness)

price         5.780190
kms_driven    6.350266
age           1.193951
power         2.579089
```

(b) Asymmetric Data Distribution

Figure 4 Feature Statistics

From the analyses using statistical measures, the features have very higher data distribution and the same for the skewness (as all kurtosis is higher than 3 and skewness is highly positive). This implies the fact towards data pre-processing will eliminate the noise from the data.

2.5.4 Encoding Features

The categorical features have been encoded to numerical values and the encoded data is shown below:

bike_name	price	city	kms_driven	owner	age	power	brand
401	35000.0	6	17654.0	0	3.0	110.0	19
295	119900.0	118	11000.0	0	4.0	350.0	17
411	600000.0	118	110.0	0	8.0	675.0	20
375	65000.0	44	16329.0	0	4.0	180.0	19
426	80000.0	44	10000.0	0	3.0	150.0	21
437	53499.0	118	25000.0	0	6.0	150.0	21
189	85000.0	118	8200.0	0	3.0	160.0	6

Figure 5 Feature-Encoded Data

2.5.5 Outlier Treatment

Several methods are available for detecting and removing outliers. A different approach has been made to detect and remove outliers by using **Kurtosis statistics.** Kurtosis is a measure of the peaked or flatness of a probability distribution (Yousef, et al., 2021). A normal distribution has a kurtosis of three. If the kurtosis is greater than three, the tails of the distribution are more peaked than those of a normal distribution, indicating that there are more extreme values. If the kurtosis is less than three, the tails are flatter than those of a normal distribution, indicating that there are fewer extreme values (Gao & Xiang, 2018). In outlier detection, kurtosis is used to identify data points that are significantly different from the rest of the data. If the kurtosis of a data point is greater than a certain threshold, it is considered an outlier (Li, et al., 2022). The advantages of Kurtosis for outlier detection are discussed below:

- It is quick and easy to calculate.
- It can be used for both univariate and multivariate data.

- It is sensitive to the shape of the distribution, making it useful in cases where the distribution is not normal.

So, using kurtosis, the following features have been seen to have outliers (shown through boxplots):

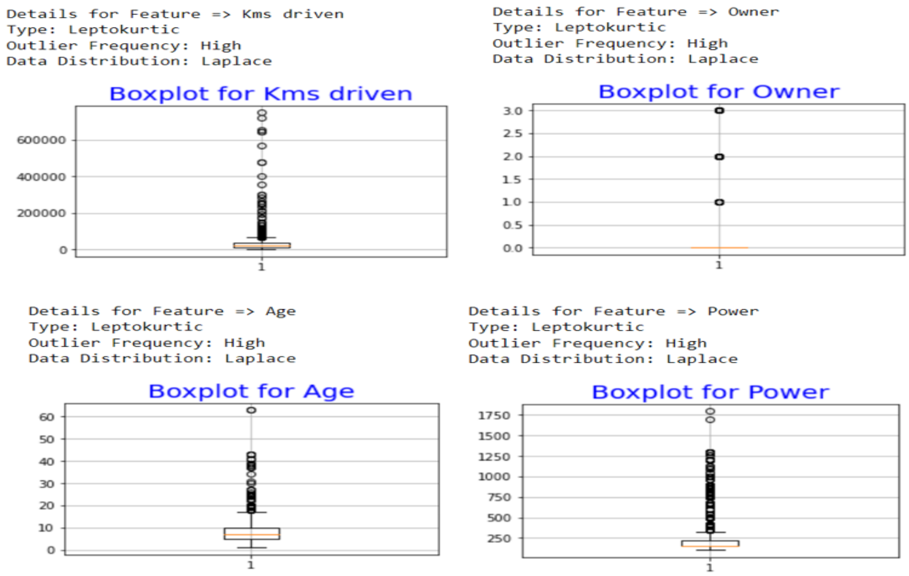

Figure 6 Outlier Detection in Features

To remove the outliers, feature normalizations have been done and the normalized data is shown below:

bike_name	city	kms_driven	owner	age	power	brand
401	6	-0.011588	-0.055787	-0.081423	-0.060889	19
295	118	-0.020460	-0.055787	-0.065294	0.080287	17
411	118	-0.034980	-0.055787	-0.000778	0.271464	20
375	44	-0.013354	-0.055787	-0.065294	-0.019713	19
426	44	-0.021793	-0.055787	-0.081423	-0.037360	21

Figure 7 Normalized Data

2.6 SELECTIONS AND OPERATIONS OF LIBRARIES

In this project for the purpose of designing artefacts, the following libraries have been used

- ***Os*** ***:To fetch the data from the local directory***
- ***numpy*** ***: To perform linear computation and array manipulation***
- ***pandas*** ***: To perform data analysis and data reading***
- ***matplotlib*** ***: For Data Visualization***
- ***scipy*** ***: To perform T-Test***

➔ *sklearn* *:To pre-process data and predict bike prices using regression models*

2.7 Selection of Artificial Intelligence Model

2.7.1 Naive Bayes

Naive Bayes is a probabilistic algorithm used for classification problems. However, it can also be used for regression problems by predicting the conditional mean of the dependent variable given the independent variables. Naive Bayes assumes that the independent variables are conditionally independent of each other given the dependent variable (Aditya, et al., 2022). This assumption allows the algorithm to make predictions quickly with limited computational resources.

