UNIVERSITY SCHOOL OF INFORMATION, COMMUNICATION AND TECHNOLOGY

Submitted in the partial fulfilment of the requirement of the degree of
MASTER OF COMPUTER APPLICATIONS (MCA (SE))

TERM PAPER REPORT

FOR

CAR PRICE PREDICTION

MENTOR:
Dr. Jaspreeti Singh
(Assistant Professor)

MENTEE:
AMITESH KUMAR SINGH
MCA (SE) - 1st year
Roll No: 04616404523

DECLARATION

This is to certify that the project report entitled “CAR PRICE PREDICTION” is done by me is an authentic work carried out for the partial fulfilment of the requirements for the award of the degree of M.C.A (S.E), under the guidance of Associate Professor Dr. Jaspreeti Singh.

The matter embodied in this project work has not been submitted earlier for the award of any degree or diploma to the best of my knowledge and belief.

Dated: 20/11/2023 AMITESH KUMAR SINGH

Certificate

Certified that the Project Report (MCA) entitled "**CAR PRICE PREDICTION**" done by the student named "**AMITESH KUMAR SINGH**" (Roll no.04616404523) is completed under my guidance.

To the best of my knowledge and belief, work done by the candidate has not been submitted for the award of any other degree.

Signature of the Guide

Date: 20/11/2023

Name of the Guide: **Dr. Jaspreeti Singh**
Designation: **Assistant Professor**
(USICT, GGSIPU)

ACKNOWLEDGEMENT

With candour and pleasure, I take the opportunity to express my sincere thanks and obligation to my esteemed guide **Dr. Jaspreeti Singh.**

It is because of her able and mature guidance and co- operation without which it would not have been possible for me tocomplete my project.

It is my pleasant duty to thank all the staff members of the computer Centre who never hesitated to help me from time to time during the project.

INDEX

INTRODUCTION

Machine Learning (ML) is a key player in modern industry and research, constantly improving computer system capabilities with the use of complex algorithms and neural network models. ML algorithms use "training data," or sample data, to build mathematical models on their own. This allows decision-making to occur without the need for explicit programming for each decision.

In the automotive sector, automobile dealerships and private individuals engage in the purchase and sale of vehicles for investment, personal use, or commercial purposes. With the ongoing problem of overvaluation or undervaluation in the auto industry and the dearth of efficient detection tools, it is critical to ensure a fair transaction. While price-to-performance ratios and other measures provide some early clues, a thorough examination and well-informed decision-making are necessary.

In addressing these problems, machine learning's promise is shown. A reliable method for precisely predicting car prices that satisfies the various demands of buyers and sellers can be developed by training machine learning models on large datasets. This study aims to utilize several machine learning techniques and incorporate them into ML models that may efficiently assist consumers in the automotive area.

Finding the perfect car with the characteristics they want within their budget is the main objective for customers. But there's no assurance that you'll be able to get the car for a reasonable price without going over budget. In a similar vein, sellers need to do extensive research to ascertain a fair market value in order to establish a decent price for their automobiles based on an informed valuation.

Furthermore, the risk of underpricing the product underscores the importance of accurate price predictions. Offering users predicted car prices empowers them to make informed decisions, ensuring fairness in transactions for both buyers and sellers.

LIMITATIONS OF PREVAILING METHODOLOGIES

There is considerable research in the field of car price prediction, but real-life solutions in this domain are limited. Digital solutions for predicting car prices are sparse, and the existing methods used by individuals and companies face several challenges.

Buyers/Customers:

When people are thinking about buying an automobile, they frequently use internet resources to research trends and acquire relevant data. But the average person might not know all the details when it comes to the true worth of an automobile. Due to a lack of understanding of the dynamics of the market, buyers may mistakenly think that the prices listed online are accurate, resulting in misinformation.

Reaching out to different auto dealerships is another popular strategy. The problem with this is that these dealers frequently impose an expense for looking up and determining the price of a particular vehicle. These price tags might be taken at face value by buyers, who might end up overpaying as a result of hidden deals between sellers and dealers.

