**Car Price Predicting Model**

Logo

Description automatically generated

**CS (AI – ML)**

|  |  |
| --- | --- |
| **(Roll. No. - 2101321530003, Name – Aditya Mishra**  **Roll. No. - 2101321530008, Name – Anuj Mishra)** | **Project Supervisor/Guide Name –**  **Mrs. Rachana Singh Sisodia** |

**Greater Noida Institute of Technology, Greater Noida**

**Dr. A.P.J. Abdul Kalam Technical University, Lucknow**

October, 2022

**Car Price Predicting Model**

**PROJECT SYNOPSIS**

For

**BACHELOR OF TECHNOLOGY**



(Roll No.- 2101321530003, SUPERVISOR/GUIDE NAME-

Name of Student- Aditya Mishra) Mrs. Rachana Singh Sisodia

(Roll No.- 2101321530007,

Name of Student- Anuj Mishra)

**Greater Noida Institute of Technology, Greater Noida**

**D**

**Dr. A.P.J. Abdul Kalam Technical University, Lucknow**

October, 2022

## **Index**

1. Cover Page
2. Declaration
3. Acknowledgement
4. Abstract
5. Index
6. Introduction
   1. Problem Statement
   2. Scope of Project
7. Tools/Environment Used
8. Analysis Document
   1. E-R diagrams/Class diagrams/any related diagrams
   2. Data flow diagrams/other similar diagrams,

.Limitations of the Project

1. Result and Future Scope of the Project

**Declaration**

I hereby declare that this project work entitled “Car Price Predicting Model”

has been prepared by me during the year 2020 – 21 under the guidance of Mrs. Rachana Singh Sisodia, Department of CS (AI – ML), GNIOT, Greater Noida in the partial fulfillment of Btech degree prescribed by the college.

I also declare that this project is the outcome of my own effort, that it has not been submitted to any other university for the award of any degree.

Date: 10/10/2022

Aditya Mishra

2101321530003

Anuj Mishra

2101321530007

**Acknowldegement**

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to Mrs. Rachana Singh Sisodia for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I would like to express my gratitude towards my parents & member of GNIOT for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

**Introduction**

### What is machine learning?

Machine learning (ML) is a type of artificial intelligence ([AI](https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence)) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning [algorithms](https://www.techtarget.com/whatis/definition/algorithm) use historical data as input to predict new output values.

[Recommendation engines](https://www.techtarget.com/whatis/definition/recommendation-engine) are a common use case for machine learning. Other popular uses include fraud detection, spam filtering, malware threat detection, [business process automation](https://www.techtarget.com/searchcio/definition/business-process-automation) (BPA) and Predictive maintenance.

### Why is machine learning important?

Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

### What are the different types of machine learning?

Classical machine learning is often categorized by how an algorithm learns to become more accurate in its predictions. There are four basic approaches:[supervised](https://www.techtarget.com/searchenterpriseai/definition/supervised-learning) learning, [unsupervised](https://www.techtarget.com/whatis/definition/unsupervised-learning) learning, semi-supervised learning and reinforcement learning. The type of algorithm data scientists choose to use depends on what type of data they want to predict.

* Supervised learning: In this type of machine learning, [data scientists](https://www.techtarget.com/searchenterpriseai/definition/data-scientist) supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.
* Unsupervised learning: This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through data sets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.
* Semi-supervised learning: This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labeled [training data](https://www.techtarget.com/searchenterpriseai/feature/Using-small-data-sets-for-machine-learning-models-sees-growth), but the model is free to explore the data on its own and develop its own understanding of the data set.
* Reinforcement learning: Data scientists typically use [reinforcement learning](https://www.techtarget.com/searchenterpriseai/definition/reinforcement-learning) to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

### How does supervised machine learning work?

Supervised machine learning requires the [data scientist](https://www.techtarget.com/searchbusinessanalytics/feature/Key-differences-of-a-data-scientist-vs-data-engineer) to train the algorithm with both labeled inputs and desired outputs. Supervised learning algorithms are good for the following tasks:

* **Binary classification:**Dividing data into two categories.
* **Multi-class classification:**Choosing between more than two types of answers.
* **Regression modeling:** Predicting continuous values.
* **Ensembling:** Combining the predictions of multiple machine learning models to produce an accurate prediction.