2.7.2 Linear Regression

Linear regression is a simple but powerful regression technique that assumes a linear relationship between the dependent variable and the independent variables (Vasu, et al., 2022). The goal of linear regression is to find the coefficients of the linear equation that best fits the data.

2.7.3 K-Neighbours

K-Neighbours is a non-parametric regression technique that makes predictions based on the values of the k-nearest neighbours in the training set. K-Neighbours is a simple and intuitive algorithm that can capture non-linear relationships between the dependent variable and the independent variables (Hatipoğlu, et al., 2021).

2.7.4 Random Forest

Random Forest is an ensemble regression technique that combines multiple decision trees to make more accurate predictions. Random Forest reduces the risk of overfitting by randomly selecting a subset of the features and the observations to train each tree (Qunzhu, et al., 2019). Random Forest is a powerful algorithm that can capture complex relationships between the dependent variable and the independent variables.

2.8 Selection of Evaluation Methods

The following metrics have been used to evaluate the performances of the selected regression models:

Table 2 Methods of Evaluation

Metrics for Evalutaion	Operations
R2 Score	It determines the closeness of the actual value and the predicted values
Mean Squared Error	It determines the closeness of the regression line set to the actual data point
Mean Absolute Error	It determines the absolute error between the actual data and predicted data.

3 EXECUTION

3.1 Data Analysis and Visualization

The feature analyses have been done to understand the feature statistics. This will also help to prepare the decision on bike purchase. The outcomes of the analyses have been presented in this section in the form of dashboards.

3.1.1 Analysis by Features

The analyses are showing that the top bikes are belonging to Delhi city followed by Bangalore. Bajaj Pulser has been seen to be the bike with the highest demand. Additionally, the bikes are available at the highest price from the first owner.

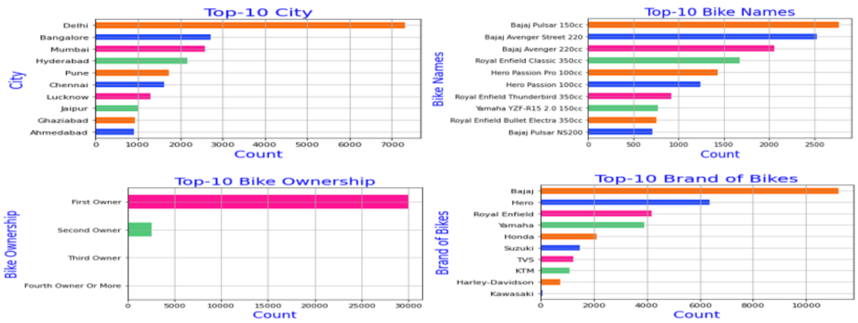

Figure 8 Analysis by Features

3.1.2 Cross Tabulation Analysis

The bi-feature analyses have been done through cross-tabulation. The outcomes are showing that the bikes (second-hand) are available from Solapur followed by Vapi where the top bikes are Yamaha & Ducati (by price) and Hero & Honda (by kilometre driven).

Figure 9 Cross-Tabulation Analysis

3.1.3 Feature Distribution Analysis

The outcomes shown below dashboard are that the bikes with variant prices are available from the third and fourth owners of the bikes. However, those bikes have been used for driving higher kilometres as well.

