Used Car finance fit and recommendation

Harishankar Ganapathy, Kunal Kumar, Manisha Nallapati, Pooja Vishwakarma, Ravi Kavidayal

***Abstract*—The problem we are tackling is twofold: providing a reliable used car price prediction for sellers, and assisting customers in checking their loan eligibility, predicting their eligible loan amount, and recommending a suitable used car. We are using different types of Machine learnings (supervised and unsupervised) and algorithms from each category of learning.**

Keywords: Machine learning, Supervised learning, Unsupervised learning,

I. INTRODUCTION

The increasing demand for used cars in the market. According to market analysis, the demand for used cars has been steadily increasing in recent years. In terms of sales volume, the used car market in India sold approximately 4.4 million units in 2020, representing a 12.5% increase from the previous year. One of the primary drivers of this trend is the rising cost of new cars, which has led many consumers to opt for used vehicles instead. Additionally, the COVID-19 pandemic has had a significant impact on the automotive industry, with many people choosing to avoid public transportation and instead purchase their own personal vehicles.

A. MOTIVATION

In recent years, the used car market has experienced a significant rise due to the increasing demand for affordable and reliable vehicles. As a result, many people are turning to used cars as an alternative to new cars. However, one of the biggest challenges that people face when buying a used car is financing. With so many financing options available, it can be difficult to determine the best fit for your needs and financial situation. This is where machine learning comes in.

Machine learning algorithms can be trained to analyse vast amounts of data and identify patterns that are not immediately apparent to humans. By analysing data such as credit history, income, employment status, and vehicle information, machine learning models can make accurate predictions about the likelihood of loan approval and the best financing options for each individual. This can be extremely helpful in providing personalized and targeted recommendations to buyers looking for used car finance.

In this project, we will use machine learning algorithms to analyse customer data and recommend the best financing options for each individual based on their financial situation and the vehicle they are interested in purchasing. Our goal is to provide buyers with a seamless and hassle-free financing experience while also ensuring that they receive the best possible loan terms and rates.

- For sellers, we will analyse a variety of factors, including the vehicle's make, model, year, condition, mileage, etc. in order to provide an accurate price prediction for their used car. This will enable sellers to price their vehicles appropriately, leading to a smoother and more efficient sales process

- For customers, our solution will enable them to quickly and easily check their loan eligibility, predict the amount of loan they may qualify for, and receive personalized recommendations for used cars that fall within their budget.

II. DEFINITION AND TERM

- **Used car finance:** A type of loan used to purchase a used car. Used car finance typically involves a down payment, monthly payments, and interest charges over a specific term.
- **Machine learning (ML):** A type of artificial intelligence that allows computers to learn and improve from experience without being explicitly programmed. ML algorithms can analyze large amounts of data and identify patterns to make predictions and decisions.
- **Recommendation system:** A type of ML algorithm that makes personalized recommendations based on user preferences and behavior. In the context of used car finance, a recommendation system can suggest the best financing options based on an individual's financial situation and the vehicle they are interested in purchasing.
- **Feature engineering:** The process of selecting and transforming the most relevant features or variables in a dataset for use in an ML model. In the context of used car finance, feature engineering could involve selecting variables such as credit score, income, and employment status, and transforming them to be used as inputs in an ML algorithm.
- **Regression analysis:** A statistical technique used to estimate the relationship between a dependent variable (such as the loan amount) and one or more

independent variables (such as income or credit score). Regression analysis can be used to develop ML models that predict loan approval likelihood and loan terms.

- **Data pre-processing:** The process of cleaning, transforming, and organizing raw data before it can be used in an ML model. Data preprocessing can involve removing missing values, scaling and normalizing features, and encoding categorical variables.

III. BACKGROUND

Supervised learning is a type of machine learning in which an algorithm learns to make predictions or decisions by being trained on labeled data. Labeled data refers to data where the desired output is known or labeled. In supervised learning, the algorithm is presented with input data and its corresponding output or label, and the algorithm learns to make predictions or decisions based on this data.

A. Supervised Learning in Loan Prediction and used car price predicction.

Some common supervised learning algorithms used for loan prediction include logistic regression, decision trees, random forests, and support vector machines. These algorithms are trained on a labeled dataset consisting of loan application data and their corresponding approval or denial status.

We have used different regression algorithms to train, fit and predict.

