University of New Haven
TAGLIATELA COLLEGE OF ENGINEERING
Electrical & Computer Engineering and Computer Science

TECHNICAL REPORT PAPER
ON
CAR PRICE PREDICITION

REPORT FEATURES

COVER PAGE

TITLE: **CAR PRICE PREDICTION**

PREPARED BY:

TEAM 10

AKSHAY KUMAR	TEAM LEAD	anagi2@unh.newhaven.edu
YESWANTH GOLLA	DATA ANALYST	cgoll2@unh.newhaven.edu
DEEPA ANABATHULA	DATA SCIENTIST	danab1@unh.newhaven.edu
CHUKWUMA IJEOMA	DATA ENGINEER	ichuk2@unh.newhaven.edu

SUBMITTED TO:

TAGLIATELA COLLEGE OF ENGINEERING

DATE:

DECEMBER 5TH, 2023

ABSTRACT

The objective of this project is to develop an accurate car price prediction model using machine learning algorithms. The model is deployed on a Flask web application, which offers users a seamless experience for obtaining estimated car prices. The model's predictive capabilities are based on key features such as vehicle mileage, manufacturing year, fuel consumption, and transmission type.

INTRODUCTION

To secure a great deal on a new car, having access to comprehensive data is crucial. This project utilizes the dataset sourced from CarDekho.com, a well-known car website, which contains a wealth of information on various vehicles, including both selling and current prices. The main objective is to use this dataset for insightful exploratory data analysis, followed by building a robust predictive model using the Random Forest method. The goal is to uncover the influential factors that determine the price of a car, providing users with a valuable tool for making informed decisions.

Data Exploration:

The first step involves a thorough exploration of the CarDekho.com dataset. Exploratory Data Analysis (EDA) is a critical process to understand the underlying patterns, distributions, and relationships within the data. Through this analysis, we gain insights into the various features that contribute to the pricing dynamics of the vehicles in the dataset. Key variables such as mileage, manufacturing year, fuel efficiency, and other relevant factors are examined to understand their impact on car prices. (Nawale, 2021)

Linear Regression:

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a “least squares” method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).

Flask Deployment:

After successful development and validation of the Random Forest model, the next step is to deploy it using Flask. This allows users to input their desired car features and receive a predicted price based on the model's analysis.

Link: https://github.com/GROUP10-DATA-ENGINEERING-UNH/group10-car-price-detection

PROJECT HIGHLIGHT

- The most impressive aspect of the project is the seamless integration of cutting-edge technology with a user-friendly interface. By incorporating a sophisticated machine learning model into a simple Flask web application, the project has not only expanded the possibilities of machine learning but also made predictive technologies more accessible to a broader range of users.

- The model empowers users to make informed decisions by offering predictions based on the input parameters. This can prove invaluable in situations where comprehending the key factors that impact outcomes is of utmost importance.

- The Random Forest model provides a deeper comprehension of the elements that impact car prices, granting users valuable insights that can prove advantageous in different scenarios, whether it's making informed buying choices or maximizing selling opportunities.

LITERATURE MODEL REVIEW

Car price prediction involves a range of modeling techniques, from traditional regression models to advanced machine learning algorithms and deep learning approaches. Each method has its advantages and limitations, and the choice of model depends on the complexity of the data relationships and interpretability requirements. (Tarique, 2020)

Linear and polynomial regression models offer simplicity and interpretability, but they may struggle with non-linear relationships. Machine learning algorithms like decision trees and random forests address non-linearity and over-fitting issues, providing improved accuracy (Abdulla , 2021). Deep learning approaches, such as neural networks, excel at capturing complex patterns but often lack interpretability.

The shift from classical regression models to more advanced machine learning and deep learning approaches reflects the ongoing pursuit of better predictive accuracy in car price prediction. Future research may focus on hybrid models that combine the strengths of different methods, achieving a balance between accuracy and interpretability in the dynamic field of predicting car prices.

METHODOLOGY

Data Collection:

- A diverse data-set containing relevant information about used cars was acquired.

- Features such as mileage, manufacturing year, fuel consumption, and transmission type were included.

```
final_dataset=df[['Year','Selling_Price','Present_Price','Kms_Driven','Fuel_Type','Seller_Type','Transmission','Owner']]
```

```
final_dataset.head()
```

	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0

Data Preprocessing:

- Missing data and outliers were handled to ensure data quality.