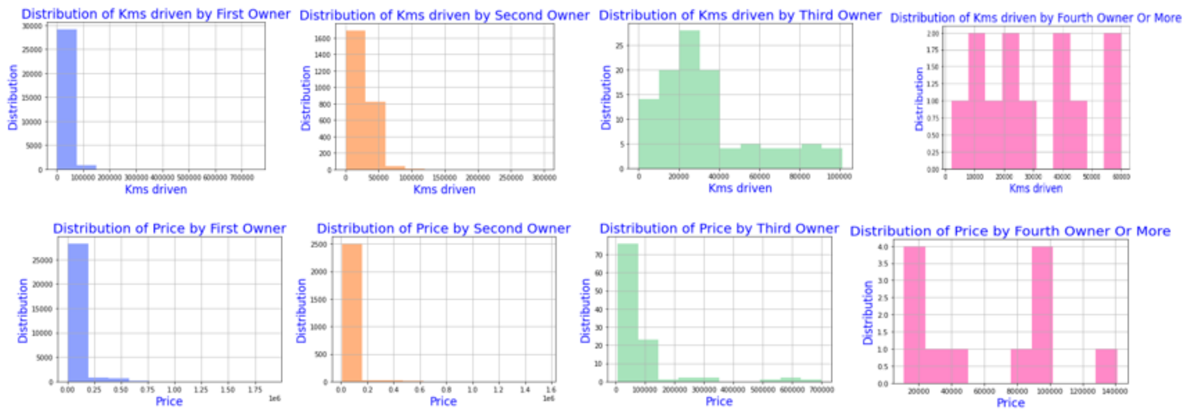

Figure 10 Feature Distribution Analysis

3.2 Hypothesis Testing

P-Test has been applied to check the p-values of the features with the prices of bikes. This step has been taken to prove the hypothesis and to observe the feature relationship. The list of p-values for all features is shown below:

Feature	T-Value	P-Value
bike name	135.648988	0.0
city	135.680665	0.0
kms driven	136.026505	0.0
owner	136.026505	0.0
age	136.026505	0.0
power	136.026505	0.0
brand	136.009607	0.0

Figure 11 P Values of Features

From the p-values of the features, it can be said that the features are highly influential and significantly related to predicting the price of bikes.

3.3 Bike Price Prediction

3.3.1 Application of Regression Models

The selected regression models have been applied to the data and the prices of bikes have been predicted. For each of the regression models, the metrics such as r2 score, MSE and MAE have been stored so that those can be compared to find the best-performing model.

Algorithm	R2 Score(Test)	R2 Score(Train)	MSE(Test)	MSE(Train)	MAE(Test)	MAE(Train)
Linear Regression	83.93	81.76	1.246112e+09	19962.562298	19662.689975	1.529832e+09
K-Neighbors	84.25	84.26	1.221708e+09	6053.257788	6880.439868	1.320212e+09
Naive Bayes	90.45	93.74	7.409609e+08	9077.185780	9964.875766	5.250045e+08
Random Forest	97.77	99.27	1.728792e+08	1156.725730	2717.785267	6.153599e+07

Figure 12 Results of Regression Models

3.3.2 Performance Evaluation

Comparison of R2

The R2 Scores for selected models have been compared concerning training and testing. It can be seen that the Random Forest model has performed the best with less model overfitting.

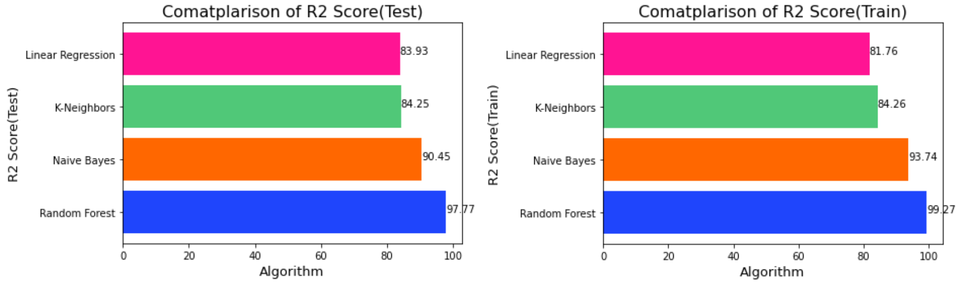

Figure 13 Comparison of R2

Comparison of MSE

The comparison of MSE is showing the same fact that the error is least for the Random Forest.

