International Journal of Scientific & Technical Development
(A Peer-Reviewed Bi-Annual Journal) ISSN: 2348-4047

PRICE PREDICTION OF USED BIKE USING MACHINE LEARNING

Raju Kumar Sah
Dept of Computer Science Engineering
Vignan's Foundation for Science, Techonology & Research, Vadlamudi, India
rajusah4054@gmail.com
S.K Gausian Muskan
Dept of Computer Science Engineering
Vignan's Foundation for Science, Technology & Research Vadlamudi, India
gousiask.2606@gmail.com
P. Upadhyay
Dept of Computer Science Engineer Vignan's Foundation for science, Technology & Research Vadlamudi, India
prashant3213@gmail.com

Abstract
Predicting bike prices accurately is essential for buyers and sellers alike in the dynamic market of biking. This study presents a practical approach utilizing machine learning techniques to forecast bike prices based on various attributes such as brand, model, kilometer, owner, condition, and city.
The methodology involves preprocessing the data to handle missing values and categorical variables, followed by the application of regression algorithms including random forest, decision tree, linear regression, and gradient boosting. Model performance is evaluated using common metrics like error or mean square and R-squared to assess prediction accuracy.
Future directions involve refining the models and incorporating additional data sources for improved predictive capabilities. We use the KNN and Naïve Bayes Algorithm and have developed a Hybrid model for predicting the price of bike easily. To predict the output or model of the dataset which can give a better accuracy from the existing model. The algorithms are Naïve Bayes, KNN (k-Nearest Neighbours) by comparing various algorithms results, Hybrid model is providing best accuracy.
Keywords—[Decision tree, Hybrid Model, KNN, Linear Regression, Machine Learning, Naïve Bayes, Random Forest].

Introduction:

Predicting bike prices accurately is a central aspect in the automotive organization, enabling consumers to make informed purchasing decisions and businesses to strategize effectively. In recent years, the advent of advanced data analytics and machine learning algorithms has revolutionized the way we forecast prices. By analyzing various factors such as market trends, economic indicators, manufacturer data, and consumer preferences.

So, to clear the problem we have a good solution the name Bike Price Prediction comes into the picture. Bike pricing is influenced by competitive factors such as pricing strategies of rival brands, promotional offers, and discounts.

Addressing these challenges requires advanced data analytics techniques, including model of machine learning that can explore big datasets, identify relevant patterns, and adapt to changing market conditions to make accurate predictions [2]. Existing model of bike price predicting results are based on parameters fixed price ranges for different bikes parameters. Bringing out most accurate or correct predictions to the users using the machine learning algorithms. The primary objective forecast the price of a given bike based on various features and characteristics such as brand, model, year of manufacture, kilometers driven, power, owner history, location, and price data. High-quality data is fundamental to model accuracy. How can we create an effective machine-learning model to foreshow the prices of bicycles based on various attributes and market trends? By harnessing vast datasets and leveraging sophisticated algorithms, practitioners can unlock invaluable insights into pricing dynamics, empowering businesses to optimize inventory management, devise competitive pricing strategies, and cultivate customer trust. Thus, the pursuit of accurate bike price prediction not only facilitates efficient market operations but also fosters a symbiotic relationship between manufacturers, retailers, and consumers in the ever-evolving landscape

of the biking industry. [7].

The paper is organized as : Section II gives the brief introduction about existing model of the bike price prediction, In section III includes methodology used in the proposed model, Section IV described the Hybrid proposed model, Section V included the result and discussion, Section VI described the conclusion and section VI includes the references.

II. RELATED WORKS

A survey of research on Bike price prediction using machine learning algorithm to improve prediction model and calculating the accuracy of a bike price which is used already in market. By using the different types of machines learning already some models have developed in the organization such as KNN, Naïve Bayes, Decision Tree etc. by Lucija Bukvić and colleagues [1] with 95 % accuracy.

Similarly other people also have developed some model with deferent-deferent accuracy such as 95 % accuracy in [1] by KNN, Naïve bayes, and Linear Regression and 55 % accuracy in [2] by using Linear Regression and Random Forest.

Similarly, some more models are there which developed by other people ex- Random Forest and Decision tree with 81 % accuracy [3].

III. METHODOLOGY

A. Data Collection

Collection of data is a process of collecting the dataset which includes attributes such as model name, price, city, kilo meter, Owner, Age, Power etc. These all attributes belong the collected dataset.