B. UnSupervised Learning for car recommendation

One way to approach this problem using unsupervised learning is to use clustering algorithms. Clustering algorithms group similar data points together based on their features or characteristics. In the case of car recommendation, the features could include factors such as customer's Age, Applicant Annual Income and Credit History

C. Figures and Tables

Table 1: Features and Label used

Type	**Features**	**Label**
Loan Prediction	Gender, Age, Married, No_Of_Dependents, Edu_Qualification, Employment_Status, Applicant_Annual_Income, Co_Applicant_Annual_Income, Applicant_Credit_History, Credit_Rating, Existing_No_of_Loans	Loan_Amount_Availed
Used car price prediction	Name, Location, Year, Kilometer_Driven, Fule_Type, Transmission, Owner_Type, Mileage, Engine, Power, Seats	Price
Car Recommendation	customer's Age, Applicant Annual Income and Credit History	N/A

IV. OBSERVATIONS

Below are the results of different supervised and unsupervised algorithms used in this project:

A. Loan Prediction :

Figure 1: Plot for different regression Model

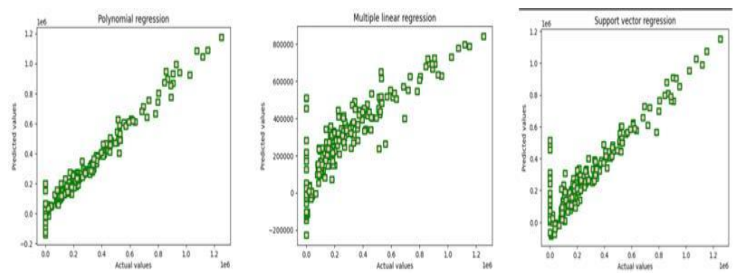


Table 2: Type of model vs R2 score

Type of Model	**R2 score**
Polynomial regression model	0.9703013316360823
Multiple regression model	0.7549258446460376
Support vector regression model	0.872788526395641

Polynomial regression performed better compared to other regression model for loan prediction.

B. Used Car Price Prediction

Figure 2: Different model vs Metrics

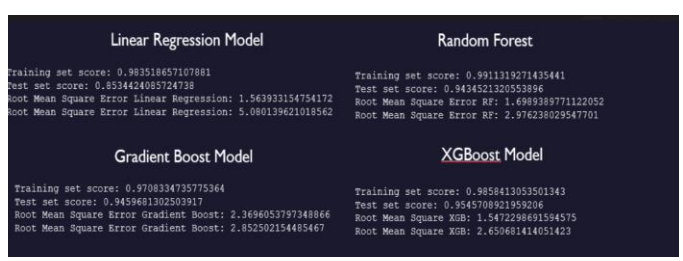


Figure 3: Linear regression Model fit (Random Forest)

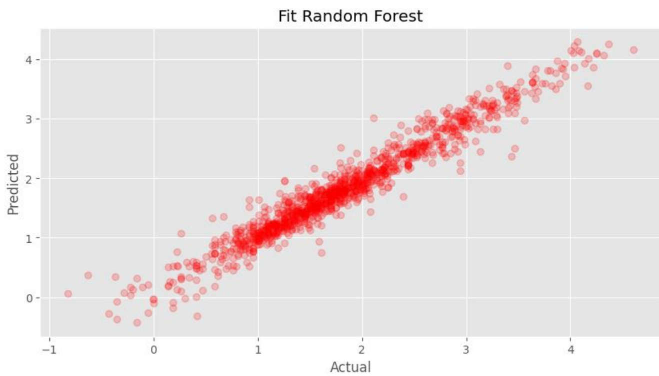


Figure 4: Linear regression Model fit (Gradient Boost)

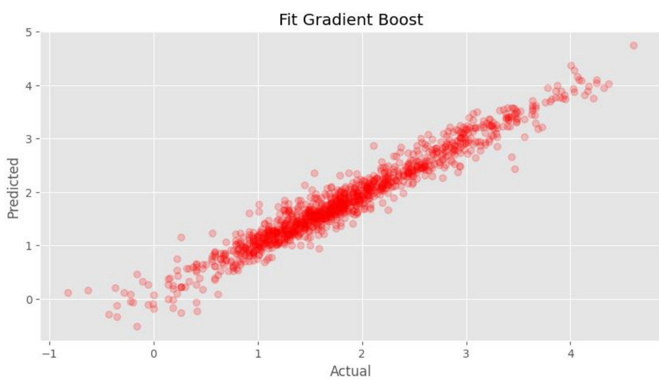


Figure 5: Linear regression Model fit (XG Boost)