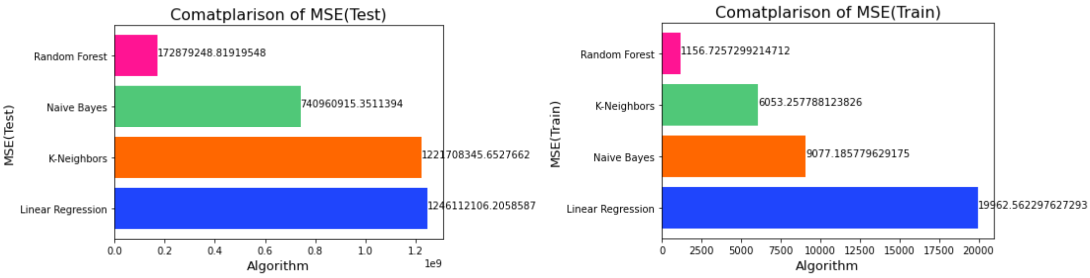

Figure 14 Comparison of MSE

Comparison of MAE

The comparison of MAE is showing the same fact that the error is least for the Random Forest.

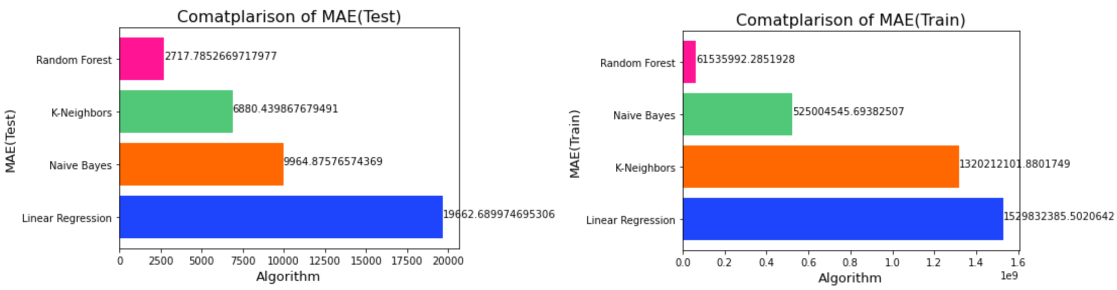

Figure 15 Comparison of MAE

3.3.3 Model Validation

From the evaluation of the models, it can be stated that Random Forest is the best-suited mode to predict bike prices. So, this model has been validated by taking 20 test data and predicting prices. The below graph has been showing the actual price versus the predicted prices of bikes.

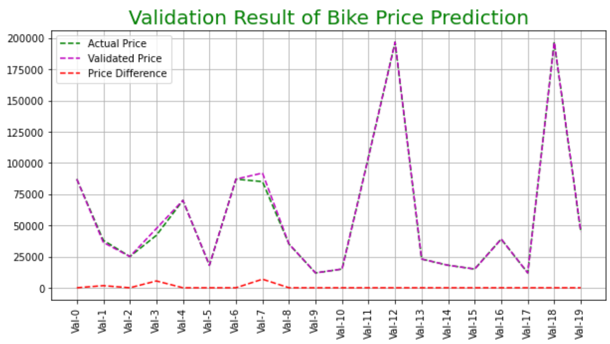

Figure 16 Validation Result of Random Forest

From the validation, it has been observed that the Validation R2 is 99.86, MAE 700.0 and MSE 3980347.6. So, the result of bike prediction can be said to be satisfactory.

4 Discussion and Conclusion

4.1 Lesson Learned

The project has been successfully conducted to predict bike prices. From the overall analytics and the application of artificial intelligence models, the following lessons have been learned:

1. The data has been pre-processed by applying necessary measures such as feature encoding (which enables the data to be applicable in AI models) and outlier removal (which enables to remove of the noise from data)
2. The kurtosis is the statistical method which helps to determine the feature having outliers precisely. This is the reason, some of the features have been normalized which have been detected with outliers using kurtosis rather that normalizing the entire data.
3. The overfitting of models is an important metric which helps in understanding the performance of the model in both training and test data.
4. Random Forest is the AI model for regression which has been seen to be performing with the highest R2 score (both training and testing) and with the least errors.