- Normalization and encoding techniques were applied for compatibility with machine learning algorithms.

```
##check missing values
df.isnull().sum()
```

```
Car_Name         0
Year             0
Selling_Price    0
Present_Price    0
Kms_Driven       0
Fuel_Type        0
Seller_Type      0
Transmission     0
Owner            0
dtype: int64
```

Model Selection:

- Various regression models, including linear regression and ensemble methods, were explored.

- The model that demonstrated optimal performance on the data-set was selected.

```
from sklearn.linear_model import LinearRegression
```
[74]

```
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import make_column_transformer
from sklearn.pipeline import make_pipeline
from sklearn.metrics import r2_score
```
[75]

Model Training:

- The data-set was split into training and testing sets.

- The chosen machine learning model was trained on the training set.

Applying Train Test Split

In [35]:
```python
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)
```

In [74]:
```python
from sklearn.linear_model import LinearRegression
```

In [75]:
```python
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import make_column_transformer
from sklearn.pipeline import make_pipeline
from sklearn.metrics import r2_score
```

Creating an OneHotEncoder object to contain all the possible categories

In [39]:
```python
ohe=OneHotEncoder()
ohe.fit(X[['name','company','fuel_type']])
```

Out[39]: OneHotEncoder()

Creating a column transformer to transform categorical columns

In [52]:
```python
column_trans=make_column_transformer((OneHotEncoder(categories=ohe.categories_),['name','company','fuel_type']),
                                    remainder='passthrough')
```

Linear Regression Model

In [54]:
```python
lr=LinearRegression()
```

Flask Integration:

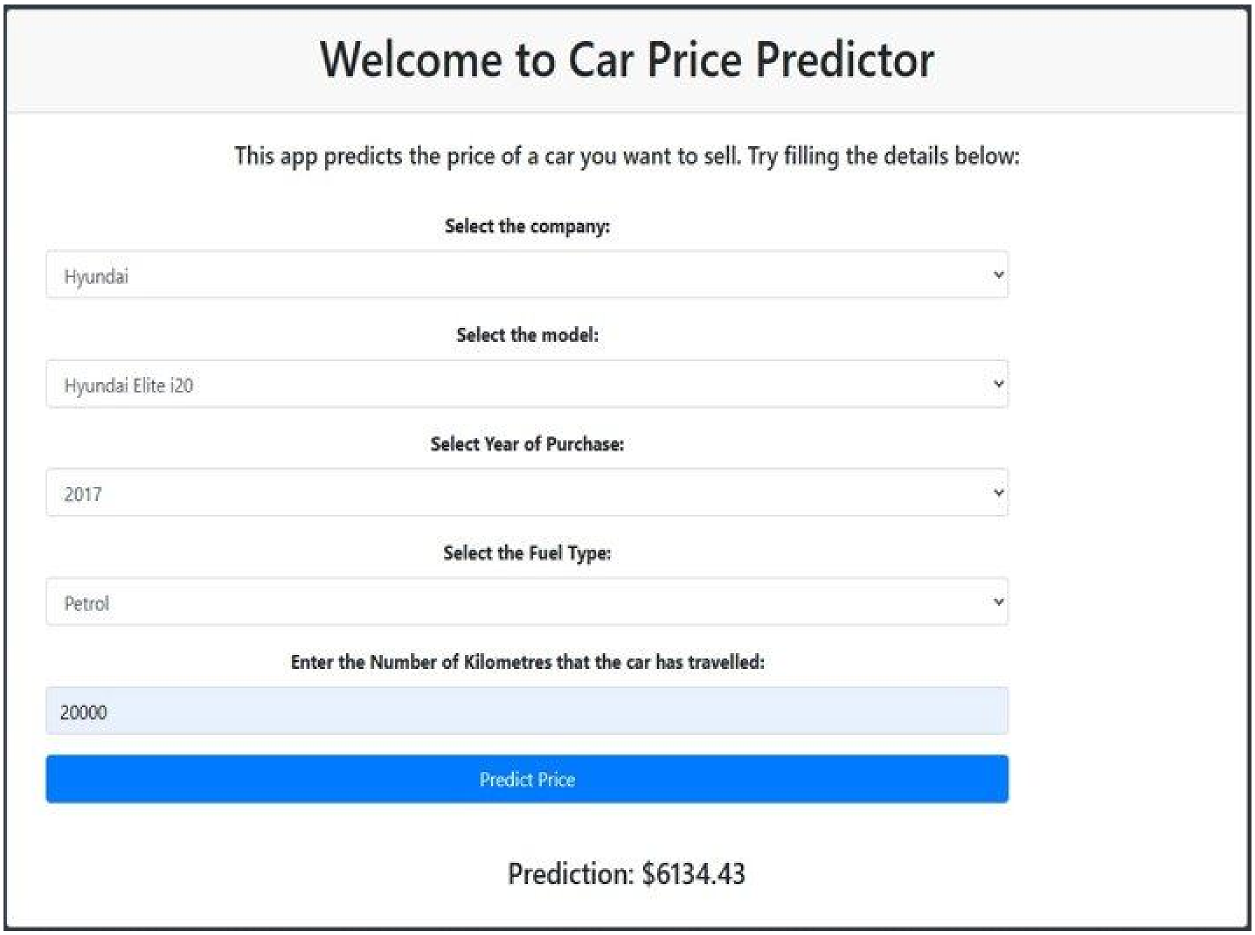
- A user-friendly interface was developed using Flask to capture user inputs.

