Harnessing Explainable AI and Machine Learning for Dual Predictive Modeling of Car Goodwill and Crop MSP in Price and Policy Forecasting

Meghna Chaudhary[1], Mohammad Afshar Alam[1], Sherin Zafar[1,*], and Kiliyanal Muhammedkunju Abubeker[2]

[1]Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi, India
[2]Department of Electronics and Communication Engineering, Amal Jyothi College of Engineering (Autonomous), Kanjirappally, Kerala, India
Email: meghnachaudhary828@gmail.com (M.C.); aalam@jamiahamdard.ac.in (M.A.A.); sherin.zafar@jamiahamdard.ac.in (S.Z.); kmabubeker82@gmail.com (K.M. A.)
*Corresponding author
Manuscript received February 12, 2025; revised April 15, 2025; accepted November 25, 2025; published March 17, 2026

***Abstract*—Accurate yet interpretable forecasting remains a major challenge in data-driven price and policy decision-making, as many machine learning models prioritize predictive performance while offering limited transparency. To address this gap, this study proposes an integrated Explainable Artificial Intelligence (XAI)—enabled machine learning framework for dual predictive modelling across two economically significant domains: automotive valuation and agricultural price policy. The framework simultaneously predicts car goodwill values and crop Minimum Support Prices (MSP), enabling cross-domain analysis while maintaining model interpretability. Comparative experiments are conducted using multiple machine learning techniques, including linear and ensemble-based models, applied to automotive and agricultural datasets. Ensemble methods demonstrate superior predictive capability in both domains. To enhance transparency and stakeholder trust, XAI techniques are incorporated to explain model behaviour and identify key influencing factors. The analysis shows that depreciation and brand-related attributes play a dominant role in car goodwill valuation, whereas climatic and cost-related factors significantly influence MSP predictions. The results confirm that integrating XAI with machine learning improves both predictive reliability and interpretability, transforming black-box models into actionable decision-support systems. The proposed dual predictive framework offers a scalable and transparent approach for market optimization and policy evaluation, highlighting the practical value of explainable AI in strategic economic planning and data-driven governance.**

***Keywords*—Explainable Artificial Intelligence (XAI), goodwill car values, crop Minimum Support Price (MSP), linear regression, random forest, decision tree, gradient boosting, XGBoost and Marketing Mix Modelling (MMM)**

I. INTRODUCTION

In the data-driven society of currently, the capacity to extract useful insights from massive amounts of data has become crucial for businesses across various sectors. Two areas where data analytics play a significant role are the automotive and agricultural industries. In the automotive sector, predicting Goodwill Car Values is essential for strategic pricing, inventory management, and overall market competitiveness [1]. Similarly, in the agricultural sector, forecasting Minimum Support Prices (MSP) for crops is vital for ensuring fair prices for farmers, stabilizing markets, and informing policy decisions. Machine Learning (ML) employs data for developing models capable of analyzing information and making autonomous decisions without human intervention. It elucidates how computers leverage past experiences to operate independently [2]. Predicting car prices is a fascinating and widely studied problem. Achieving accurate car price predictions necessitates expert knowledge, as prices typically depend on numerous unique features and factors [3]. Key determinants include the brand, vehicle model name, kilometers driven etc. Additionally, the type of fuel used and the car's fuel consumption per mile significantly impact its price [4], especially given frequent fluctuations in fuel prices. Accurate car value prediction has significantly simplified the car-buying process for consumers, requiring minimal effort and expertise. With numerous brands regularly introducing new models at premium prices, many customers find it financially challenging to purchase new cars. Consequently, reliability used car value forecast that precisely assesses an automobile's worth is needed worldwide, facilitating informed purchasing decisions. To achieve this, we applied machine learning techniques, specifically regression algorithms, which provide continuous value outputs. Algorithms such as Random Forest and Linear Regression were employed to enhance accuracy. Utilizing data from an online repository, we processed and compared the performance of various algorithms to determine the most effective model for predicting used car prices. Given the challenging economic conditions, it is anticipated that the sales of second-hand imported (reconditioned) cars and used cars will increase. The sales of new cars have notably decreased in 2020 and 2021 due to pandemic conditions. Predicting the resale value of a car is a complex task, as the value of used cars depends on numerous factors. Unfortunately, comprehensive information about all these factors is not always available, requiring buyers to make purchasing decisions based on a limited set of variables. This research focuses on developing machine learning models to accurately predict the price of used cars based on their features, thereby facilitating informed purchases. This research implements and evaluates various machine learning algorithms, including Linear Regression, Random Forest, Decision Tree, Gradient Boosting and XGBoost, comparing their performance to select the most effective model.

Significant advancements in machine learning offer limitless possibilities. Machine learning has evolved alongside vast data advancements, creating opportunities to determine, evaluate, as well as understand complex processes. Numerous scholars and professionals in contemporary agriculture are putting their theories to the test on a larger scale, leading to more accurate and consistent predictions. Contemporary farming can discover numerous methods for storing water, using nutrients as well as energy more efficiently, and respond to various changes in the environment. Few learning-based usages in farming include crop yield forecasting, crop disease identification, weed detection, plant variety recognition. Assumption of Yield, a critical aspect of precision agriculture, holds great importance for yield measurement, evaluation, harvest supply coordination, and crop management to enhance productivity [5]. Another significant responsibility in agriculture is controlling pests and crop diseases in outdoor and nursery environments [6]. The most commonly used pest control practice involves regularly spraying pesticides over farming areas, which leads to substantial financial and environmental costs. ML is needed for precision agriculture, optimizing agrochemical usage in terms of timing and location [7]. Weeds pose a major threat to crop production, and their precise detection is crucial for agricultural sustainability, as weeds are challenging to distinguish from crops [8]. Machine learning algorithms, combined with sensors, enable accurate weed detection and segregation with minimal effort and no environmental impact [9]. Additionally, applications have been developed to recognize features related to crop quality. Accurate harvest characteristics of quality can be identified and separated, which can raise product pricing and reduce losses [10]. This research implements various machine learning models for forecasting crop MSP. The introduction of machine learning algorithms has transformed the agriculture industry by offering formerly unattainable accuracy and insights for MSP forecasting for crops, which is vital for ensuring fair prices for farmers and stabilizing markets.

The predictive accuracy of machine learning models is essential for supporting informed strategic decisions; however, limited interpretability often leads to reduced user confidence. Explainable Artificial Intelligence addresses this limitation by clarifying how predictions are generated, thereby improving transparency and trust. The proposed dual prediction approach enables a comprehensive understanding of the factors influencing both car values and crop prices.

The primary objective of this study is to develop and evaluate machine learning models for predicting Goodwill Car Values and Crop Minimum Support Prices while integrating XAI techniques to enhance model interpretability and adaptability. The study further aims to assess model effectiveness within the context of Marketing Mix Modelling and to generate actionable insights for strategic decision-making in the automotive and agricultural sectors. The integration of XAI with machine learning in dual prediction tasks demonstrates the potential of advanced analytics to optimize marketing strategies and resource allocation. The findings of this research contribute to improved predictive accuracy and interpretability, strengthening strategic planning and stakeholder trust across both domains. By addressing challenges related to accuracy and transparency, the study supports more effective and accountable decision-making processes, ultimately contributing to sustainable development.

Calculating the Goodwill Car Value factor plays a crucial role in determining the accurate market value of used vehicles by accounting for brand reputation, customer perception, and historical performance. It supports informed decision-making for buyers and sellers, ensures transparency in financial reporting for businesses, provides valuable insights into market trends and consumer preferences, and assists in identifying and managing potential risks.

There are certain key technological challenges and constraints that are associated with predicting Goodwill Car Values and Crop Minimum Support Prices (MSP) and that affect model performance and reliability. One of the primary challenges lies in the accessibility and reliability of data, as accurate predictions depend on the availability of comprehensive, consistent, and high-quality datasets. Incomplete, noisy, or biased data can significantly degrade predictive accuracy and limit model robustness.

Another critical challenge involves feature design and selection. Identifying relevant and informative features is essential for effective prediction, while redundant or irrelevant features may increase model complexity and reduce accuracy. In addition, the development and deployment of machine learning models often require substantial computational resources, making efficient resource utilization an important consideration, particularly for large-scale or real-time applications.

Ensuring that predictive models generalize well to unseen data remains a persistent challenge. Overfitting can lead to strong performance on training data but poor reliability in practical deployment. Furthermore, the interpretability of complex machine learning models presents a significant limitation, as opaque decision-making processes can hinder user understanding and acceptance. Enhancing model transparency and explainability is therefore crucial for building stakeholder trust and supporting informed decision-making.

By addressing these technological challenges, the options can be provided for further machine learning research and development, ultimately boosting prediction accuracy and dependability in both automotive and agricultural sectors.

The format of this research is as follows: Section II examines relevant research in the area of predicting used car prices and crop MSP evaluation. Section III outlines the method and materials including marketing mix modelling and explainable artificial intelligence that are used in performing the research along with various machine learning algorithms used to evaluate their performance in predicting used car prices and crop MSP values. Section IV outlines the conclusion of the research carried out. Section V discusses the study's future directions.

II. Literature Review

This section offers a thorough analysis of existing literature on the implementation of Explainable AI and ML approaches to predict Goodwill Car Values and Crop Minimum Support Prices (MSP) within the framework of Marketing Mix Modelling (MMM). The review highlights key studies, methodologies, and findings that have contributed to the advancement of dual prediction models in these domains. By synthesizing the current state of study, this section seeks to fill in the vacancies in the available information and provide the foundation for the proposed research.

In Ref. [11], Gegic *et al.* initially examined a variety of distinct attributes to ensure accurate and precise estimate. Making use of three machine learning techniques—Random Forest, Support Vector Machine, and Artificial Neural Network—they created a model to forecast used automobile prices. To determine which algorithm best suited the provided dataset, the authors evaluated the performance of several different algorithms. A Java application was created using the finished prediction model. The accuracy of the model, when tested using test data, was 87.38%. In Ref. [12], Hudon *et al.* use techniques like Lasso, Logistic Regression analysis, and regression tree models to build a statistical framework to estimate a used car's pricing based on past client data and a set of attributes. They also analyze forecast accuracy of different models to identify the most accurate algorithm for calculating the car's price. Pudaruth [13] looked into using supervised machine learning methods to forecast used car prices in Mauritius. Historical information gathered from daily newspapers served as the basis for the forecasts. The predictions are made using a variety of methods, including decision trees, k-Nearest-Neighbor estimation, Naive Bayes, and multiple linear regression analysis. The methods providing the best performance are then determined by evaluating and comparing these forecasts, what appeared to be a straightforward problem proved to be quite difficult to address accurately. The performance of the four employed approaches was equivalent.