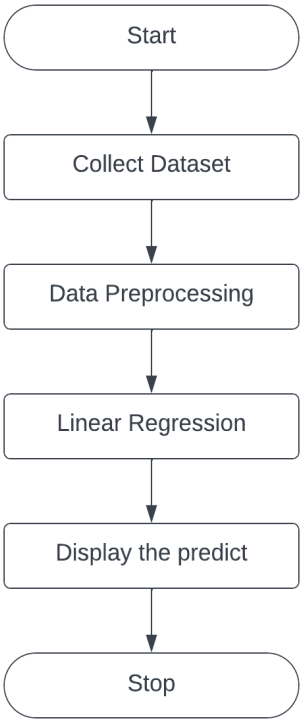
Sellers/Dealerships:

When someone chooses to sell their car, they usually do an internet comparison with a large number of other vehicles. Finding the best deal by hand-comparing prices with several different cars can be laborious and error-prone. Bigger dealerships that carry a wide range of car models frequently designate specific people to handle product pricing. But depending on human judgment leaves room for biases and mistakes.

Using computer-based solutions that make use of artificial intelligence, machine learning, and data analytics could simplify the procedure in this case. Based on a number of variables, including the make, model, mileage, and market trends, Automated Valuation Models (AVMs) for automobiles may offer more impartial and accurate pricing. When compared to human-driven methods, such technology-driven solutions may save costs, increase accuracy, and save time.

METHODOLOGY

This project aims to employ a regression-based approach for predicting continuous variables, specifically focusing on the target variable 'Price,' within a supervised machine learning framework. The selected algorithm for this task is Linear Regression, deemed adequate for forecasting the value of 'Price' by determining the Y-intercept and slope of the regression line, accounting for inherent noise in the data.



Block Diagram of the System

Data Gathering and Arrangement:

1. **Collect Relevant Data:** Compile a thorough dataset of sales of used cars, taking into account attributes such as location, make, model, year, mileage, and condition. Websites run by governments, auto dealerships, and internet marketplaces are examples of data sources.
2. **Data Cleaning:** Eliminate outliers, inconsistent data, and missing values from the data. Use suitable techniques to replace missing values, such as mean or median imputation. Use methods such as the z-score or interquartile range (IQR) to find and eliminate outliers.
3. **Data Transformation:** Adjust all variables to a common range in order to standardize the data. By doing this, it is ensured that all features are on a similar scale and that numerical features do not dominate the model.

4. **Data Splitting:** Separate the data into sets for testing and training. The testing set (20%) assesses the model's performance on unobserved data, while the training set (usually 80%) is used to train the model.

Modeling with Linear Regression:

1. **Model Selection:** Select a suitable linear regression algorithm, like ridge regression or ordinary least squares (OLS). A linear relationship between the car's features and price is established by these algorithms.
2. **Model Training:** Using the training set of data, train the chosen linear regression model. For every unit change in each feature, the model learns the coefficients that indicate the change in car price.
3. **Model Evaluation:** Assess the effectiveness of the trained model using the test data. Root mean square error (RMSE) and mean absolute error (MAE) are examples of common metrics. These metrics calculate the discrepancy between actual and projected car costs.

Determination and Improvement of the Model:

1. **Coefficient Analysis:** Examine the model's coefficients to determine how each feature affects the cost of a car. Significant statistical correlations suggest a robust association between the attribute and cost.
2. **Feature Selection:** Take into account getting rid of features that don't really add much to the cost of the car. This can lessen overfitting and enhance the model's interpretability.
3. **Hyperparameter Tuning:** To maximize the model's performance, adjust its hyperparameters, such as the regularization parameters. Methods such as grid search and random search can be used to identify the ideal values for hyperparameters.

Deployment and Prediction:

1. **Deployment of Model:** The model can be used to make predictions on fresh car data after it has been trained and improved. In order to get an estimated price, this entails feeding the model's features into the car.
2. **Real-time Predictions:** Incorporate the model into websites or applications to offer real-time estimates of car prices. This makes it simple for customers to obtain price quotes for various car models and combinations.