B. Data Preprocessing:

Addressing outliers, discrepancies, missing numbers, and other data issues requires data cleaning. Make sure that numerical properties are comparable by standardizing or normalizing them. Use methods like label encoding or one-hot encoding to encode categorical information.

Learn more about the variables' distributions and properties by exploring the datasets with descriptive statistics and visualization.

C. Feature Selection:

Find important characteristics linked to find problems using statistical approaches (e.g., feature importance ratings, correlation analysis). To choose useful predictors, use domain expertise and advice from experts. As the number of features increases, the data becomes increasingly sparse in the feature space, making it harder for the model to learn effectively. Feature selection can help mitigate this issue by reducing the number of irrelevant or redundant features.

D. Model Training:

Separate the cleaned-up data into two sets: one for assessing its efficacy training and the other for training the model. Individual prediction models may be built using decision tree and random forest classifiers to detect trends in the data. Use methods like KNN, Random forest to fine-tune the model's hyperparameters. To enhance overall accuracy, use ensemble learning methods like Voting Classifier to aggregate predictions from many base classifiers. To make the models even more resilient and good at finding the good accuracy of the bike price prediction.

E. Machine Learning model

Apply the different type of machine learning model for create a new model based on its accuracy value such as KNN, Naïve Bayes, Random Forest, Linear Regression, Decision Tree and proposed model such as Hybrid algorithm.

F. Final Model of the dataset with accuracy.

After the machine learning model, we can see the result of the dataset with accuracy of the algorithm which help us to predict the price of a bike easily.

IV. PROPOSED MODLE

Proposed Model

Hybrid NB-KNN is a novel machine learning algorithm

that combines the strengths of Naive Bayes and K Nearest Neighbours (KNN) classifiers.

It leverages Naive Bayes for probabilistic classification and KNN for local pattern recognition.

The algorithm computes class probabilities using Naive Bayes and incorporates them as weights in the KNN prediction mechanism, resulting in a more robust and accurate classification model of a bike price prediction.

Workflow

Training Phase:

The Naive Bayes classifier is trained on the dataset to learn the probability distributions of features given each class.

The K Nearest Neighbours classifier is trained on the same dataset to learn local patterns in the feature space.

Prediction Phase:

During prediction, the Naive Bayes classifier calculates the class probabilities for each data point.

These probabilities are used as weights for the KNN algorithm, where each neighbour's class is weighted by its probability.

The weighted predictions from KNN are normalized to ensure a valid probability distribution.

The final prediction is built up by selecting the class with the highest weighted sum of probabilities.

Proposed Model has been shown in figure 1.

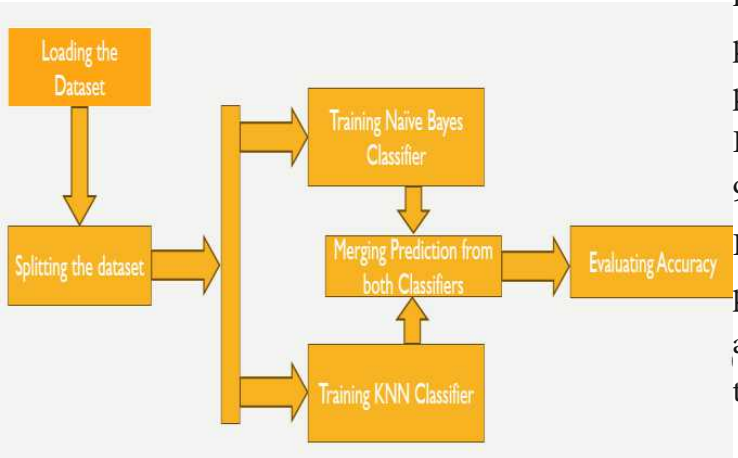

Figure 1. Proposed model

Step1. First, we need to load the data set which are consisting a large amount of data with different type of attributes and such as kilometer, owner name, model name of the bike etc.

Step2. Splitting the dataset for an unbiased evaluation of prediction of accuracy.

Step 3. Training the naïve bayes classifier for individual accuracy.

Step4. Training the KNN classifier for again its individual accuracy.

Step5. Merging prediction from both classifier and finally check the accuracy percentage.

Advantage

Combines the advantages of probabilistic classification (Naive Bayes) and local pattern recognition (KNN).

Robust to noisy data and capable of handling both categorical and numerical features.

Can capture complex decision boundaries and adapt to various types of datasets.

V. RESULT AND DISCUSSIION.

By using the KNN algorithm which provides 93% accuracy for predict the price of a bike. By using the Naïve Bayes algorithm has been developed a model which gives 27% accuracy which is very less percentage of predicting the accurate price of the bike.