Samruddhi and Kumar [14] suggested a model based on supervised machine learning to analyze used automobile prices using the method known as the KNN (known as K-N Neighbor) prediction algorithm. Data from the site Kaggle was used to train the model. The data was analyzed using different training parameters and test ratios throughout the experiment. Consequently, the suggested model was determined to be the optimal model after achieving an accuracy of almost 85%. To forecast used automobile prices, Pal *et al.* [15] used Random Forest, a supervised learning technique. Following a comprehensive exploratory data study to ascertain how each feature affected price, the model was chosen. To train the data, a Random Forest with 500 different decision trees was constructed. Training accuracy was 95.82% and testing accuracy was 83.63%, according to the experimental data. By concentrating on the most linked features, the model predicts automobile prices with accuracy. Chong *et al.* [16] offered a thorough analysis of technological applications for palm oil production, including subjects such yield prediction, Biomass Above Ground (AGB) and production of carbon calculation, tree calculating, recognizing changes, age estimations, and more. They pointed up prospective areas for further study and offered potential fixes. However, methods for predicting palm oil yield were not the main focus of their review. Young [17] addressed important techniques used recently in surveys, remote sensing, official statistics generation, and combining them with administrative, meteorological, and other data. In addition to highlighting the uncertainties associated with these projections, their study focused on ways to improve current crop production prediction techniques. The paper examined crop production prediction methods over wide geographic regions; however, it did not address machine learning techniques that are widely employed to forecast crop yield. Additionally, this article might not be helpful to users seeking accurate crop production forecast models for particular crops.

A comprehensive review of a variety of characteristics and prediction algorithms was conducted by Klompenburg *et al.* [18]. But rather than critical analysis, identifying research gaps, and making recommendations, the study concentrated more on information extraction. Chlingaryan *et al.* [19] reviewed the use of machine learning for nitrogen status estimation. Their research came to the conclusion that quick developments in machine learning and sensing technologies might result in cost-effective agricultural solutions. Elavarasan *et al.* [20] evaluated machine learning frameworks that are pertinent to predicting crop yield, with a primary focus on climate characteristics. They recommended looking for more thorough standards for crop output. To determine the mechanisms causing voids in palm oil production, another review [21] examined the body of research regarding palm oil output from a physiological standpoint. Liakos *et al.* [22] examined how machine learning is being used in agriculture by looking at articles on land, crop, livestock, and water management. Li *et al.* [23] reviewed methods for determining fruit maturity in order to improve harvesting time and yield forecasting.

Above preceding discussion highlights the need for a thorough review paper that fills in the gaps in the body of current review literature. This study contributes to the existing literature by addressing key research gaps in interpretable dual-domain prediction through a unified machine learning and Explainable Artificial Intelligence framework. The research highlights the importance of systematically modelling the Goodwill Car Value factor and examines the benefits and limitations of various feature sets and learning algorithms for reliable automotive valuation. It further identifies current and emerging technological challenges in dual prediction, particularly those related to data quality, model generalization, and interpretability, thereby extending the scope for future research in explainable machine learning applications. In the agricultural domain, the study clarifies the fundamental aspects of crop yield and MSP prediction processes and provides a critical assessment of contemporary machine learning approaches used for this purpose. By synthesizing insights from both domains, the work advances understanding of how transparent predictive models can support more accountable decision-making in marketing and policy contexts. Overall, the paper offers an integrated perspective that bridges methodological gaps between accuracy and explainability, establishing a foundation for scalable and trustworthy AI-driven forecasting in economic and agricultural systems.

III. Materials and Methods

The study employs a robust dataset comprising economic indicators, market trends, and sector-specific variables. For the automotive sector, the focus is on predicting Goodwill Car Values, utilizing features such as car make, model, age, mileage, and market demand. In the agricultural sector, the model aims to forecast Crop MSP, incorporating variables like historical crop yields, weather conditions, cost of production, and market prices. Various ML algorithms, including Linear Regression, Decision Tree, Random Forest, Gradient Boosting, and XGBoost, are evaluated for their predictive capabilities. XAI techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are applied to ensure the models' decisions are comprehensible and transparent. To employ Explainable Artificial Intelligence (XAI) and machine learning techniques for dual prediction of Goodwill Car Values and Crop MSP (Minimum Support Prices), the following methods and materials are typically used.

A. Methods

1) Data collection and preprocessing

The methodology begins with the collection of historical data on both car values and crop yields in order to establish a reliable foundation for predictive modelling. Relevant features are then integrated from the gathered information, including automotive attributes such as make, model, year, and mileage, as well as agricultural factors such as weather conditions and soil characteristics. Following this, a thorough data cleaning process is conducted to remove inconsistencies, missing values, and potential discrepancies, ensuring that the final dataset is accurate and suitable for model development.

2) Model selection and training

Appropriate machine learning algorithms are selected to suit the predictive objectives of the study, including techniques such as Linear Regression and XGBoost. The pre-processed data are then used to train and refine these models so that they can effectively learn underlying patterns and relationships within the datasets.

3) Explainable AI techniques

Explainability is achieved through the use of SHapley Additive exPlanations, which illustrate how each individual feature contributes to the model's predictions and overall forecasting behavior. In addition, Local Interpretable Model-agnostic Explanations are applied to provide case-specific justifications for particular predictions, enabling a clearer understanding of model decisions at the local level.

4) Model evaluation

The developed models are evaluated using appropriate performance metrics, such as the R^2 score and other relevant indicators, to measure their predictive accuracy and reliability. In addition, cross-validation techniques are employed to ensure that the models generalize effectively to unseen data and maintain consistent performance across different data subsets.

5) Estimating and evaluation

The trained models are utilized to perform dual prediction by estimating both Goodwill Car Values and Crop Minimum Support Prices. Furthermore, sensitivity analysis is conducted to examine how variations in different input variables influence the resulting predictions.

B. Materials

The study relies on high-performance computational resources to support the training and evaluation of machine learning models. A range of software tools is employed, including machine learning frameworks such as scikit-learn, TensorFlow, and PyTorch, along with explainable AI libraries like SHAP and LIME. The analysis utilizes datasets containing historical information on car values and crop yields to develop and validate the predictive models. In addition, visualization tools such as Matplotlib and Seaborn are used to present model outputs and interpretability results in a clear and meaningful manner.

By combining these methods and materials, researchers can effectively harness XAI and machine learning methods to increase reliability and analysis interpretability in Marketing Area.

C. Fundamental Aspects of the Dual Prediction Process

In the context of a predictive process, fundamental aspects are the critical steps and considerations necessary to build an accurate and reliable prediction model [24]. Essentially, they are the foundational building blocks upon which the entire process is constructed [25].

Fig. 1 outlines the key steps involved in using XAI and machine learning techniques for dual prediction of Goodwill Car Values and Crop MSP. It highlights the processes of data collection and preprocessing, model selection and training, incorporation of explainable AI techniques, prediction and analysis, performance evaluation and optimization, and finally, implementation and decision-making.

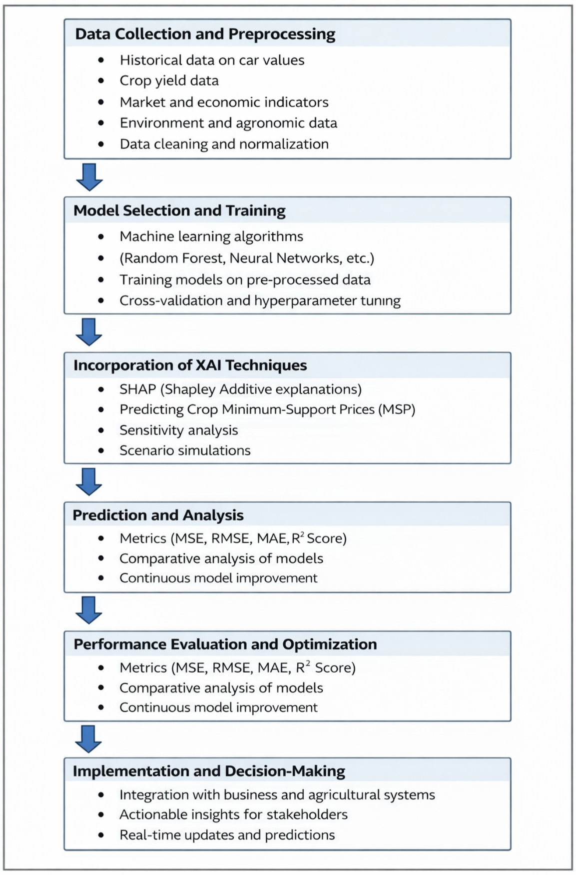


Fig. 1. Detailed XAI and ML multi-stage workflow.

D. Explainable AI

The term "Explainable Artificial Intelligence" (XAI) describes a collection of procedures and techniques that enable people to comprehend and evaluate the results of machine learning algorithms [26]. The goal of XAI is to close the disparity between the intricacy of artificial intelligence models and the requirement for accountability, transparency, and trust in judgments made by AI [27].

1) Fundamental aspects of XAI

Explainable Artificial Intelligence strengthens this study by ensuring transparency in the predictive models, enabling users to clearly understand how decisions regarding car goodwill values and crop MSP are generated. Such transparency improves confidence in data-driven recommendations and supports the practical adoption of the proposed framework. Interpretability further enhances this value by translating complex model behavior into understandable feature-level explanations, allowing stakeholders to recognize which economic and agricultural factors drive predictions. The framework also promotes accountability by making algorithmic decisions traceable and justifiable, which is essential when outputs influence financial planning and agricultural policy. In addition, explainability supports regulatory and ethical compliance by providing clear reasoning for AI-driven outcomes, aligning the system with emerging governance requirements. Together, these aspects demonstrate that XAI is not only a technical enhancement but a necessary component for deploying trustworthy and responsible dual prediction models in real-world decision environments.

Fig. 2 outlines the key operations undertaken during the XAI process: gathering of data and training of systems, model prediction, applying explainability layer and interpreting the results.

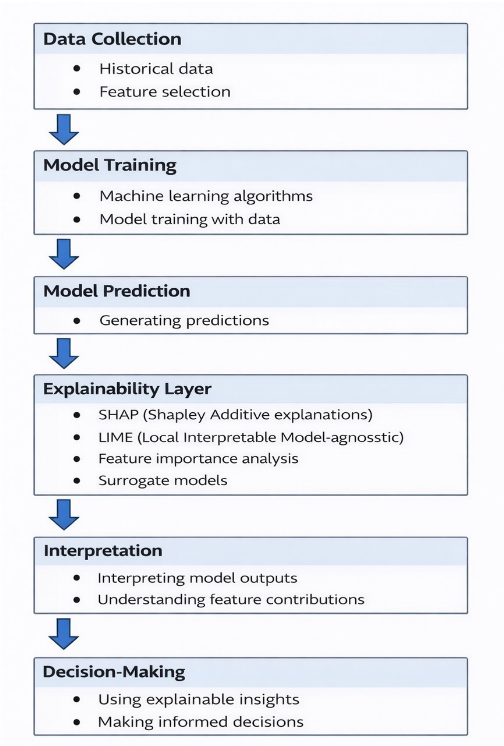


Fig. 2. A diagram illustrating the process of utilizing XAI and ML techniques during two scenarios.