IMPLEMENTATION

Dataset

The dataset utilized in this analysis was obtained from the online marketplace "Quikr," which specializes in the buying and selling of automobiles. The dataset includes attributes like Model, Year, Price, Kms_driven, and fuel type that are essential for predictive analysis. Because of the size of the dataset, preprocessing operations like feature engineering and deduplication are expected to be required in order to improve the dataset's quality and usefulness.

- Import Libraries: The code begins by importing the necessary libraries. `pandas` is used for data manipulation, and `scikit-learn` is employed for machine learning tasks.
- Load the Dataset: The dataset containing information about cars, including features such as 'Year,' 'Kms_driven,' 'fuel_type,' and the target variable 'Price,' is loaded into a pandas DataFrame (`df`).
- Select Features and Target Variable: Relevant features for prediction, in this case, 'Year,' 'Kms_driven,' and 'fuel_type,' are selected and stored in the variable `X`. The target variable, 'Price,' is stored in `y`.
- Split the Dataset: The dataset is split into training and testing sets using the `train_test_split` function from scikit-learn. This is crucial for evaluating the model's performance.

Model

```
x=car.drop(columns='Price')
y=car['Price']
```

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

Train and Test Code snippet

- Create and Train the Model: A linear regression model is created using the `LinearRegression` class. The model is trained on the training set (`X_train` and `y_train`) using the `fit` method.
- Make Predictions: The trained model is used to make predictions on the test set (`X_test`), and the predicted values are stored in `y_pred`.
- Visualize Predictions: The code includes a simple visualization using matplotlib to plot the actual prices against the predicted prices. This allows for a qualitative assessment of the model's performance.
- Print Metrics: Finally, the R-squared value are printed to quantify the model's accuracy.

RESULT

```
lr=LinearRegression()
```

```
pipe=make_pipeline(column_trans,lr)
```

```
pipe.fit(x_train,y_train)
y_pred=pipe.predict(x_test)
```

```
r2_score(y_test,y_pred)
```

0.6860850839287258

predict specific car price

```
pipe.predict(pd.DataFrame(columns=x_test.columns,data=np.array(['Hyundai Santro Xing','Hyundai',2017,100,'Diesel']).reshape(1,5)))
```

array([515651.07573796])

CONCLUSION

For a lot of people, owning the car of their dreams is a major objective. Our goal is to enable customers to make educated decisions about cars at fair and accurate prices by integrating our sophisticated model, shielding them from potential abuse by dishonest dealers.

Large car companies can also benefit greatly from this model, which gives them accurate forecasts that help them create the best possible pricing strategies. This helps them avoid needless difficulties and streamlines their business operations, which ultimately saves them a lot of time and money. In a healthy market, the significance of accurate car pricing cannot be emphasized, and our model is dedicated to advancing fairness and transparency throughout the car-buying process.

FUTURE WORK

Find the Best Algorithms to Increase Model Accuracy: Exploring diverse machine learning algorithms. We will compare the performance of different models such as decision trees, random forests, and deep learning architectures. This comprehensive assessment aims to identify the most effective algorithm for our specific task, ensuring accurate and robust predictions in the dynamic realm of car pricing.

Constant Improvement of Prediction Model Score: Focus on continuously improving the prediction model score by exploring various random states. This involves systematically testing different randomization seeds to ensure robustness and identify the configuration that yields the best performance. By iterating on random states, we aim to enhance the model's reliability and effectiveness in capturing underlying data patterns.

Deploy Prediction App on Streamlit: Plan to deploy our car price prediction model using Streamlit, a popular Python web application library. This user-friendly app will enable individuals to input car details and receive real-time pricing estimates. The model will be packaged for Streamlit compatibility, hosted on a server for easy access, and refined based on user feedback for enhanced usability.

REFERENCE

1. Flach, P. (2015). Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge

2. DATA : https://www.kaggle.com/datasets/harshalchaudhari96/quicker-car/

3. https://www.javatpoint.com/machine-learning