|  |
| --- |
| TECHNICAL REPORT TEMPLATE |

|  |
| --- |
| Electrical & Computer Engineering & Computer Science (ECECS) |

|  |  |
| --- | --- |
| SPRING 22 |  |



Contents

[Project Name 2](#_Toc96341549)

[Executive Summary 2](#_Toc96341550)

[Technical Report 3](#_Toc96341551)

[Highlights of Project 3](file:///C:\Users\ardia\Desktop\TEACHING\SPRING%2022\DSCI6002%20Data%20Science\Projects\Technical%20report%20Template.docx#_Toc96341552)

[Submitted on: 3](file:///C:\Users\ardia\Desktop\TEACHING\SPRING%2022\DSCI6002%20Data%20Science\Projects\Technical%20report%20Template.docx#_Toc96341553)

[Abstract 4](#_Toc96341554)

[Methodology 5](#_Toc96341555)

[Results Section 5](#_Toc96341556)

[Discussion 6](#_Toc96341557)

|  |
| --- |
| Forecasting the Value of Pre-Owned Vehicles |

|  |  |  |
| --- | --- | --- |
| Executive Summary This project addresses the dynamic pricing challenge in the pre-owned vehicle market. The aim is to model the cost of cars using various independent variables, thereby understanding the relationship between these variables and price variations. Key technologies employed include Python 3.7, Jupyter Notebook, Flask Framework, HTML, CSS, JS, and AWS Cloud. The project leverages advanced data engineering and machine learning techniques, with a focus on the Random Forest Model for its accuracy in predicting vehicle prices. | | |
| person at a table writing in a notebook with people around | | |
| **Team Members:**  Chaitanya Gopinadh Madala  Divya Mamuru  SaiKumar Dacharla  Nikhith Krishna Vinduru | **Questions?**  Contact :  [sdach13@unh.newhaven.edu](mailto:sdach13@unh.newhaven.edu)  [nvind1@newhaven.edu](mailto:nvind1@newhaven.edu)  [dmamu2@unh.newhaven.edu](mailto:dmamu2@unh.newhaven.edu) |  |

|  |
| --- |
| Technical Report |

|  |  |
| --- | --- |
| Forecasting the Value of Pre-Owned Vehicles | Additionally, cars sold by dealers were found to have higher selling prices than those sold by individuals. These insights are crucial for stakeholders in the pre-owned vehicle market, aiding in informed decision-making.  Submitted On:  05-12-2023 |
| Highlights of Project The project on forecasting the value of pre-owned vehicles delivered pivotal insights into the used car market, revealing key factors influencing vehicle pricing. Advanced data engineering and machine learning techniques were utilized, with the Random Forest Model identified for its predictive accuracy. Analysis showed a direct correlation between a car's current value and its selling price; vehicles with higher present values fetched higher selling prices. The age of a car was inversely related to its price, with newer cars commanding higher prices. Fuel type also played a role, with diesel cars selling at a premium. Transmission type was another price determinant; automatic cars were pricier than manual ones. |

## Abstract

This project addresses the dynamic pricing challenge in the pre-owned vehicle market. The aim is to model the cost of cars using various independent variables, thereby understanding the relationship between these variables and price variations. Key technologies employed include Python 3.7, Jupyter Notebook, Flask Framework, HTML, CSS, JS, and AWS Cloud. The project leverages advanced data engineering and machine learning techniques, with a focus on the Random Forest Model for its accuracy in predicting vehicle prices.

**Git Hub Link:** **https://github.com/orgs/SaiDivyaKrishnaChai/teams**

Cover Page

Forecasting the Value of Pre-Owned Vehicles

Chaitanya Gopinadh Madala

Divya Mamuru

SaiKumar Dacharla

Nikhith Krishna Vinduru

A logo of a university of new haven

Description automatically generated

Executive Summary

Introductory Section

The pre-owned vehicle market faces the challenge of accurately determining vehicle values, influenced by factors like make, model, age, mileage, and condition. This project aims to develop a scalable model that provides real-time, accurate valuation of used vehicles. This model is crucial for both buyers and sellers, assisting them in making informed decisions.

Review of available research

## The project presents a comprehensive analysis of factors influencing the resale value of pre-owned vehicles, utilizing a data-driven approach to model price variations. Advanced technologies such as Python, Jupyter Notebook, and AWS Cloud were adeptly employed to process and analyze data, with the Random Forest Model emerging as a superior predictive tool due to its accuracy. The project's findings are robust, with clear evidence showing the impact of mileage, vehicle age, engine volume, and ownership history on selling prices. The visualizations, including heatmaps and distribution charts, effectively convey the nuances of the used car market, illustrating the depreciation of value with increased mileage and age, and the preference for newer, less-used vehicles. Overall, the project stands out for its practical applications in aiding informed decision-making for buyers and sellers in the pre-owned vehicle market.

## 

## Methodology

The methodology encompasses gathering data from various sources including historical sales records, online listings, and manufacturer specifications. This data is then processed into a cohesive format for training the forecasting model. The project focuses on selecting relevant features that significantly impact vehicle valuation. The Random Forest Model is emphasized for its high accuracy in dealing with complex datasets and multiple variables.

## Results Section

## A screenshot of a color scheme Description automatically generated

There are three variables: Mileage, EngineV (presumably Engine Volume), and log\_price. The color intensity reflects the strength of the correlation between these variables. The correlation coefficient is a measure that determines the degree to which two variables move in relation to each other, with values ranging from -1 to 1. A value of 1 implies a perfect positive correlation, -1 a perfect negative correlation, and 0 no correlation.

