***Final AI Assignment***

**-Car Price Prediction (Proposal and analysis of results)**

Jess George- 4349612

Devika Samji- 4348590

ADMN5016-W23-101-2609-K: Applied Artificial Intelligence and Machine Learning

Business Analytics International

Prof. Joseph Santarcangelo

**Business Problem**

We are modelling the Car price dataset to predict the price of cars. It can be used to understand how exactly the prices vary with the independent variables and can accordingly manipulate the business strategy to meet certain price levels. Thus, the target variable is price, and the predictor variables include fuel type, engine type, horsepower, citympg, etc. (categorical and numerical). We are determining which variables are significant in predicting the price of a car and how well those variables describe the price of a car. Below is the summary of variables in the dataset.

Table

Description automatically generated

**Savings and Benefits of Car price prediction**

* Better pricing strategy and increased sales: It can help a car seller give right price of cars to potential buyers and maximise profits.
* Reduced time to sell: It helps in reducing the time taken to sell a car. If a car is priced too high, then, then it can take more time to get sold and if it is priced too low, then it can minimise the profits. Therefore, car price prediction helps business to decrease the time for a car sale.
* Improved customer satisfaction: Selling cars at the right price, make buyers happy and can lead to positive word of mouth recommendations.
* Competitive advantage: A seller who provides accurate car price can stand out from competitors and can attract more buyers.

**Performance of Machine learning models:**

1. Linear Regression

A linear regression model was performed to predict the price of the cars. This model was selected because the target variable is numeric and continuous. Standardisation of features was done before implementing the model since linear regression assumes normal distribution of dataset. It provided R squared value of 0.86 which indicates that approximately 87% of the variance in the dependent variable can be explained by the independent variables, which is a good fit for the dataset.

1. K-Nearest Neighbours (KNN)

Another model chosen was K-Nearest Neighbours as it can handle both categorical and numerical data, and it can work well for datasets with nonlinear relationships between the features and the target variable. It provided R squared value of 0.72 which shows that 72% of variance in the price can be predicted by the given features. While comparing with linear regression, the accuracy of this model was not good enough.

1. Random Forest Regressor

We also used the model, Random Forest Regressor as it combines multiple decision trees to give the final prediction. This helps to reduce overfitting of data and improves accuracy. The results of the model showed an accuracy of 90% which means that the model performed better than linear regression and KNN to predict the car price. Overall, the model was able to explain the large proportion of variance in the target variable but there were significant errors. So, using hyperparametric tuning was important to avoid overfitting and improve accuracy. We used the max\_depth to give the depth of decision trees in the model, n\_estimators to include the number of trees in the model and criterion as MSE (Mean Squared error). This made the model to reduce errors to an extent and improve the accuracy to 91%.

**Risks and limitations**

There are several disadvantages or risks associated with the above business proposal. The machine learning models can only predict prices of cars by using the available data. It cannot consider various factors like market volatility, geographical and temporal variation, etc. Changes in supply and demand, consumer preferences, and economic conditions are the factors included in market volatility. Avoiding these factors can lead to inaccurate price predictions. Also, the price of cars can vary according to various location and geographical regions. Moreover, it is also impacted by subjective factors such as the condition of the car, its history, or its sentimental value to the owner.

**Github link for the python notebook:**

[**https://github.com/JessPhil**](https://github.com/JessPhil)