Artificial Intelligence is an integral part of all major e-commerce companies today. With the evolution of the information industry and extensive research in the field of AI in the past two decades, businesses have started to explore the ways to automate various activities using state of the art Machine Learning algorithms and Deep Neural Networks. Many IT giants and start-ups have already taken a big leap in this field and have dedicated teams and resources for research and development of cutting edge AI applications. Online retail platforms today are extensively driven by AI-powered algorithms and applications. Activities ranging from inventory management and quality checking at the warehouse to product recommendation and sales demographics on the website, all employ machine learning at various scales.

## **What is price forecasting and how is it done**

**Price forecasting** is predicting a commodity/product/service price by evaluating various factors like its characteristics, demand, seasonal trends, other commodities’ prices (i.e. fuel), offers from numerous suppliers, etc.

Price forecasting may be a feature of consumer-facing travel apps, such as Trainline or Hopper, used to increase customer loyalty and engagement. At the same time, other businesses may also use information about future prices. Entrepreneurs may need to define an optimal time to buy a commodity to adjust prices of products or services that require a commodity (lumber, coffee, gold), or evaluate the investment appeal of fixed assets.

Price prediction can be formulated as a regression task. Regression analysis is a statistical technique used to estimate the relationship between a dependent/target variable (electricity price, flight fare, property price, etc.) and single or multiple independent (interdependent) variables AKA predictors that impact the target variable. Regression analysis also lets researchers determine how much these predictors influence a target variable. In regression, a target variable is always numeric.

In general, price forecasting is done by the means of descriptive and predictive analytics.

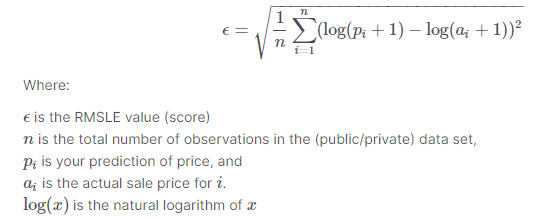
# ML problem

A Car company has provided user-inputted text descriptions of its products, including details like product category name, brand name, item condition and other required data or feature of a car. Using this data, we have to come up with a model that predicts the price of a car listed on car company as accurately as possible. This looks like a standard regression problem.

Tools/Environment Used

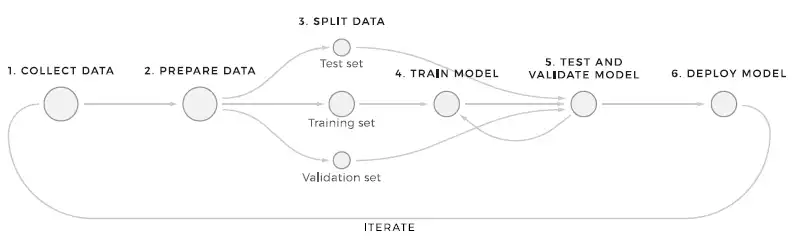
# Performance metrics and Business constraints

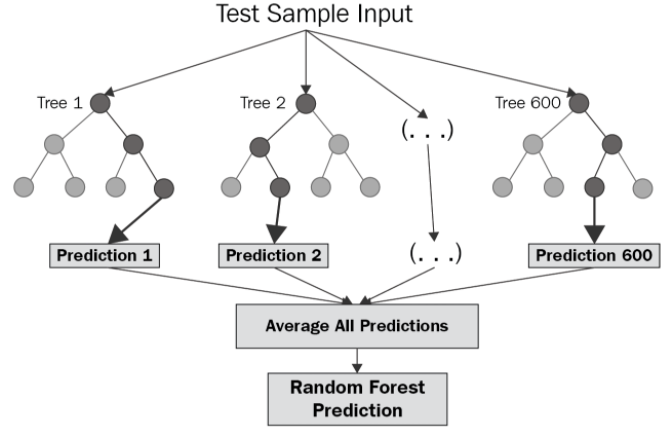
We must have a yardstick to measure how good or bad our model’s performance is. In machine learning terminology, we call this yardstick performance metric or simply metric. There are various metrics to measure the performance of a regression model, e.g. mean absolute error, mean squared error, mean squared logarithmic error, maximum residual error, median absolute error, coefficient of determination(R²), etc.  
For this problem, Kaggle uses **Root Mean Squared Logarithmic Error(RMSLE)**. Lesser the RMSLE, better is our model. RMSLE is calculated as



