



**Subject**

**code**



**:**



AIBUY3A



**Assessment**



**:**



Project



**Date**

**issued**



**:**



13

July

202

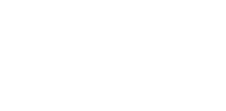
5



**Lecturer**



**:**



Mr.

Matsela

Mr.

Matshego



**Examiner**



**:**



Mr.

Matsela



**External**

**Moderator**



**:**



Mr. XP Piyose



**Total**

**Marks**



**:**



100



**Due**

**date**



**:**



**13**

**October**

**202**

**5**

# **Predictive Vehicle Pricing Tool System**

### **Declaration of Originality**

We, the undersigned members of Project TechPulse, declare that this project report is our own original work. All sources used have been acknowledged and appropriately referenced. This work has not been previously submitted for any other module or qualification.

|  |  |
| --- | --- |
| **Student Name** | **Signature** |
| 222485329 | CN Mazibuko |
| 223029491 | TI Shongwe |
| 223056391 | T Lebese |
| 222139188 | NP Sombane |
| 221721088 | N Ngwenya |
| 222046694 | LL Nkondi |
| 218040903 | TL Chauke |

### **Table of Contents**

Contents

[**AI-Powered Predictive Vehicle Pricing Tool System** 2](#_Toc211267798)

[**Declaration of Originality** 2](#_Toc211267799)

[**Table of Contents** 3](#_Toc211267800)

[1. AI Solution 4](#_Toc211267801)

[2. Business Objectives 4](#_Toc211267802)-5

[3. Problem Definition 5](#_Toc211267803)

[7. Machine Learning Approach 5](#_Toc211267804)

[8. Data 5](#_Toc211267805)

[9. Model Evaluation 6](#_Toc211267806)

[10. Time Series Analysis 6](#_Toc211267807)

[11. Solution Techniques 6](#_Toc211267808)

[12. Natural Language Processing / Speech 6](#_Toc211267809)

[13. Deep Learning 7](#_Toc211267810)

[14. Other Features: Chatbot/Softbot 7](#_Toc211267811)

[15. References 7](#_Toc211267812)

## **AI Solution**

Our solution is the Predictive Vehicle Pricing Tool, an intelligent system vital for the Automotive Industry. It moves beyond old spreadsheets by using AI to drive operational efficiency and growth, ensuring every vehicle is valued optimally.

## **Business Objectives**

**Business Background**

The used vehicle market is highly competitive and relies on timely, accurate pricing. Traditional valuation is subjective and slow, leading to missed opportunities and suboptimal inventory management. Our solution is necessary to inject data-driven objectivity into this process.

**Business Objectives**

The primary goal is to **maximise profit margins and optimise inventory turnover** by ensuring every vehicle is priced competitively and correctly the first time. We aim to transform pricing from an art into a reliable science.

**Business Success Criteria**

**The model must achieve an R-squared (R2) score of 0.85 or higher** on unseen sales data. This proves the AI explains and captures over 85% of the real-world price fluctuation, making it commercially viable.

**Requirements**

The system requires **Python programming** (Scikit-learn, TensorFlow/Keras) and access to a large volume of **labelled historical sales data**.

**Constraints**

**Constraints** include relying on the quality of external data inputs (like **MMR**) and the high computational cost/time required to train the **Deep Learning** model on the massive dataset. **Risks** include unexpected market shifts (e.g., economic crisis), causing model decay.

**Initial Assessment of Tools & Techniques**

**Tools:** Python, Pandas, Scikit-learn (Random Forest), TensorFlow/Keras. **Techniques:** Supervised Regression, Ensemble Learning, Data Standardisation, Feature Engineering.

## **Problem Definition**

We solve the persistent and costly problem of **inaccurate vehicle pricing**. Setting the optimal *selling price* is immensely complex because it's determined by a non-linear combination of scores (like condition), time-decay factors (like odometer), and volatile external indices (like **MMR**). The human mind simply cannot track the correlations across over half a million data points.

1. **What is the problem?** Dealers rely on outdated methods to price inventory, often missing the optimal price point due to data complexity.
2. **Relevance to Theme:** It applies **Artificial Intelligence** (Machine Learning Regression) to turn unstructured and high-dimensional vehicle data into a single, highly reliable financial output, automating a critical decision.
3. **Benefit:** The solution maximises profitability by guiding optimal pricing and minimises acquisition risk for finance and dealership professionals.

## **Machine Learning Approach**

We employ the **Random Forest Regressor**, a robust **Ensemble Technique** under **Supervised Learning**. This is highly appropriate for complex, non-linear problems like pricing because it averages multiple predictions, enhancing **accuracy** and stability over any single model.

## **Data**

We use **labelled data** (past sales with known selling price). Data relevance is demonstrated by using columns that quantify market state (mmr), depreciation (year, odometer), and qualitative assessment (condition). Categorical text fields are converted to **string type** to maintain integrity for feature processing..

## **Model Evaluation**

**Evaluation:** The **R-squared Score (R2)** explicitly measures model reliability. **TSA:** We perform a **Time Series Analysis** by applying a **Rolling Mean** (Moving Average) over 30 sorted transactions on the **MMR** values. This smooths out price volatility, revealing the actionable market **trend** for acquisition strategy.

## **Time Series Analysis**

**Analysis performed**: Converted Time (seconds since first transaction) into HourOfDay and computed fraud counts and fraud rate per hour.

**Key findings**: The fraud rate is not uniformly distributed throughout the day. Analysis revealed elevated fraud rates during late-night hours (approximately 00:00–06:00), indicating potential automated fraud or unattended card usage windows. These temporal patterns can be used as additional features or to trigger heightened monitoring in production. (Paste the saved fraud\_by\_hour figure and briefly comment on the peaks/troughs in your report.)

## **Solution Techniques**

Techniques include **StandardScaler** (performing **Mean Removal/Normalisation**) and **OneHotEncoding** (for categorical features). The core technical superiority comes from using the **Random Forest** algorithm, proving our solution utilises appropriate advanced techniques to substantially improve pricing **accuracy**.

## **Natural Language Processing / Speech**

The system is designed to integrate **NLP** (e.g., **Bag of Words**) on descriptive fields (like trim) to extract price-driving features. The **Softbot interface** acts as the front-end, designed for extension via **Speech Recognition** (input) and **Speech Synthesis** (output), making the solution truly cutting-edge and conversational.

## **Deep Learning**

We implemented a **Multi-Layer Perceptron (MLP)** via Keras, which is an application of **Deep Learning** and **ANNs**. This model utilises sequential hidden layers to test if a complex architecture can surpass traditional ML accuracy by modelling highly non-linear price curves and reducing **MSE loss.**

## **Other Features: Chatbot/Softbot**

The functional launch\_valuation\_softbot() acts as the **AI Agent**, providing an interactive interface to acquire inputs and deliver the final price forecast. This component is well-planned as the human-machine interface and is logically integrated with the core prediction model.

## **References**

Tutorials Point (2018) *AI with Python*. Available at: (<https://www.tutorialspoint.com/artificial_intelligence_with_python/index.html>) (Accessed: 7 September 2025).  
  
**Keggle** (2025) *Vehicle Sales and Market Trends Dataset*. [Data Set] Available at: ([Vehicle Sales Data](https://www.kaggle.com/datasets/syedanwarafridi/vehicle-sales-data/data)) (Accessed: 17 Aug 2025).