By developing the Hybrid model with the help of merging KNN & Naïve Bayes Algorithm that is providing 97% accuracy which can give an accurate price of the bike compare to existing KNN and Naïve Bayes algorithm and more existing model also. Example 95 % in 2022 [1], 55% in 2017 [4] etc.

In literature survey each model has less accuracy than proposed model and hybrid mode is better compare to all literature survey result so, that hybrid model can give the accurate price of the bike for a buyer and a seller.

Comparison with existing model has been shown in fig 2.

COMPARISON WITH EXISTING MODEL

Machine Learning Model	Accuracy
KNN [1]	95%
Linear Regression.[2] Random Forest. [2]	84% 81%
Naïve Bayes.[1]	27%
Random Forest and decision tree.[3]	81%
Proposed Model	97 %

Figure 2. Comparison with existing model

As we can see in the image that in every Model accuracy is

good if compare with naïve bayes which is very less but after merging the both algorithm approach of KNN and Naïve Bayes we are getting 97 % accuracy for predicting the price of a bike on the given dataset.

In the comparison existing model by using the KNN algorithm accuracy is 95 % which is developed by Lucija Bukvic on his dataset. But on taken dataset by using the same approach and algorithm existing model is providing only 93 % accuracy as you can see in the final result.

Data set Information has been shown in figure 3.

• **Data Collection and representation**

bike_name	price	city	kms_driver	owner	age	power
TVS Star Cit	35000	Ahmedabac	17654	First Owner	3	110
Royal Enfiel	119900	Delhi	11000	First Owner	4	350
Triumph Da	600000	Delhi	110	First Owner	8	675
TVS Apache	65000	Bangalore	16329	First Owner	4	180
Yamaha FZ:	80000	Bangalore	10000	First Owner	3	150
Yamaha FZs	53499	Delhi	25000	First Owner	6	150
Honda CB H	85000	Delhi	8200	First Owner	3	160
Hero Splen	45000	Delhi	12645	First Owner	3	100
Royal Enfiel	145000	Bangalore	9190	First Owner	3	350

Figure 3. Data Information

Collected dataset is consisting some attributes such as name of the bike, model of the bike, owner of bike, city, kilo meter , age, power and price etc. on this set of attributes, proposed model providing about 97 % accuracy which developed by using the two existing model first is KNN algorithm and other Naïve bayes algorithm.

Data Visualization: It has been shown in figure 4.

This is a plot of predicted values v/s the values of the prices that are present in the dataset for the model using the KNN algorithm.

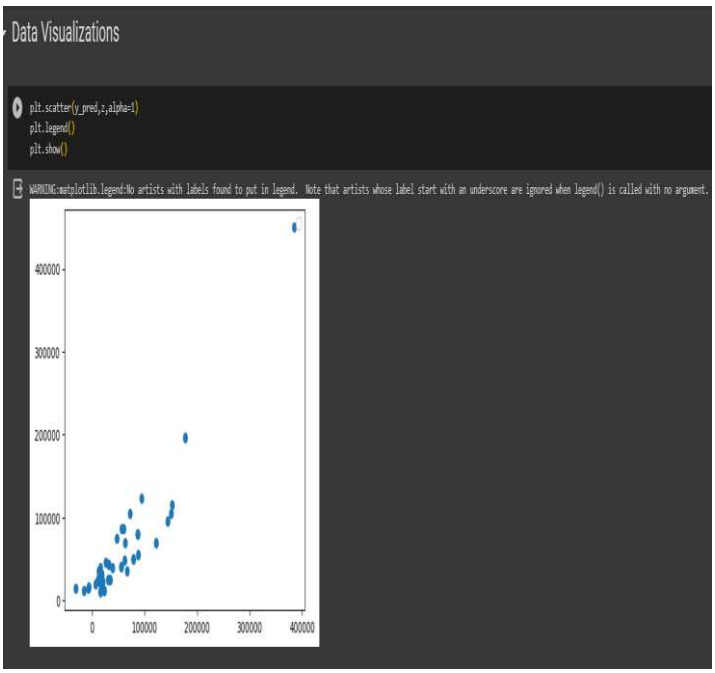

Figure 4. Data Visualization

Result of the Propose Model has been shown in figure 5.

Using many different machine-learning models on the data set they are K-Nearest Neighbours, Linear Regression, Naïve Bayes Classifiers, Random Forest, and Decision tree we've calculated the vailed Accuracy of the given dataset but our proposed model is giving some good accuracy about 97 % which can predict the price of a bike easily.

