MAKERERE UNIVERSITY

COLLEGE OF COMPUTING AND INFORMATION SCIENCES

MSc in Computer Science

MCS7103: Machine Learning

AI-Powered Price Intelligence for Car Buyers in Uganda Report

By: Kigula Jesse James

NAME	REGISTRATION NO.	STUDENT NO.
KIGULA JESSE JAMES	2025/HD05/26347U	2500726347

Table of Contents

1 Introduction

Buying a car online can be confusing because different sellers offer different prices for similar vehicles. Websites such as Jiji, Autorec and BeForward contain many car listings, but there is no easy way for buyers to know if a price is fair or too high. This project uses Machine Learning (ML) to help solve this problem.

The main goal of this work is to build:

- a model that predicts a fair price for a car,
- a model that tells whether a price is underpriced, fair or overpriced,
- and a system that recommends a good price range and shows similar cars.

These tools can help buyers make better decisions and understand the market more clearly.

2 Methodology

The project followed three main steps: collecting data, preparing the data, and building machine learning models.

2.1 Data Collection

Vehicle data was collected by scraping three popular websites: Jiji.ug, Autorec.co.jp, and BeForward.jp. Python scripts were used to extract information such as make, model, year, mileage, engine size, fuel type, transmission and price. Prices from BeForward were listed in USD, so they were converted to Uganda Shillings (UGX) using a fixed rate of 3600 UGX per USD.

2.2 Data Cleaning

Because the data came from different websites, it had many differences and missing values. The following cleaning steps were applied:

- Standardizing column names across all datasets,
- Converting mileage, engine size and year to numbers,
- Replacing missing numeric values with the median,
- Filling missing text values with simple placeholders,
- Matching Autorec and Jiji data to the richer BeForward format,
- Oversampling the Autorec dataset to balance the data.

After cleaning, all datasets were combined into one file called `cars_merged.csv`.

2.3 Price Prediction Model

A regression model was built to estimate the fair market price of a car. Several models were tested, including:

- Linear Regression,
- Decision Tree,
- Random Forest,
- Gradient Boosting.

The data was split into training and testing sets. Categorical features (such as make and model) were converted using one-hot encoding, while numeric features were scaled.

The best model was saved as `price_prediction_model.pkl`.

2.4 Overpricing Classification Model

To check if a car is underpriced, fair or overpriced, a second model was trained. First, the median price was calculated for every make–model–year group. Using this median, three labels were created:

- Underpriced: price is more than 10% below the median,
- Fair: price is within 10% of the median,
- Overpriced: price is more than 10% above the median.

Several classifiers were tested, and Random Forest gave the best results. The final model was saved as `overpricing_model.pkl`.

2.5 Recommendation Engine

A simple recommendation system was built to use both models. It gives:

- a predicted fair price,
- a suggested price range (low to high),
- a label showing if the given price is fair or not,
- and a list of similar cars for comparison.

This system can be used in a website or mobile app.

3 Results

3.1 Price Prediction Results

The regression models were compared, and Random Forest and Gradient Boosting performed the best. They produced reasonable error levels and showed that car price depends strongly on factors such as year, mileage, engine size and model. However, when I removed the Oversampled data for autorec, Gradient Boosting performed far better. The model generated from it was used for recommendation.

3.2 Overpricing Classification Results

The classification model was able to correctly label cars as underpriced, fair or overpriced. Most mistakes happened between "fair" and "overpriced," which is expected because real prices can vary.

3.3 Recommendation Engine Output

The recommendation engine successfully provided:

- fair price estimates,
- suggested price ranges,
- price labels,
- and similar car listings.

These results show that the system is useful for buyers who want quick and reliable price guidance.

4 Conclusion

This project showed that machine learning can help understand car prices in the Ugandan online market. By collecting and cleaning data from three websites and training prediction and classification models, it is possible to give buyers clear information about fair prices and overpricing.

In the future, the system can be extended to:

- run on a web server,
- support more car websites,

Overall, this approach can greatly support car buyers and sellers in making informed decisions.