- The trained model was integrated into the Flask application for real-time predictions.

```
C:\Users\babun\AppData\Local\Programs\Python\Python37\lib\site
ckle estimator Pipeline from version 0.22 when using version 0
. Use at your own risk.
  UserWarning)
 * Serving Flask app "application" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a p
   Use a production WSGI server instead.
 * Debug mode: off
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
127.0.0.1 - - [05/Dec/2023 22:28:53] "GET / HTTP/1.1" 200 -
[658422.22397539]
127.0.0.1 - - [05/Dec/2023 22:29:04] "POST /predict HTTP/1.1"
127.0.0.1 - - [05/Dec/2023 22:31:35] "GET / HTTP/1.1" 200 -
[466689.80957898]
127.0.0.1 - - [05/Dec/2023 22:32:14] "POST /predict HTTP/1.1"
127.0.0.1 - - [05/Dec/2023 22:33:13] "GET / HTTP/1.1" 200 -
[350552.28766014]
127.0.0.1 - - [05/Dec/2023 22:34:28] "POST /predict HTTP/1.1"
[350552.28766014]
127.0.0.1 - - [05/Dec/2023 22:35:19] "POST /predict HTTP/1.1"
```

RESULT

The model achieved commendable accuracy in predicting car prices. Users can input details such as mileage, manufacturing year, fuel consumption, and transmission type, receiving accurate and timely predictions.



CONCLUSION

The outcomes highlight the venture's prosperity, uncovering excellent precision in anticipating vehicle costs through the blend of cutting-edge innovations. Users can effortlessly navigate the complexities of estimating car values based on multiple features thanks to the user interface, which was designed for simplicity but is equipped with the sophistication of machine learning.

To sum up, this venture is a remarkable instance of the convergence of user-centric design and innovation. The creation of a sturdy, precise, and easily accessible framework for forecasting automobile prices showcases the essence of technological progress and suggests a future where predictive analytics will not be restricted to specialists but rather become a tool that empowers all. This initiative, with its robust plan, compelling results, and customer-oriented approach, paves the way for a new era in the democratization of AI applications.

REFERENCES

References

Abdulla , A. (2021). Used Cars Price Prediction and Valuation using Data Mining Techniques.

Akhtar, T. (2020, November 23). Retrieved from Predicting Car Price using Machine Learning: https://towardsdatascience.com/predicting-car-price-using-machine-learning-8d2df3898f16

Nawale, T. (2021, March 30). *Used Carprice Prediction -Complete Machine Learning Project*. Retrieved from https://medium.com/geekculture/used-carprice-prediction-complete-machine-learning-project-d25559cf2d2a

Tarique, A. (2020, November 30). *Predicting Car Price using Machine Learning*. Retrieved from https://towardsdatascience.com/predicting-car-price-using-machine-learning-8d2df389f16