Analysis Document

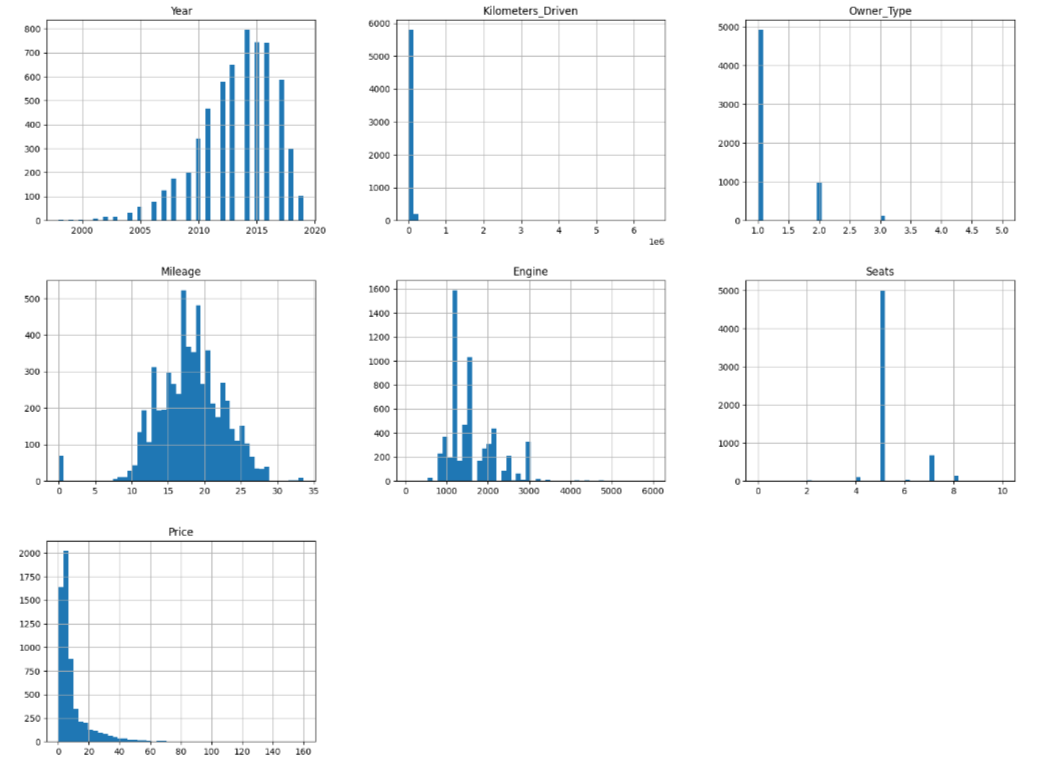
1.





Data Overview

2.



Timeline, box and whisker chart

Description automatically generated

Result

We will built a Price predicting module based on supervised machine learning, with the help of various regression tools.

Though this is the most simple model we’ve to built, the final predictors still seem to have high correlations. One can go ahead and remove some of these features, though that will affect the adjusted-r2 score significantly.

Thus, for now, the final model consists of the 7 variables mentioned above.

# Future Work

* Using deep learning was productive and yielded a very good score on test data. More complex models such as LSTMs and Convolutional Neural Nets can be tried.
* We can experiment with more complex MLPs by adding additional layers and larger number of units in hidden layers.
* Other vectorization schemes such as [Wordbatch](https://github.com/anttttti/Wordbatch" \t "_blank) can be experimented with ML models.
* Regression models like [FTRL](https://datatable.readthedocs.io/en/latest/ftrl.html) and [FM\_FTRL](https://medium.com/@dhirajreddy13/factorization-machines-and-follow-the-regression-leader-for-dummies-7657652dce69) can also be tried.

# Future Scope

The scope of Machine Learning is not limited to the investment sector. Rather, it is expanding across all fields such as banking and finance, information technology, media & entertainment, gaming, and the automotive industry.

What is the scope of data science in future? Especially in a developing country like India, there is scope for huge data-related operations such as data scientists, data analytics, big data engineers, big data managers, and data architects. Take up this Advanced Certification in Data Science and AI to make a lucrative career as a data scientist in India.