Looking at the heatmap:

There is a strong negative correlation between Mileage and log\_price, indicated by a darker color and a value of approximately -0.61. This suggests that as the mileage on a vehicle increases, the logarithm of its price tends to decrease.

The correlation between EngineV and log\_price is moderately positive, around 0.41, suggesting that vehicles with larger engine volumes tend to have higher prices.

There is a negligible correlation between Mileage and EngineV, shown by the neutral color and the correlation coefficient being close to 0. This indicates that there is no significant linear relationship between a vehicle’s mileage and its engine volume.

A comparison of a graph

Description automatically generated with medium confidence

In the first chart, "Selling Price Vs Owner," the selling price is plotted against the number of previous owners a vehicle has had. The categories of owners are labeled as 0, 1, and 3. It can be observed that vehicles with 0 previous owners, which are likely new or like-new, fetch the highest selling prices, whereas the selling price seems to decrease with the number of owners, with vehicles that have had 3 previous owners exhibiting the lowest selling prices. This trend indicates that the perceived value of a vehicle diminishes with an increasing number of owners, which could be due to factors like increased wear and tear or the perception of reduced reliability.

The second chart, "Selling Price Vs Car Age," shows the selling price in relation to the car age, with the ages ranging from 2 to 17 years. The chart indicates a general trend of decreasing selling price with increasing car age. This is consistent with the common understanding that the value of a car depreciates over time. Interestingly, there is notable variance in selling prices among cars of the same age, which suggests that other factors, in addition to age, also significantly influence the selling price. The error bars indicate the variability of the selling price within each age group, highlighting that while age is a significant factor, individual vehicle conditions and other attributes can cause substantial differences in pricing.

Both charts together suggest that newer, less-used vehicles tend to have a higher selling price, reflecting the common consumer preference for newer and less-owned vehicles in the pre-owned car market. These visualizations are useful for buyers and sellers to understand how the number of previous owners and the age of a car could affect its market value.

A graph of a graph

Description automatically generated with medium confidence

The first plot, "Selling Price Vs Present Price," is a scatter plot with a trend line, indicating a positive correlation between a vehicle's present price and its selling price. This suggests that vehicles with a higher present value tend to be sold for more, which aligns with market expectations where a vehicle's current market value influences its resale value.

The middle plot, "Distribution of Selling Price," is a kernel density estimation (KDE) that shows the distribution of vehicle selling prices. The KDE plot reveals a distribution with multiple peaks, suggesting that the selling prices of vehicles might have several common values or clusters, possibly indicating different groups or segments in the vehicle market, such as economy, mid-range, and luxury vehicles.

The third plot, "Distribution of KMS Driven," is another KDE plot illustrating the distribution of the kilometers driven by the vehicles. The distribution appears to be right skewed, with most of the data concentrated on the left, indicating that a larger number of vehicles have lower mileage, with fewer vehicles having been driven for longer distances. This skewness could reflect that vehicles with lower mileage are more common in the market or that they are more likely to be sold.

The model reveals several key insights:

- The current value of a car significantly influences its selling price, with a direct correlation observed.

- The age of a car negatively impacts its selling price, with older vehicles commanding lower prices.

- Diesel-fueled cars have higher selling prices compared to other fuel types.

- Automatic cars tend to have higher prices than manual cars.

- Vehicles sold by dealers fetch higher selling prices than those sold by individuals.

## Discussion

In our analysis, the Random Forest Model emerged as a standout performer, demonstrating superior accuracy in predicting the selling prices of pre-owned vehicles. This model's efficacy lies in its ability to handle large datasets with multiple input variables effectively, making it particularly suited for our complex dataset encompassing diverse factors such as age, make, model, and mileage of vehicles. Its robustness against overfitting, coupled with its capacity for capturing nonlinear relationships and interactions between variables, significantly contributed to its higher predictive accuracy. Furthermore, the Random Forest Model's feature importance analysis provided valuable insights into which factors most strongly influenced vehicle valuation, guiding our data-driven decision-making process. These attributes underscore the model's suitability for addressing the intricate dynamics of the pre-owned vehicle market, making it a reliable tool for both our predictive analytics and strategic planning purposes.

## Conclusion

The culmination of our project on 'Forecasting the Value of Pre-Owned Vehicles' has yielded significant insights into the factors influencing the pricing of used cars. Our findings reveal a complex interplay of variables such as a vehicle's age, fuel type, transmission type, and the nature of the seller, each playing a crucial role in determining its market value. Notably, the current value and age of a vehicle are inversely related to its selling price, highlighting the depreciation factor in the pre-owned market. Diesel cars generally command higher prices, indicating a market preference for fuel efficiency and longevity. Additionally, the distinction between manual and automatic transmission vehicles has shown a clear preference in pricing, with automatic cars often fetching higher prices.

The study also unveiled the impact of the seller type on the selling price, with vehicles sold by dealers commanding higher prices than those sold by individuals. This aspect underscores the trust and reliability factor associated with dealer-sold vehicles.

Our use of the Random Forest Model has demonstrated a robust approach to addressing the multifaceted nature of this market. The model's accuracy in predicting prices, based on a comprehensive set of variables, provides a solid foundation for buyers and sellers to make informed decisions.

The implications of our study are far-reaching. For buyers, this provides a transparent framework to assess the fair market value of pre-owned vehicles, enabling better negotiation and purchasing decisions. For sellers, including dealers and individuals, these insights can guide pricing strategies to align with market trends and buyer preferences. Furthermore, for market analysts and automotive industry stakeholders, this research offers a detailed understanding of market dynamics, assisting in strategic planning and forecasting.

In conclusion, our project not only enhances the transparency and efficiency of the pre-owned vehicle market but also paves the way for data-driven strategies that can adapt to evolving market conditions, benefiting all stakeholders involved.