5. Finally, the validation of the model (Random Forest) has helped with the determination of the final performance on the unseen data (test data) to predict price.

4.2 Hypothesis Test Result

The hypothesis test has been done by employing a p-test. From the p values of the features with bike prices, it has been seen that values for owner type and bike age with bike prices are less than 0.05. It redirects the fact towards the acceptance of an alternative hypothesis.

4.3 Limitations

The limitations of the present project are discussed below:

1. The Kurtosis method has been applied for the detection of outliers and no other methods have been applied and tested.
2. Several regression models are available in Artificial Intelligence whereas four modes have been selected and applied.
3. To evaluate the performances, three metrics have been used.
4. Feature selection has not been done.

4.4 Future Recommendations

The project can be enhanced further, and the recommendations are as follows:

1. Deep learning models such as Long Short-Term Memory can be used to predict bike prices.
2. Different methods of outlier detection methods such as PCA can be used.
3. The feature selection process can be applied to select important features however lessening the number of features may create an issue as the total number of data features is few.

References

Aditya, E. et al., 2022. New Student Prediction Using Algorithm Naive Bayes And Regression Analysis In Universitas Potensi Utama. *4th International Conference on Cybernetics and Intelligent System (ICORIS)*, pp. 1-6.

Cai, S. et al., 2019. An efficient outlier detection approach on weighted data stream based on minimal rare pattern mining. *China Communications*, 16(10), pp. 83 - 99.

Chen, S., 2021. Correlation analysis of financial indicators and stock price fluctuations based on artificial intelligence system. *International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 1-4.

Domala, S. S., 2021. *Used Bikes Prices in India.* [Online] Available at: https://www.kaggle.com/datasets/saisaathvik/used-bikes-prices-in-india

Feng, Y. & Wang, S., 2017. A forecast for bicycle rental demand based on random forests and multiple linear regression. *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pp. 1-6.

Gao, Q. & Xiang, J., 2018. A Method Using EEMD and L-Kurtosis to Detect Faults in Roller Bearings. *Prognostics and System Health Management Conference (PHM-Chongqing)*, pp. 1-5.

Gao, Y. et al., 2022. Short-Term Traffic Speed Forecasting Using a Deep Learning Method Based on Multitemporal Traffic Flow Volume. *IEEE Access*, pp. 82384-82395.

Hatipoğlu, Ş., Belgrat, M. A., Degirmenci, A. & Karal, Ö., 2021. Prediction of Unemployment Rates in Turkey by k-Nearest Neighbor Regression Analysis. *Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1-4.

Holmgrena, J., Aspegrena, S. & Dahlströma, J., 2017. Prediction of bicycle counter data using regression. *2nd edition of the International Workshop on Data Mining on IoT System*, Volume 113, pp. 502-507.

Li, T., Peng, D. & Shi, S., 2022. Outlier-Robust Superpixel-Level CFAR Detector With Truncated Clutter for Single Look Complex SAR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 5261 - 5274.

Mandhare, H. C. & Idate, S. R., 2017. A comparative study of cluster based outlier detection, distance based outlier detection and density based outlier detection techniques. *International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1-5.

Qunzhu, T. et al., 2019. Improvement of Random Forest Cascade Regression Algorithm and Its Application in Fatigue Detection. *IEEE 2nd International Conference on Electronics Technology (ICET),* pp. 1-5.

Reddy, R. T. & Malathi, P., 2022. Linear Regression and Artificial Neural Networks based Efficient Sales Forecasting Model with Increased Prediction Accuracy. *4th International Conference on Inventive Research in Computing Applications (ICIRCA),* pp. 1100-1105.

Sharma, P., Vijayvargiya, A., Singh, B. & Kumar, R., 2022. Data Driven Temperature Estimation of PMSM with Regression Models. *Second International Conference on Power, Control and Computing Technologies* , pp. 1-6.

Vasu, V. N., Surendran, Saravanan & Madhusundar, 2022. Prediction of Defective Products Using Logistic Regression Algorithm against Linear Regression Algorithm for Better Accuracy. *International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT),* pp. 1-5.

Yousef, W. A., Traoré, I. & Briguglio, W., 2021. UN-AVOIDS: Unsupervised and Nonparametric Approach for Visualizing Outliers and Invariant Detection Scoring. *IEEE Transactions on Information Forensics and Security*, pp. 5195 - 5210.