2) Explainable AI important techniques

a) SHAP (SHapley Additive exPlanations) algorithm

The following formula is used to determine the SHAP value for individual attribute x_i provided a model for prediction f and a particular instance x with characteristics x_1, $x_2,\ldots,$ x_n. For each feature x_i, calculate its marginal contribution to the prediction $f(x)$ by contrasting the forecast alongside and without the characteristic x_i. The mean of the marginal contributions of feature x_i across all potential feature subsets is its Shapley value (Eq. (1)).

Mathematically, Shapley value is determined as follows:

$$\phi_i(f,x) = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}\left(f(x_s \cup \{x_i\}) - f(x_s)\right) \tag{1}$$

where:

- N: the collection of all features is N.
- S: a group of features that do not include x_i.
- $|S|$: the subset S's feature count.
- $f(x_S)$: the prediction of the model when only features in S are used.
- $f(x_S \cup \{x_i\})$: the model's prediction when S and x_i features are employed.

b) LIME (Local Interpretable Model-agnostic Explanations) algorithm

The LIME algorithm provides local justifications for each prediction produced by intricate machine learning models. Here's a simplified explanation of the mathematical algorithm behind LIME Algorithm.

- *Local perturbation:* Generate a new dataset by perturbing the instance x_x for which we want an explanation. This involves creating new samples around x by randomly perturbing its features within a defined neighbourhood.
- *Model training on perturbed data:* Use the involved sample to train an interpretable model. This simpler model is used to approximate the complex model's local behaviour surrounding the instance x.
- *Determining the significance of the attributes:* Determine each feature's significance in the comprehensible model. This is accomplished by calculating the relative contribution of each feature to the interpretable model's prediction.
- *Formation of descriptions:* Creating an explanation for the complex model's prediction for instance x using the feature importance values. The description identifies the characteristics that had the biggest impact on the outcome.

Let's denote the complex model as f and the instance we want to explain as x. The steps can be summarized as follows

- *Perturbation:* Generate a new dataset D by perturbing x within a neighbourhood ϵ.
- *Interpretable model training:* Train an interpretable model g on D to approximate f locally (Eq. (2)).

$$g(x) \approx f(x) \tag{2}$$

- *Feature importance:* For each attribute x_i in the comprehensible model g, determine the feature importance β_i.
- *Explanation:* The explanation for the prediction $f(x)$ is given by the feature importance values β_i.

E. Mathematical Representation: Applied Machine Learning Algorithms

1) Linear regression

Linear regression is a machine learning algorithm used to model the relationship between a dependent variable (Y) and one or more independent variables (X). It helps in predicting outcomes based on input data.

Mathematical formula for Simple Linear Regression is as follows (Eq. (3)):

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (3)$$

where:
- Y: Dependent variable (the value we want to predict)
- X: Independent variable (the predictor)
- β_0: Intercept (value of Y when $X = 0$)
- β_1: Slope coefficient (change in Y for a unit change in X)
- ε: Error term (accounts for variability not explained by X)

This helps in modelling more complex relationships between input features and the target variable.

2) Decision tree

A Decision Tree is a supervised machine learning algorithm that helps in classification and regression tasks. It works by repeatedly splitting the data based on certain conditions until we reach a decision.

The mathematical formulas used in Decision Trees is Entropy, which measures uncertainty in data. It (Eq. (4)) determines how mixed the data is. If the data is pure (all samples belong to one class), entropy is low; if mixed, entropy is high.

$$E(S) = -\sum_{i=1}^{c} p_i \log_2 p_i \quad (4)$$

where:
- $E(S)$: Entropy of the dataset S
- c: Number of classes
- p_i: Proportion of instances belonging to class i
- $\log_2$: Logarithm base 2 (used to measure information in bits)

Decision Trees are widely used for tasks like spam detection, medical diagnosis, and customer segmentation.

3) Random forest

Random Forest is a machine learning algorithm that builds multiple decision trees and combines their outputs for improved accuracy and robustness. It is widely used for classification and regression tasks. It operates by constructing multiple decision trees and making predictions based on majority voting (for classification) or averaging (for regression).

Mathematical formulation of Random Forest (regression) (Eq. (5)):

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x) \quad (5)$$

Mathematical formulation (Eq. (6)) of Random Forest (classification):

$$\hat{y} = mode\left(h_1(x), h_2(x), \ldots, h_T(x)\right) \quad (6)$$

where:
- $\hat{y}$: Final predicted output (either a value or class)
- T: Total number of decision trees in the forest
- $h_t(x)$: The prediction from the t-th decision tree
- x: Input feature vector

The key advantage of Random Forest is that it handles large datasets well and reduces overfitting compared to individual Decision Trees.

4) Gradient boosting

Gradient Boosting is a machine learning technique that builds multiple weak learners (usually Decision Trees) sequentially, each one improving upon the previous model by correcting errors. It minimizes the loss function using gradient descent.

Mathematical formulation of Gradient Boosting (Eq. (7)):

$$F_M(x) = F_0(x) + \sum_{m=1}^{M} v \cdot \gamma_m \cdot h_m(x) \quad (7)$$

where:
- $F_M(x)$: Final prediction after M boosting rounds
- $F_0(x)$: Initial model (e.g., the mean value for regression)
- $h_m(x)$: The prediction of the m-th weak learner
- γ_m: Weight (scaling factor) assigned to the m-th learner based on its fit to the residuals
- v: Learning rate (typically between 0.01 and 0.3), controls the impact of each learner
- $\sum(m = 1 \text{ to } M)$: Total number of boosting iterations or weak learners

Gradient Boosting optimizes a loss function by combining multiple weak models ($f\_m(x)$) into a strong learner. The key advantage of Gradient Boosting is that it handles complex patterns well and works for classification and regression tasks.

5) XGBoost

Extreme Gradient Boosting (XGBoost) is a powerful and optimized gradient boosting algorithm used for classification and regression tasks. It improves upon standard Gradient Boosting by incorporating regularization, efficient computation, and better handling of missing values.

Mathematical formulation of XGBoost (Eq. (8)):

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \; f_k \in \mathcal{F} \quad (8)$$

where:
- $\hat{y}_i$: predicted value for the i-th sample
- K: total number of trees
- f_k: individual regression tree
- x_i: input features for the i-th sample
- F: space of regression trees (functions $f(x) = wq(x)$, where q is a function mapping features to leaf index and wq are leaf weights).

The key advantage of XGBoost is that it handles missing values efficiently and provide fast training with parallel processing.

F. Data Collection

This study utilizes data (dataset 1) one from publicly available datasets and another data (dataset 2) collected from the portal available at government website. However, the dataset 1 is later modified by the author and additional fields are added in the obtained data. The fields (attributes) added have a great significance in this study. The datasets for Goodwill Car Values and Crop Minimum Support Prices (MSP) are crucial for both the automotive and agricultural sectors, respectively. They provide valuable insights for predictive modelling, helping to accurately forecast car resale values and crop prices, thus informing better decision-making. For car dealers and buyers, understanding market trends and key features impacting car value can lead to more informed transactions. Meanwhile, crop MSP datasets support government policy-making, ensuring fair pricing for farmers and fostering economic stability. Together, these datasets drive transparency, accountability, and strategic planning in their respective domains, ultimately benefiting businesses and consumers alike. The data sources and significance are as follows

1) Goodwill car values dataset

The Online Repository is used for obtaining the first dataset utilized in this work. The dataset belongs to the Cars24 company. With its headquarters located in Gurgaon, Haryana, Cars24 is a multinational Indian internet marketplace for used cars. This organization aims to revolutionize used car market by providing a seamless platform for buying and selling pre-owned vehicles. Cars24 offers a range of services, including car sales, resale paperwork, loans, and vehicle inspections. The dataset consists of 8010 rows and 17 columns. It covers various cars resale factors like Car Name, Year, Distance, Owner, Purchase Price, Engine in cc, Power in bhp, Mileage (Km/l or Km/kg), Fuel, Location, Drive, Type, Selling Price, Depreciation Rate, Annual Depreciation Multiple Owners and Adjusted Depreciation Rate/Goodwill. The primary objective of the Goodwill Car Values data is to enable accurate prediction of used car prices by analyzing several features. This dataset is instrumental in developing machine learning models that help businesses and consumers make informed decisions about buying and selling used cars, ensuring fair pricing and transparency in the market.

The Goodwill Car Values dataset is significant for several reasons:

- *Predictive modelling:* It is widely used in projects to determine used autos' resale value. By analyzing features for example brand reputation, features, and more, models can provide accurate price predictions.
- *Market insights:* The dataset helps businesses and consumers understand market trends and the factors influencing car prices. This can guide purchasing and selling decisions, as well as inform marketing strategies.
- *Feature importance:* It allows for the identification of key features that contribute to a car's value, helping manufacturers and dealers focus on aspects that add the most value to their vehicles.

We acknowledge the domain-specific biases present in the car dataset. The data, sourced from a single platform in India, may reflect specific brand preferences, car models, and pricing structures, potentially limiting generalizability. To mitigate this, we conducted Exploratory Data Analysis (EDA) to identify and remove outliers that could skew predictions. However, the model's applicability beyond this dataset remains a consideration for future research.

Additionally, factors such as location, fuel type, and mileage may introduce inherent biases in pricing predictions. While our model accounts for these factors, external economic trends and regional pricing variations might influence car prices in ways not captured by the dataset. Future work could explore incorporating macroeconomic indicators or user behavior trends to further refine the model's predictive capabilities. These considerations strengthen the methodological rigor of our study and ensure that the reported results accurately reflect model performance in the given context

2) Crop MSP dataset

The Crop MSP dataset is collected from the portal available at the government website and using government crop related document (Swaminathan Report, available at the government site). The government crop MSP (Minimum Support Price) related dataset is an important asset in understanding agricultural economics and policy-making. The dataset typically provides insights regarding the MSP for various crops, production costs, and market prices. The data is often accessible through platforms like the Open Government Data (OGD) Platform India and the National Portal of India, providing transparency and aiding stakeholders in making informed decisions. There are 10 columns and 19689 rows in the dataset. It covers various agricultural production and marketing related factors for example Name of the crop, Year of the crop, Crop Area, Crop Production, Crop Yield etc. The Crop MSP (Minimum Support Price) data is crucial for ensuring fair pricing and economic stability in the agricultural sector. It aids government policy-making, enabling the establishment of price to protect farmers' income and livelihoods. The dataset also supports economic analysis, helping to understand the impact of MSP on production and market prices. By promoting transparency and accountability, it enables farmers to decide on crop cultivation with expertise. Additionally, researchers use this dataset to study trends, improve agricultural practices, and advance crop yield research, while businesses and consumers gain insights into market trends and influencing factors, facilitating better purchasing and marketing strategies.