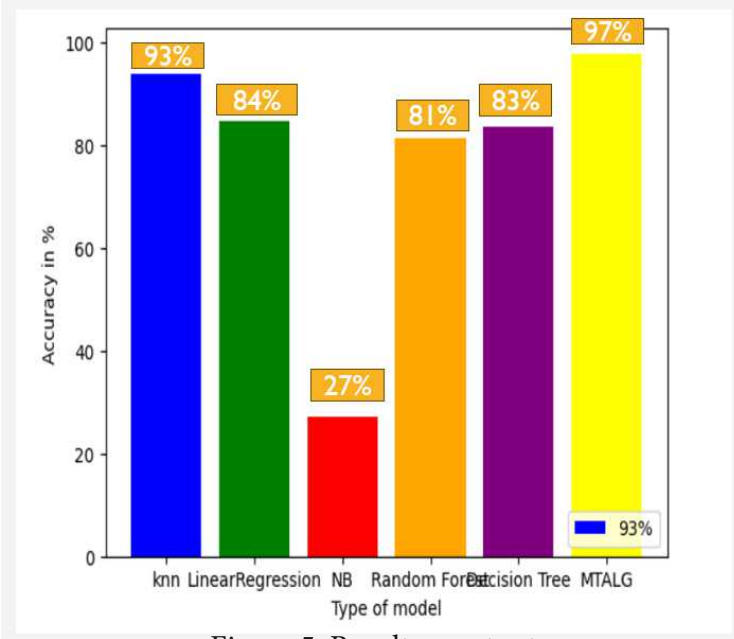

Figure 5. Result or output

VI. CONCLUSION

Predicting bike prices using machine learning can yield valuable insights and accurate estimates. Through the analysis of various features such as brand, model, mileage, year of manufacture, and additional functionalities, machine learning algorithms can learn patterns and correlations to make predictions. Techniques like regression, decision trees, or neural networks can be employed depending on the complexity of the problem and the available data.

Hoping that We've learned a lot about dealing with data, especially on this data set. Exploring through the entire data-making analysis using different kinds of classifiers for predicting the accurate answers.

Tried different kinds of classifiers, at last, dealt with an accuracy of 93% using the K neighbors Classifiers and an accuracy of 84% using the linear regression model.

Finally using hybrid algorithm, we getting an accurate accuracy for predict the price of bike that accuracy is 97% which can give an accurate price of bike.

VII. REFERENCES

[1] Bukvić, L., Pašagić Škrinjar, J., Fratrović, T. and Abramović, B., 2022. Price prediction and classification of used-vehicles using supervised machine learning. Sustainability, 14(24), p.17034.

[2] Chen, C., Hao, L. and Xu, C., 2017, May. Comparative analysis of used car price evaluation models. In AIP Conference Proceedings (Vol. 1839, No. 1). AIP Publishing.

[3] Louppe, G., 2014. Understanding random forests. Cornell University Library, 10.

[4] Yin, Y.C., Lee, C.S. and Wong, Y.P., 2012. Demand prediction of bicycle sharing systems. URL http://cs229. stanford. edu/proj2014/YuchunYin, ChiShuenLee, Yu-PoWong, DemandPredictionofBicycleSharingSystems. pdf.

[5] Ileberi, Emmanuel, Yanxia Sun, and Zenghui Wang. "A machine learning based credit card fraud detection using the GA algorithm for feature selection." Journal of Big Data 9.1 (2022): 24.

[6] Wang, B. and Kim, I., 2018. Short-term prediction for bike-sharing service using machine learning. Transportation research procedia, 34, pp.171-178.

[7] Kim, T.Y., Park, M.J., Shin, J. and Oh, S., 2022. Prediction of Bike Share Demand by Machine Learning: Role of Vehicle Accident as the New Feature. International Journal of Business Analytics (IJBAN), 9(1), pp.1-16.

[8] Prada, S.J., Sri, A.G., Venkateswarlu, B., Vineesha, C. and Teja, P.L., 2020. Bike Buyer Prediction. International Journal of Computer Engineering and Technology, 11(3).

[9] Bhagirath, Mittal, N. and Kumar, S., 2021. Online Resale Bike Price Prediction in Indian Market. In Innovations in Cyber Physical Systems: Select Proceedings of ICICPS 2020 (pp. 157-167). Springer Singapore.

[10] Yang, R.R., Chen, S. and Chou, E., 2018. AI blue book: vehicle price prediction using visual features. arXiv preprint arXiv:1803.11227.