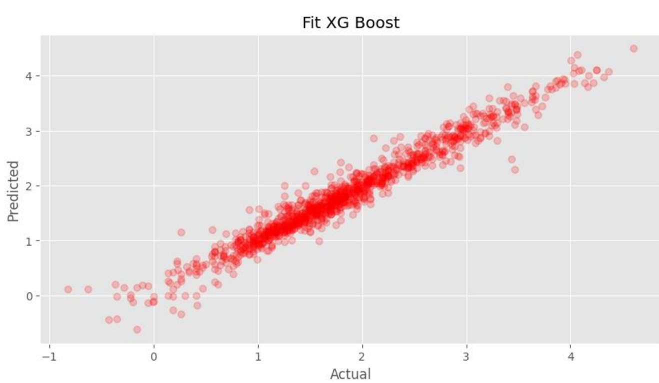
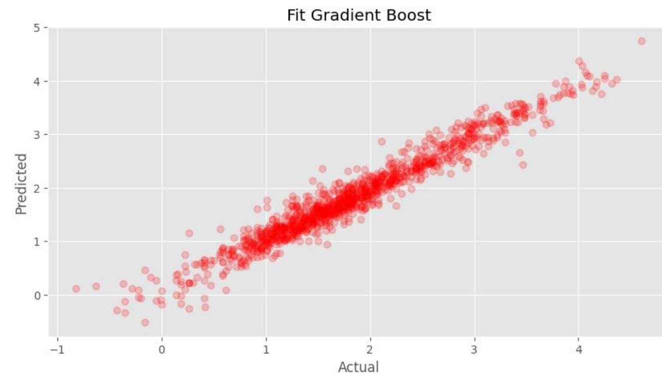


Figure 6: Linear regression Model fit (Gradient Boost)



Gradient boost model preformed good compared to other tested models for used car price prediction.

C. Car recommender System

Figure 7: K- Mean cluster elbow method result (value for feature "profession" is self-employed)

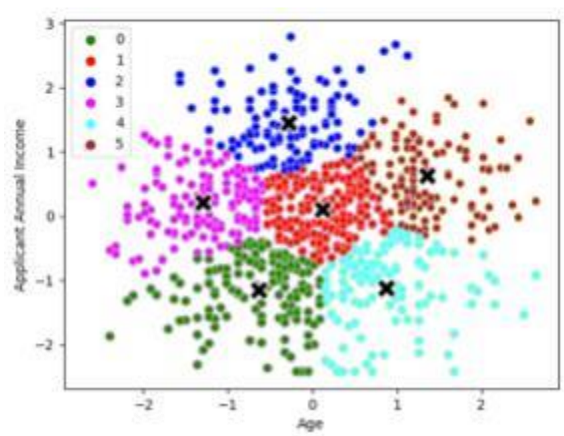


Figure 8: K- Mean cluster elbow method result (value for feature "profession" is Software engineer)

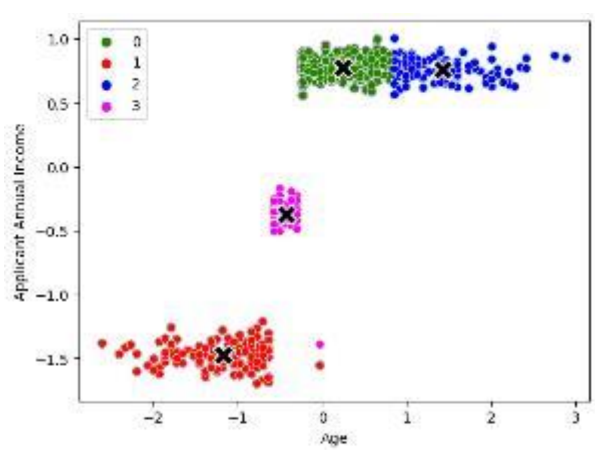
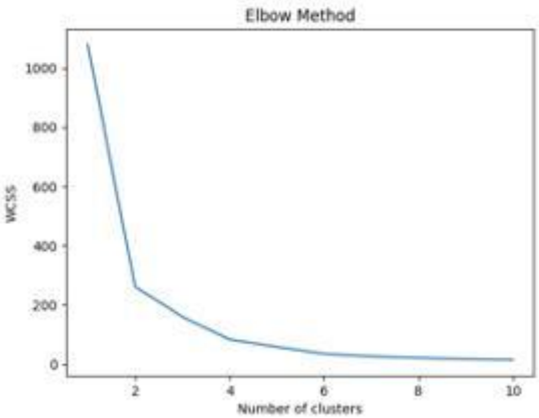


Figure 9: Representation of elbow metrics



References

[1] https://www.kaggle.com/code/vjchoudhary7/kmeans-clustering-in-customer-segmentation

[2] https://www.javatpoint.com/hierarchical-clustering-in-machine-learning

[3] https://realpython.com/k-means-clustering-python/

[4] https://scikitlearn.org/stable/auto_examples/svm/plot_svm_regression.html