We acknowledge the domain-specific biases present in the MSP dataset. The dataset is country-specific (India's MSP system) and is influenced by government policy, inflation, and economic conditions. These factors might introduce trends that affect MSP predictions over time. To mitigate potential biases, we conducted Exploratory Data Analysis (EDA) to identify and remove anomalies or outliers that could skew predictions. Additionally, we ensured that temporal trends were taken into account by maintaining the integrity of year-wise data splits, preventing data leakage between training and testing sets.

Given the nature of MSP determination, incorporating macroeconomic indicators, policy decisions, and additional agronomic factors could further enhance model performance. Future work could explore time-series forecasting techniques or policy-based modelling approaches to better capture these dynamics. These considerations strengthen the methodological rigor of our study and ensure that the reported results accurately reflect model performance in the given context [28].

Both the dataset (dataset 1 and dataset 2) collected involves sensitive marketing data. Ethical guidelines are adhered to for data protection regulations to maintain privacy, confidentiality, ensuring that data usage does not lead to discriminatory practices and adhere to all relevant data protection regulations.

G. The Reasons Behind the Selection of the Datasets

The Cars24 dataset was selected because it represents a widely used platform for pre-owned vehicle transactions and contains relevant indicators such as customer demand, depreciation behavior, and market trends that support the estimation of goodwill car values. Similarly, the government crop dataset provides a reliable foundation for forecasting

Minimum Support Prices, as it includes historical MSP patterns, crop yield information, and regional variations that reflect real agricultural conditions [29]. To address potential biases within these datasets, a structured preprocessing approach was adopted in which anomalies, missing values, and inconsistencies were identified and cleaned to prevent distortion of model predictions. In addition, fairness-oriented techniques were considered to reduce pricing and policy prediction bias through methods such as debiasing strategies and fairness-aware learning mechanisms. Explainable Artificial Intelligence further supported openness by enabling the detection of biased model behavior through SHAP and LIME analyses, ensuring that the resulting predictions remained interpretable and equitable [30].

IV. Result and Discussion

For the Goodwill Car Values prediction, the dataset (dataset 1) belongs to the Cars24 company containing sale of the 8010 cars. Cars24 is an online platform based in India that specializes in buying and selling used cars. It offers a wide range of certified pre-owned vehicles that have been thoroughly inspected on various parameters to ensure quality. Cars24 provides services such as car valuation, financing options, vehicle history reports, and even assistance with tasks like RC transfer and insurance. The goal is to determine impact of various marketing parameters or other key performance indicators on the sale of the cars. The dataset 2 is obtained from the government website and comprises of 19,689 crops. The goal is to evaluate how variables like Fertilizer, Pesticide, Rainfall, Area, and State impact the MSP. The machine learning is applied to the Marketing Mix Model over dataset 2 to understand how different inputs—be it price, place, or promotional efforts—affect sales outcomes.

A. Results for Dataset 1

The below Table 1 presents the results obtained with respect to the Dataset 1. The train/test split ration related to this dataset is 80:20. The k-fold cross-validation was employed and the number of folds (*k*) utilized is 10.

Table 1. Descriptive statistics of numerical features in the dataset

Features	Mean	Std Dev	Min	25%	Median	75%	Max
Year	2016.99	2.86	2010	2015	2017	2019	2023
Distance (km)	52,636	29,183	0	30,751	50,382	71,764	971,212
Owner	1.30	0.51	1	1	1	2	4
Purchase Price (Lakh ₹)	8.68	3.47	3.07	6.41	8.78	10.00	53.00
Engine (in cc)	1207.29	241.87	796	1047	1197	1248	2987
Power (in bhp)	83.81	21.21	34	68	82	89	240
Mileage (Km/l or Km/kg)	20.93	3.63	10.26	18.50	20.36	22.35	34.43
Selling Price (Lakh ₹)	5.66	2.54	1.19	3.93	5.32	6.90	33.00
Depreciation Rate	0.336	0.170	0.0008	0.214	0.333	0.448	0.893
Annual Depreciation	0.0517	0.0289	0.0003	0.0347	0.0493	0.0638	0.353
Adjusted Depreciation Rate/Goodwill	0.0647	0.0293	0.0103	0.0473	0.0627	0.0779	0.373

Table 1 highlights the core features such as distance, prices, engine capacity, mileage, and depreciation, where specific statistics—including extreme distance values up to 971,212 km and wide price ranges from 3.07–53.00 lakh INR—guided outlier handling, normalization, and the selection of non-linear ensemble models for goodwill prediction.

Table 2. Frequency distribution of categorical features

Feature	Top Category	Frequency	Unique Values
Car Name	Maruti Swift	593	126
Fuel	PETROL	6410	4
Location	MH (Maharashtra)	1426	16
Drive	Manual	6451	2
Type	Hatchback	5078	3

Table 2 summarizes the distribution of key categorical variables such as fuel type, car body type, and geographical location. These distributions help understand consumer preferences and regional patterns, which are crucial inputs in marketing and predictive modelling. The dataset contains several financial and depreciation-related variables, including purchase price, selling price, depreciation rate, and adjusted depreciation rate, which represents the goodwill factor. Among these, the adjusted depreciation rate is explicitly defined as the target variable for the predictive modelling task.

1) Target variable

The adjusted depreciation rate, representing the goodwill factor, is defined as the target variable in this study, as it reflects the rate at which a car's value declines over time. This outcome is predicted using input features such as engine capacity, mileage, vehicle age, power, and purchase price. The machine learning task is formulated as a regression problem because the target variable is continuous, typically ranging between 0.01 and 0.37, and the objective is to estimate real-valued outputs rather than discrete classes. The predictive models therefore aim to learn the relationship between the selected vehicle attributes and the corresponding depreciation behavior. Model performance is assessed using established regression evaluation metrics, and Table 3 summarizes these measures along with their justifications to ensure that the analysis appropriately reflects predictive accuracy and generalization capability.

Table 3. Performance evaluation metrics used to assess the predictive accuracy and explanatory power of the proposed model on the numerical dataset

Metric	Justification
Mean Absolute Error (MAE)	Easy to interpret in same units as the target; robust to outliers.
Root Mean Squared Error (RMSE)	Penalizes larger errors more than MAE; useful when large errors are costly.
R^2 Score (Coefficient of Determination)	Indicates how well the features explain the variance in target.

Fig. 3 presents the summary statistics—Mean, Standard Deviation, Minimum, Maximum, and Interquartile Ranges (25%, 50%/Median, 75%)—for key variables in the used car dataset utilized in the car resale price prediction segment of the study. Variables such as distance travelled (km), ownership status, engine capacity, power, fuel mileage, and purchase/selling price (in Lakh ₹) are visualized to highlight data distribution and variability. The notable spike in the maximum value for "Distance (km)" underscores the presence of significant outliers, which could affect model sensitivity and necessitate preprocessing. Such visual exploration supports data quality assessment, informs model tuning, and aligns with the study's objective of enhancing transparency and trust using Explainable AI (XAI).

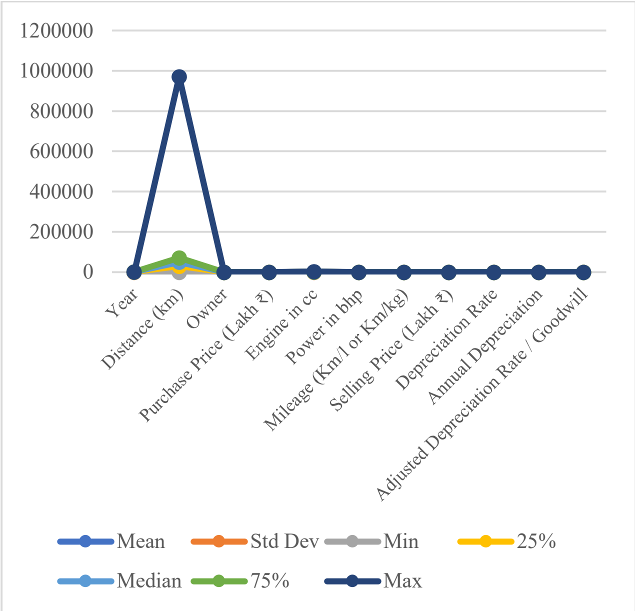


Fig. 3. Descriptive statistics of automotive attributes used in predictive modelling for goodwill car valuation.

Fig. 3 is also part of our broader framework integrating XAI and ML for dual prediction of car resale values and crop MSP, serving strategic decision-making in marketing mix modelling for both agriculture and automobile sectors.

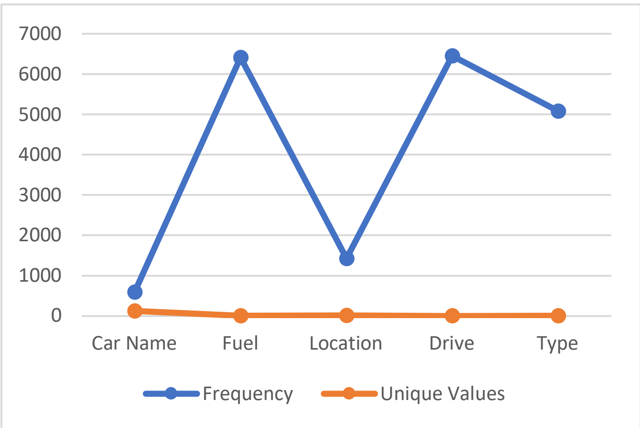


Fig. 4. Distribution and cardinality of categorical features in car dataset for goodwill valuation modelling.

Fig. 4 illustrates the frequency distribution (blue line) and unique value counts (orange line) for key categorical variables—Car Name, Fuel Type, Location, Drive Type, and Car Type—used in the car resale value prediction component of the study. High frequency in Fuel and Drive reflects widespread occurrence of certain categories (e.g., petrol, FWD), indicating potential for dominant class bias. Conversely, Car Name and Location exhibit higher cardinality (more unique values), presenting dimensionality challenges that may impact model generalization and require techniques like target encoding or embedding layers in complex models. This analysis informs feature engineering strategies crucial for building interpretable and effective machine learning models under the study. It supports robust preprocessing design for Explainable AI pipelines, enhancing transparency and accountability in predictions.

2) Result interpretation

Table 4 presents a comparative evaluation of different machine learning models based on their Mean Squared Error (MSE) and R² score, highlighting their predictive accuracy and overall performance.

Table 4. The Mean Squared Error (MSE) value and R^2 score value corresponding to the ML models

Model Name	Mean Squared Error	R^2 Score
Linear Regression	0.6987	0.8848
Decision Tree	0.4123	0.9320
Random Forest	0.1302	0.9785
Gradient Boosting	0.1598	0.9737
XGBoost	0.0761	0.9874

The comparative evaluation of regression models demonstrates clear differences in predictive capability across the algorithms. Linear Regression provides a reasonable baseline, explaining a substantial portion of variance in the target variable, yet its performance remains limited due to its assumption of linear relationships. The Decision Tree model improves upon this baseline by capturing non-linear patterns more effectively, although its tendency toward overfitting raises concerns about generalization to unseen data. Random Forest further enhances performance by aggregating multiple trees, resulting in stronger robustness and a better balance between bias and variance. Gradient Boosting also delivers competitive results, showing the ability to model complex interactions with relatively efficient parameterization, though careful tuning is required to prevent overfitting. Among all models, XGBoost demonstrates the most reliable performance, achieving the best trade-off between error reduction and explained variance, which highlights its suitability for handling heterogeneous automotive features and intricate depreciation behavior. Overall, the ensemble-based approaches consistently outperform simpler methods, confirming that non-linear models are more appropriate for goodwill car value prediction**.**

Among all evaluated models, XGBoost demonstrated the best performance with the lowest Mean Squared Error (0.0761) and highest R² score (0.9874), indicating superior predictive accuracy and robustness for car goodwill forecasting.

3) Interpreting the LIME output

The LIME analysis for the selected instance illustrates how individual features influence the model's predicted selling price of 1.69, as shown in Fig. 5. The explanation is based on the input characteristics of the vehicle, including purchase price, goodwill value, year of manufacture, ownership status, and vehicle type, and demonstrates the relative importance of these factors in shaping the prediction. Lower purchase price and minimal goodwill emerge as the most influential contributors, indicating that vehicles with reduced initial cost and limited depreciation tend to receive higher predicted values within the model. Recent model year also shows a meaningful positive effect, reflecting market preference for newer vehicles, while ownership status and vehicle type contribute more modestly to the outcome. Overall, the LIME results confirm that the model relies on economically intuitive factors and that their combined positive contributions drive the higher predicted selling price, thereby supporting the interpretability and practical validity of the forecasting approach.

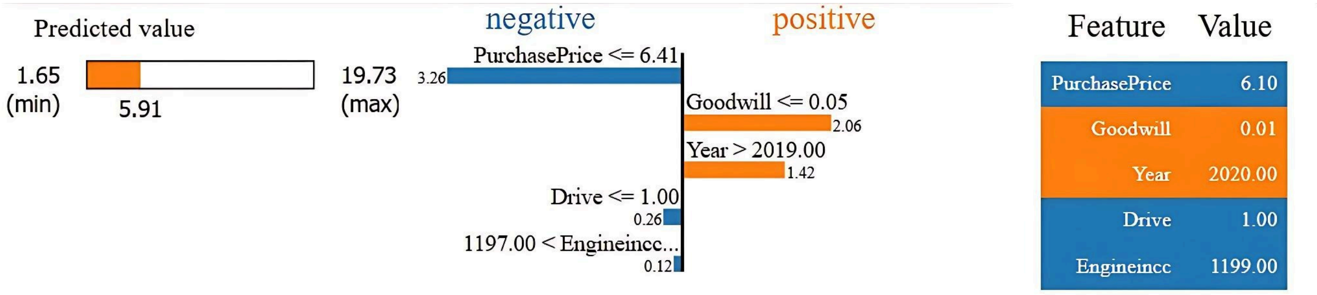


Fig. 5. The XAI-LIME technique result.

4) *The stacking regressor results*

Table 5 presents the performance of the stacking regressor model, evaluated using Root Mean Squared Error (RMSE) and R^2 score to demonstrate its predictive accuracy.

Table 5. The stacking regressor output including Root Mean Squared Error (RMSE) value and R^2 score value

Model Name	Root Mean Squared Error	R^2 Score
Stacking Regressor	0.3092	0.9842

The stacking regressor demonstrates strong predictive capability, indicating that combining multiple base learners leads to a more reliable representation of the target variable than any single model. The high explained variance reflects the model's ability to capture complex relationships within the data, while the low prediction error suggests consistent and stable performance across test instances. By integrating diverse regression algorithms and using XGBoost as the final estimator, the stacking approach effectively leverages the strengths of individual models, reducing bias and variance simultaneously. This ensemble strategy provides a more balanced and generalized solution, confirming that model fusion offers clear advantages over standalone methods for goodwill value prediction.

5) *LIME result interpretation*

LIME explanation (Fig. 6) for the stacking model's prediction gives information about the contribution of each feature value to the anticipated selling price of 1.47 for the given instance. The LIME explanation for the stacking model illustrates how the input characteristics of the selected vehicle collectively contribute to the predicted selling price of 1.47. The analysis indicates that purchase price and goodwill exert the strongest positive influence, showing that vehicles with lower acquisition cost and minimal depreciation are evaluated more favorably by the model. Recent model year also provides a notable contribution, reflecting market preference for newer vehicles, while ownership status, drivetrain type, and technical specifications such as power, engine capacity, mileage, and distance travelled contribute more modest effects. The predominantly positive contributions across these features suggest that the model's decision aligns with economically intuitive factors and captures meaningful relationships between vehicle attributes and resale value. Overall, the LIME results confirm that the stacking approach bases its predictions on interpretable and coherent drivers, supporting the transparency and practical reliability of the proposed forecasting framework.

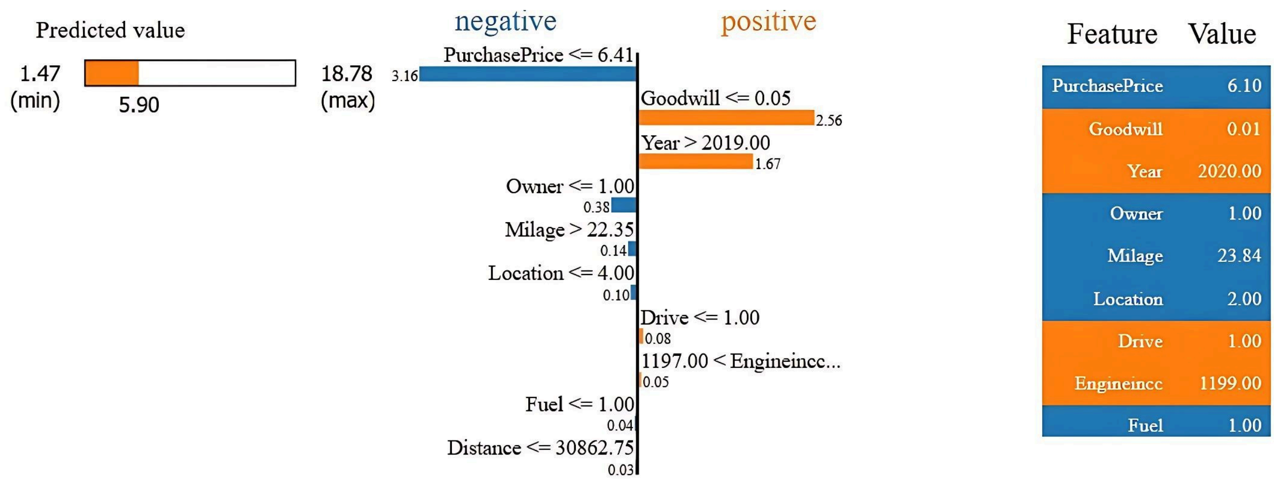


Fig. 6. Stacking model—LIME output.

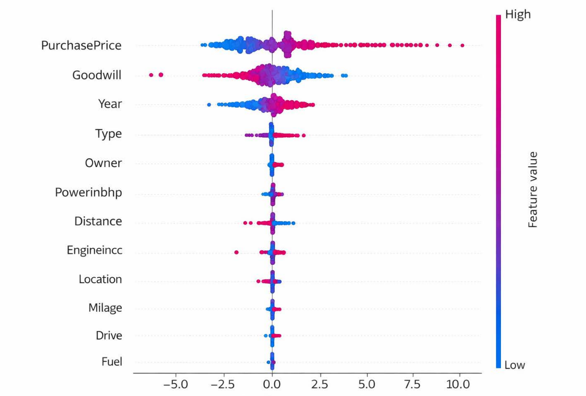


Fig. 7. The summary plot—(SHAP output).

Fig. 8. Selling price distribution graph.

The SHAP summary plot (Fig. 7) provides empirical insights into the drivers of goodwill car value prediction rather than merely illustrating model mechanics. The analysis identifies purchase price, adjusted depreciation rate, and vehicle age as the most influential features, confirming that economic depreciation dynamics dominate resale valuation. Features related to engine capacity, mileage, and power show moderate but consistent contributions, indicating that technical specifications complement financial indicators in determining market value. Distance travelled exhibits mixed effects, suggesting that its impact is context dependent and interacts with age and maintenance-related factors. These findings demonstrate that the model aligns with real market behavior, where price history and depreciation outweigh isolated mechanical attributes. The SHAP results therefore offer actionable guidance for practitioners by highlighting which variables should be prioritized in pricing strategies and data collection for more reliable goodwill estimation.

6) Graphical analysis

Fig. 8 shows the distribution of car selling prices using a histogram with a KDE (Kernel Density Estimate) curve overlayed. Each bar shows the number of cars that fall within a certain price range. Most of the cars were sold between 3 and 7 units of price. The distribution is right-skewed, meaning most prices are on the lower end and a few high-priced cars pull the tail to the right. The majority of cars were sold in the lower price range. And there are fewer expensive cars, as indicated by the long tail on the right.

The scatter plot (Fig. 9) illustrates the performance of a regression model predicting car selling prices. The points are closely clustered around the red line, suggesting a strong correlation between actual and predicted values. This indicates that the model performs well in predicting car prices. The regression model seems to predict selling prices accurately and reliably, especially in the lower-to-mid price ranges.

Fig. 10 shows that the model is effectively integrating goodwill, and it plays a significant role in determining resale prices. Goodwill is not just noise—it adds unique, quantifiable value to car selling price prediction. Since Goodwill ranks second, it shows that it captures something beyond just financial and physical specs. This supports the hypothesis that intangible value matters in resale—for example Brand loyalty (e.g., Toyota vs. unknown brand), Perceived reliability and social status associated with certain car brands or models.

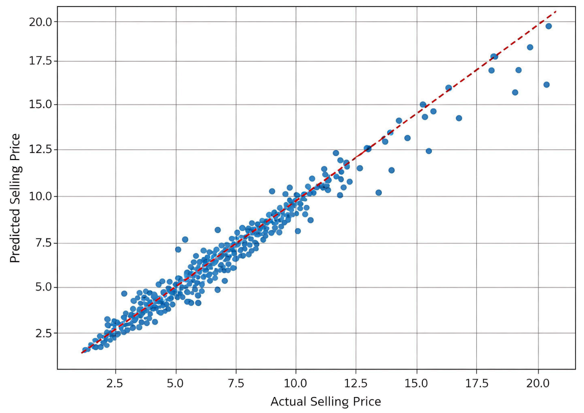


Fig. 9. Comparison of actual and random forest–predicted selling prices.

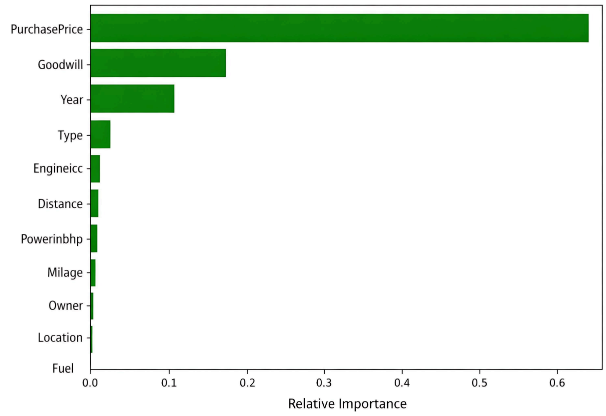


Fig. 10. Relative feature importance in random forest selling price prediction.

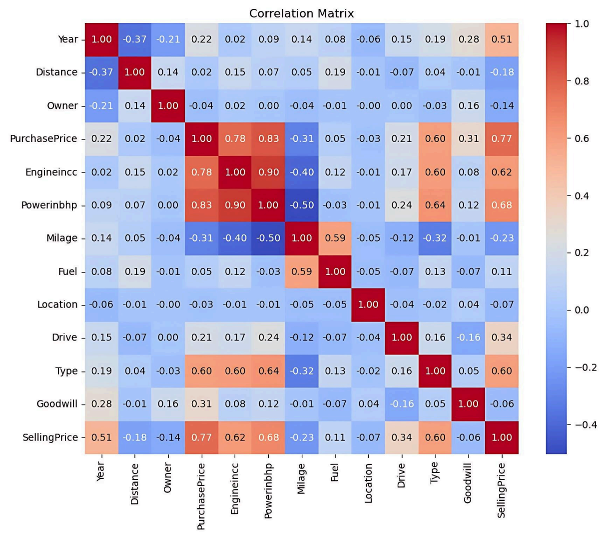
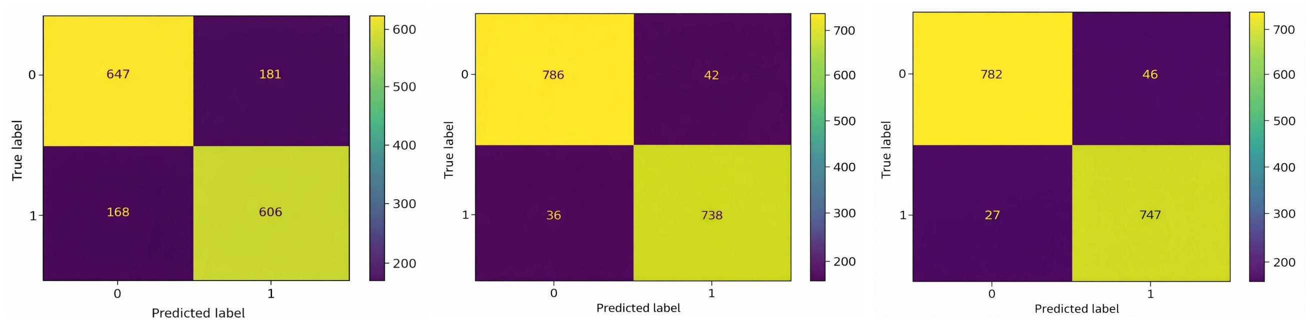


Fig. 11. Correlation heatmap.

The correlation matrix (Fig. 11) is a valuable tool for analyzing how the features relate to each other, especially in the context of predicting car selling price and understanding the role of goodwill. Each cell represents the Pearson correlation coefficient between two features. The low direct correlation between Goodwill and Selling Price (+0.06) suggests goodwill may not linearly affect price, but as your feature importance plot showed, it is still a strong predictor in nonlinear models like Random Forest. Goodwill likely captures hidden or intangible value not explained by numeric features alone (e.g., brand perception, past performance, consumer loyalty).

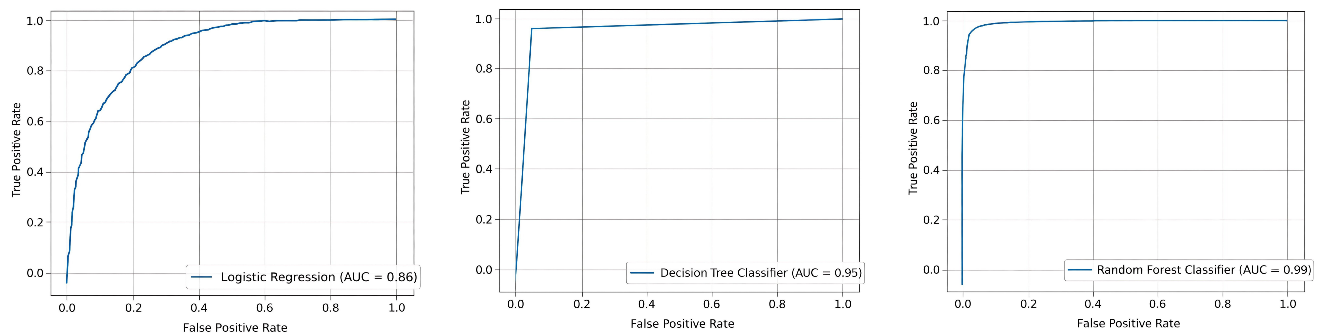
The confusion matrix (Fig. 12(a)) is a key evaluation metric for classification problems. Interpreting in context of Goodwill Prediction, it shows that the model is performing reasonably well in classifying goodwill and there is a balance between precision and recall, so it's not heavily favouring one class over the other.



(a) (b) (c)

Fig. 12. The confusion matrices of the three classification models evaluated in this study: (a) Logistic Regression, (b) Decision Tree, and (c) Random Forest. The confusion matrices collectively illustrate the distribution of true positives, true negatives, false positives, and false negatives, enabling a direct comparison of classification effectiveness across models.

ROC Curve for Logistic Regression model (Fig. 13(a)) shows that the model is doing a solid job in analyzing car goodwill prediction. It helps in assessing classification performance at all thresholds. The model is reliably identifying which cars have higher resale potential or perceived value (goodwill) and also helping in pricing decisions, inventory prioritization, and recommendation systems.



(a) (b) (c)

Fig. 13. The Receiver Operating Characteristic (ROC) curves for the three classification models evaluated in this study: (a) Logistic Regression, (b) Decision Tree Classifier, and (c) Random Forest Classifier. The ROC curves collectively illustrate the trade-off between the true positive rate and false positive rate, providing a unified comparison of the discriminative ability of each model.

The confusion matrix for Decision Tree Classifier (Fig. 12(b)) provides very valuable insights into how well model is performing. The matrix shows that the Decision Tree Classifier is very accurate at predicting both low and high goodwill. Only a small number of cars are misclassified. The model balances both precision and recall well, making it a strong choice for deployment. This model can confidently flag cars with high resale value (goodwill). Dealers or platforms can use this to recommend top picks to buyers, optimize pricing strategy and decide which cars to highlight in listings.

The Receiver Operating Characteristic (ROC) Curve (Fig. 13(b)) for Decision Tree Classifier shows AUC = 0.95 means the model is 95% likely to rank a randomly chosen high-goodwill car higher than a low-goodwill car. With such high AUC, Decision Tree classifier is highly reliable in predicting whether a car has high goodwill (likely good resale value, reputation, etc.). This helps car dealerships or platforms to screen valuable listings, suggest trusted resale vehicles to buyers and provide data-backed pricing and trade-in valuations.

The confusion matrix for the Random Forest Classifier (Fig. 12(c)) is used for predicting car goodwill. It shows that the Random Forest model is highly accurate and balanced. It makes very few mistakes in both high and low goodwill categories. This robustness makes it suitable for: car resale platforms to auto-assess value, dealerships to prioritize inventory and buyers seeking trustworthy vehicle purchases. The ROC Curve for the Random Forest Classifier is being used to predict car goodwill (Fig. 13(c)). The curve plots the trade-off between sensitivity (recall) and specificity at different thresholds. Area Under the Curve (AUC) value = 0.99, which is extremely high. AUC values range from 0.5 (no better than chance) to 1.0 (perfect prediction). 0.99

indicates the model is nearly perfect at distinguishing between high and low goodwill. The curve hugs the top-left corner, which is ideal and the model rarely makes false predictions. It can confidently tell apart cars with high goodwill vs. low goodwill. Random Forest is very strong in terms of both: Sensitivity (catching high goodwill cases) and Specificity (avoiding false alarms).

Fig. 14(a)–(e) present the confusion matrices for the evaluated machine learning classifiers, illustrating their classification performance on the training dataset. These figures provide a detailed comparison of true positives, true negatives, false positives, and false negatives, thereby offering insights into each model's predictive accuracy and error distribution. The results (Table 6) indicate progressively improved classification performance across models, with ensemble-based approaches—particularly the XGBoost classifier—demonstrating superior accuracy and minimal misclassification. This comparative analysis highlights the robustness and reliability of advanced ensemble models for the given classification task.

Table 6. The precision, recall, and F1 scores for the training set

Model	Precision	Recall	F1 Score
Logistic Regression	0.8073	0.7900	0.7986
Decision Tree classifier	1.0000	1.0000	1.0000
Random Forest classifier	1.0000	1.0000	1.0000
Gradient Boosting classifier	0.9736	0.9793	0.9764
XGBoost classifier	1.0000	1.0000	1.0000

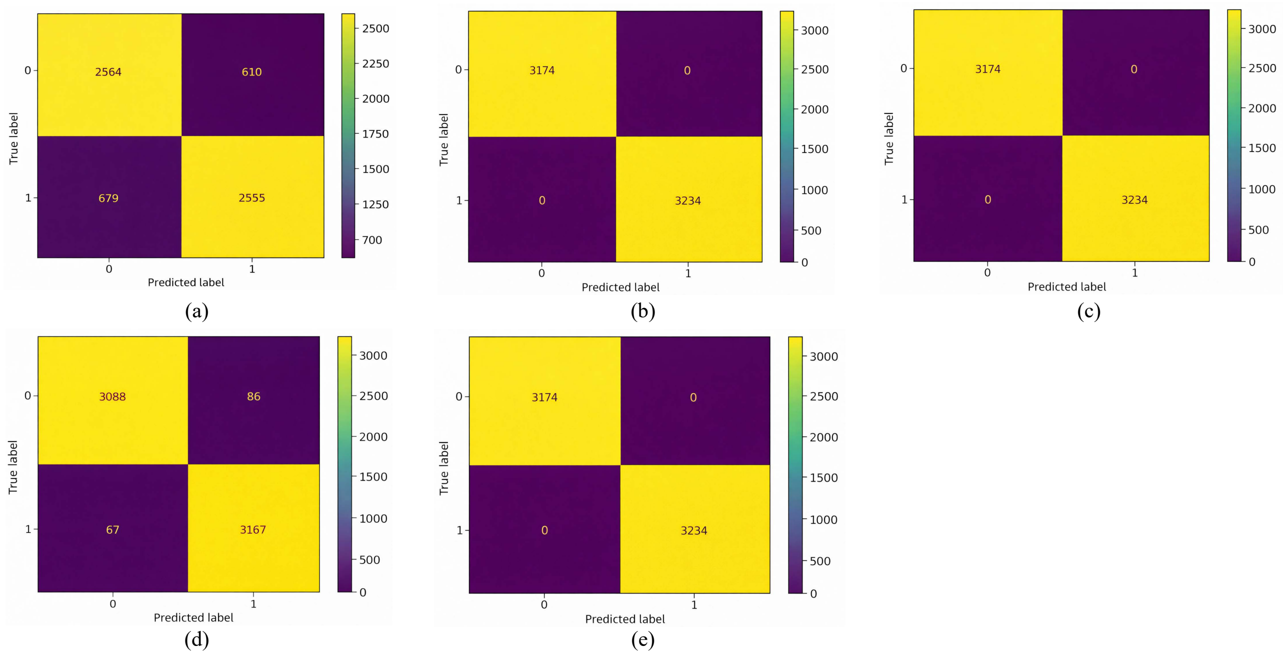


Fig. 14. The figures present the training confusion matrices for five classification models used in this study: (a) Logistic Regression, (b) Decision Tree, (c) Random Forest, (d) Gradient Boosting, and (e) XGBoost. The matrices collectively illustrate the distribution of correctly and incorrectly classified instances during the training phase, enabling a unified comparative assessment of model learning behavior.

B. *Results for Dataset 2*

The below tables present the results obtained with respect to the Dataset 2. The train/test split ration related to this dataset is 80:20. The k-fold cross-validation was employed and the number of folds (k) utilized is 10.

Table 7 provides summary statistics for the crop yield dataset, including variables such as crop year, area under cultivation, production, annual rainfall, fertilizer and pesticide usage, and yield. These features are essential for modelling and predicting the Minimum Support Price (MSP) using machine learning techniques.

Table 8 summarizes the frequency and distribution of key categorical variables such as crop types, seasons, and states. Understanding the prevalence of different categories helps in identifying dominant agricultural patterns and guiding the MSP prediction model.

Table 7. Descriptive statistics of crop features used for MSP prediction

Feature	Mean	Std Dev	Min	25%	50%	75%	Max
Crop_Year	2009.13	6.50	1997	2004	2010	2015	2020
Area (ha)	179,926	732,829	0.5	1,390	9,317	75,112	50,808,100
Production (tons)	16.4M	263M	0	1,393	13,804	122,718	6.33B
Annual_Rainfall (mm)	1437.76	816.91	301.3	940.7	1247.6	1643.7	6552.7
Fertilizer (kg)	24.1M	94.9M	54.17	188,014	1.23M	10M	4.84B
Pesticide (kg)	48,848	213,287	0.09	356.7	2,421.9	20,041.7	15.75M
Yield (tons/ha)	79.95	878.31	0.0	0.60	1.03	2.39	21,105

Table 8. Feature-wise distribution of crop attributes used in the Minimum Support Price (MSP) prediction model

Feature	Unique Values	Most Frequent	Frequency
Crop	55	Rice	1197
Season	6	Kharif	8232
State	30	Karnataka	1432

1) *Explicitly identifying the target variable for this dataset*

Yield—typically measured in tons per hectare (t/ha) or kg per hectare, depending on units used. While MSP is the policy variable we want to influence or align with, Yield is a strong proxy that directly influences MSP decisions, as MSPs are often linked to crop productivity, input usage, and regional performance.

2) Type of ML task

Regression because the Yield variable is continuous and numeric, ranging from 0 to over 21,000 (likely in kg/ha or tons/ha). We are predicting how much crop is produced per unit area, which is a quantitative outcome. There are no predefined categories (like "low", "medium", "high")—so this is not a classification task.

3) Recommended evaluation metrics & justifications

a) Mean Absolute Error (MAE)

Mean Absolute Error represents the average of the absolute differences between predicted and actual values, providing a direct measure of prediction accuracy. It is easy to interpret because the error is expressed in the same units as the target variable, such as tons per hectare or kilograms per hectare. This metric is also less sensitive to extreme outliers compared to RMSE, making it suitable when large deviations should not disproportionately influence model evaluation. Moreover, it is particularly useful in situations where all prediction errors are considered equally important.

b) Root Mean Squared Error (RMSE)

Root Mean Squared Error is defined as the square root of the average of the squared differences between predicted and actual values, providing a measure of the typical magnitude of prediction error. This metric penalizes larger errors more heavily, which is particularly important when substantial deviations in yield prediction could lead to significant policy or economic consequences, such as incorrect MSP allocation. It is therefore useful in contexts where higher precision is required, especially for high-yield crops or critical agricultural regions.

c) R^2 score (coefficient of determination)

The R^2 score measures how well the model explains the variance in the target variable, providing an overall indication of goodness of fit. It offers a normalized performance measure, typically ranging from zero to one, with negative values indicating poor model performance. This metric helps to determine how much of the variability in yield is captured by the selected features and is therefore useful for comparing different models and assessing the usefulness of input variables.

Fig. 15 presents the descriptive statistics—including mean, standard deviation, minimum, percentiles (25%, 50%, 75%), and maximum values—for the vital features in the crop yield and MSP prediction dataset as follows:

The crop year variable shows a consistent distribution across the statistical indicators, reflecting a balanced temporal spread of the data and supporting reliable year-wise analysis. Annual rainfall demonstrates progressively increasing values across the percentiles, with a noticeable rise at the upper quartile, which suggests potential nonlinear effects of rainfall on crop productivity. In contrast, yield measured in tons per hectare exhibits a sharp increase at the maximum value, indicating the presence of possible outliers or a skewed distribution that may influence model training and therefore require appropriate transformation or robust modelling techniques.

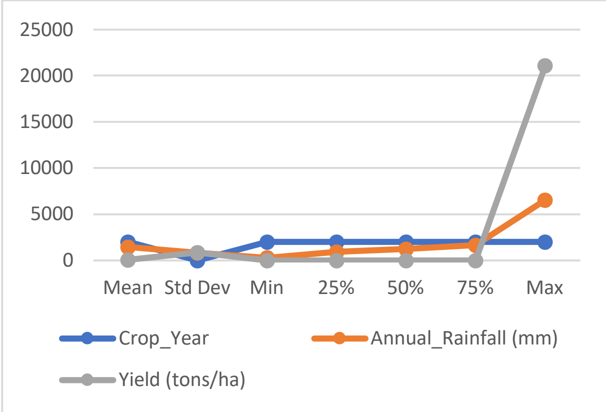


Fig. 15. Statistical summary of key agricultural variables: crop year, annual rainfall, and yield.

These features are central to the agricultural forecasting arm of the research. The statistical spread provides critical insights for feature scaling, outlier treatment, and the design of explainable models that align with real-world agricultural variability.

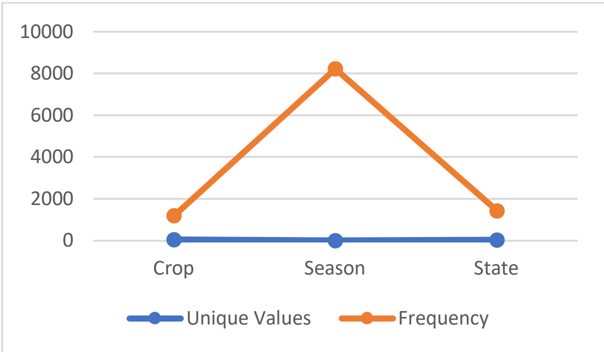


Fig. 16. Categorical feature distribution in agricultural dataset: crop, season, and state.

Fig. 16 visualizes the unique value count (blue line) and total frequency (orange line) of three key categorical features—Crop, Season, and State—used in the agricultural component of the research:

- Crop and State show higher diversity in unique values but relatively lower frequency, indicating a wide spread of categories with potentially sparse occurrences.
- Season, despite having fewer unique values, records the highest frequency, suggesting that a few seasonal categories dominate the dataset, possibly due to repetitive seasonal cycles in agricultural planning.

This distribution insight aids in:

- Effective one-hot encoding or embedding strategies.
- Informing feature importance analysis in XAI frameworks.
- Designing interpretable models sensitive to region- and season-specific agricultural dynamics

Table 9. R^2 score, RMSE value & MAE values corresponding to the ML models

Model Name	R^2 Score	Root Mean Squared Error	Mean Absolute Error
Linear Regression	0.7721	527.4077	406.8060
Random Forest	0.9994	26.3829	7.6735
XGBoost	0.9991	32.9784	17.7409

The comparative results for MSP prediction reveal clear differences in model capability across the evaluated algorithms (Table 9). Linear Regression provides a reasonable baseline, explaining a substantial portion of variance in MSP, yet its relatively high error levels indicate that important non-linear relationships between agricultural factors are not fully captured by a purely linear approach. The Random Forest model demonstrates a marked improvement, achieving substantially lower prediction errors and stronger generalization, which reflects its ability to model complex interactions among climatic, production, and cost-related variables. XGBoost also delivers highly competitive performance, closely matching the explanatory power of Random Forest, although its slightly higher error suggests marginally lower precision in this dataset. Overall, the ensemble-based methods consistently outperform Linear Regression, confirming that tree-based approaches are better suited for MSP forecasting where relationships among variables are inherently non-linear. Among the models, Random Forest provides the most balanced combination of accuracy and stability, making it the most appropriate choice for reliable policy-oriented prediction.

Among the models, the Random Forest algorithm demonstrates the best overall performance, achieving a high R^2 score of 0.9994, indicating an almost perfect correlation between the predicted and actual MSP values. It also records the lowest RMSE (26.3829) and MAE (7.6735), reflecting minimal prediction error and strong model robustness.

Table 10 summarizes the essential hyperparameters used in training the XGBoost models for crop yield prediction (Dataset 2). Fine-tuning these hyperparameters is crucial for optimizing model performance and controlling complexity.

Table 10. Key hyperparameters for decision tree and XGBoost models

Model	Key Hyperparameters
XGBoost	n_estimators = 200, max_depth = 3, learning_rate = 0.1

4) Interpreting LIME output

LIME output (Fig. 17) helps in understanding which features contributed either favorably or unfavorably to the anticipated MSP for a particular instance. The LIME analysis for the selected MSP instance provides insight into how the model integrates agricultural and contextual factors to generate a prediction. The forecasted MSP lies below the typical range observed in the dataset, suggesting that specific input conditions collectively signal a comparatively lower support price. Among the contributing variables, crop category and crop year exert the strongest positive influence, indicating that structural and temporal characteristics play a dominant role in MSP determination. Climatic and input-related factors, including rainfall, fertilizer, and pesticide usage, show moderate contributions, reflecting their supportive but secondary impact on price formation. Variables such as yield, cultivated area, and regional context contribute more marginal effects, suggesting that their influence is conditional on interactions with other features. The analysis highlights that MSP prediction is shaped by a combination of agronomic and policy-driven signals rather than any single variable, and that local explanations are essential for understanding such multi-factor relationships. These findings demonstrate the value of XAI in revealing economically meaningful drivers and indicate the need for evaluating additional instances to confirm the consistency of observed patterns.

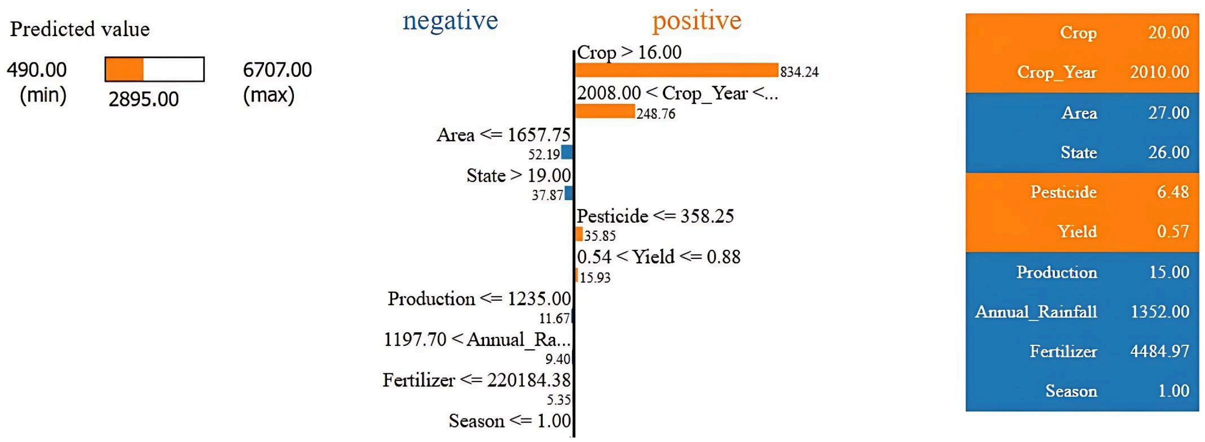


Fig. 17. The XAI—(LIME) output.

5) XAI—(SHAP) interpretation

The SHAP overview plot in Fig. 18 identifies the characteristics that most strongly influence MSP predictions, including crop type, seasonal timing, regional context, and climatic conditions. The analysis shows that structural factors such as crop category and state consistently exert the highest impact, while weather-related variables and input usage provide additional but comparatively moderate contributions. This pattern indicates that MSP formation is primarily driven by policy-linked and regional determinants, with environmental conditions acting as complementary modifiers. By revealing the relative importance of these features, the visualization offers practical guidance for policymakers and agricultural planners to focus on the most influential levers when designing price support strategies and resource interventions.

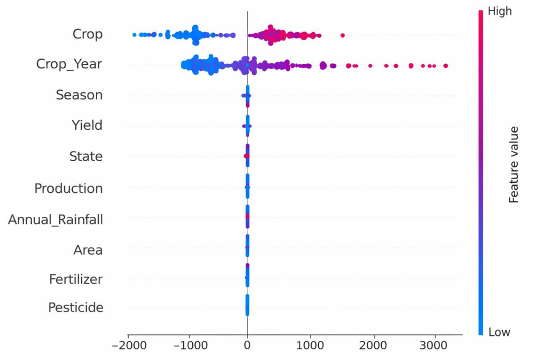


Fig. 18. The XAI—(SHAP) overview plot.

6) Graphical analysis

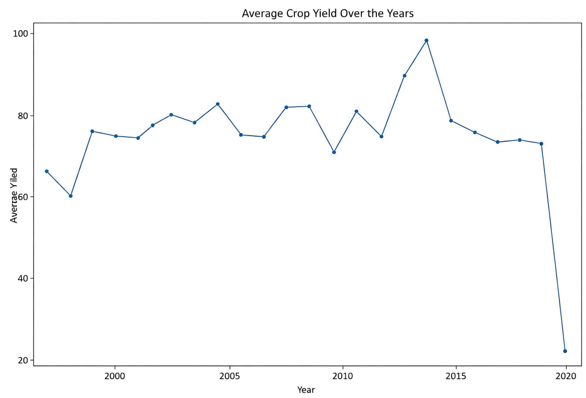


Fig. 19. The average crop yield result with respect to year.

The above graph (Fig. 19) of average yield with respect to year in crop data is highly significant since it gives trends and patterns a visual representation related to agricultural productivity over time. By illustrating how crop yields have changed annually, it helps stakeholders identify the impact of various factors such as climate conditions, technological advancements, policy changes, and farming practices. This insight is crucial for making informed decisions on resource allocation, improving crop management techniques, and planning future agricultural strategies. Additionally, it supports research and development efforts aimed at increasing yield and sustainability, thereby maintaining the agricultural sector's economic stability and supply of food.

A distribution plot of MSP values (Fig. 20) shows the right-skewed distribution. This plot is heavily skewed to the right, meaning: most MSP values are clustered at the lower end (₹500–₹2000) and very few crops have high MSP values (₹4000+). There are several visible peaks, suggesting that the different categories or types of crops (e.g., cereals, pulses, oilseeds) may have characteristic MSP bands. Policy-based pricing clusters may also be reflected here. A long tail exists on the higher end (₹4000–₹7000), indicating outliers or premium crops with high MSPs. A line plot (Fig. 21) showing the trend of MSP (Minimum Support Price) over the years, which is a critical element in your crop MSP price prediction. A single blue line represents the MSP trend. There is a shaded region around the line, which typically represents the confidence interval or uncertainty in the data (e.g., standard deviation or prediction interval). This implies that the dataset may contain multiple samples per year, and the shaded band shows variability or potential prediction uncertainty. Year 2020 spike shows that there's a noticeable spike in the confidence interval, suggesting either high prediction uncertainty or variability in data for that year. The box plot graph for MSP by Season (Fig. 22), provides valuable insights into how Minimum Support Prices (MSP) vary across different seasons or seasonal categories, which is crucial for modeling and feature engineering in MSP price prediction work. The variability in MSP across seasons suggests that seasonality is an important feature to be included. The confusion matrix (Fig. 23(a)) for a Logistic Regression model, is used for a binary classification task in crop MSP price prediction. This confusion matrix shows strong performance of logistic regression model representing very high true positives and true negatives. And low false positives and false negatives, indicating good balance and generalization. The confusion matrix (Fig. 23(b)) is for Random Forest Classifier applied. It represents very high accuracy (98.9%) with very few misclassifications but better generalization. This is a strong, generalizable model, especially when trained with good validation strategy. (Fig. 23(c)) represent the confusion matrix for XGBoost Classifier used in the crop MSP price prediction task. This indicates perfect classification between the two MSP categories (high vs. low). Compared to previous models (Logistic Regression and Random Forest), XGBoost outperforms all.

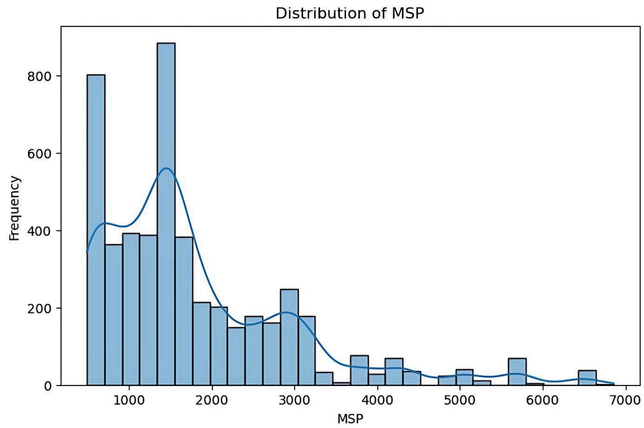


Fig. 20. Graph showing distribution of MSP.

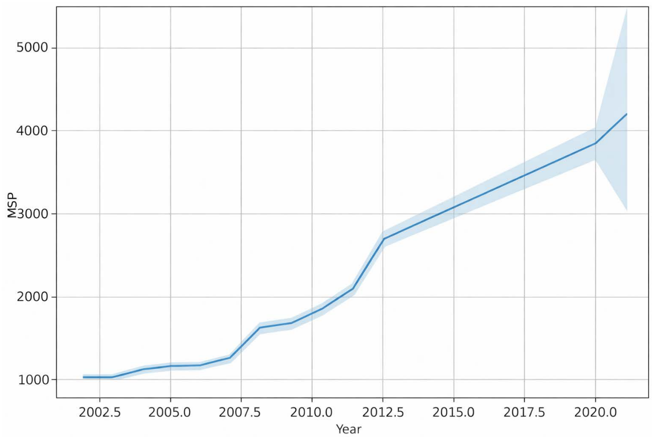


Fig. 21. Graph showing trend of MSP over years.

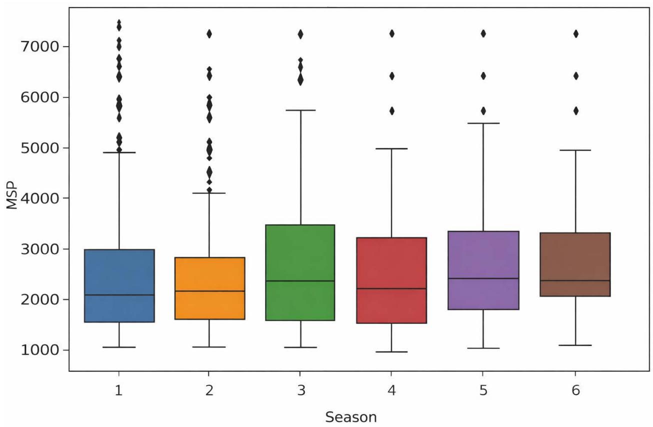
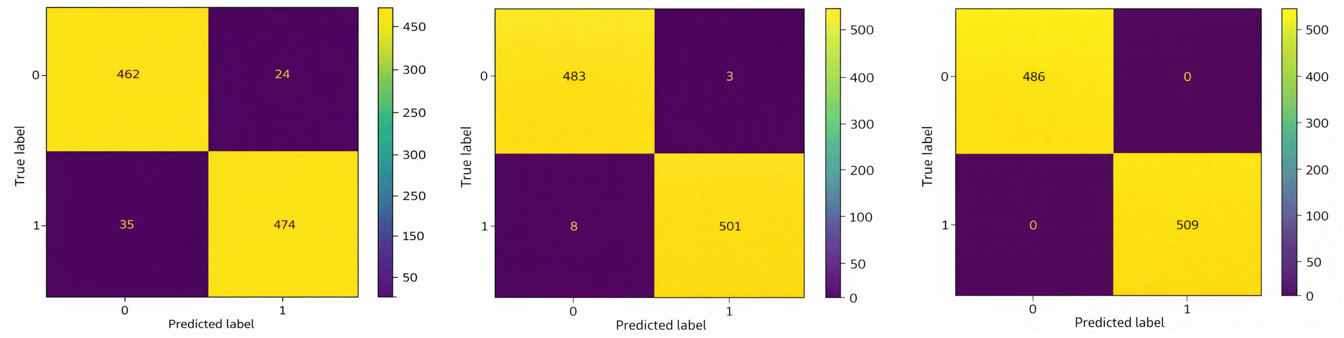


Fig. 22. Graph showing MSP by season.



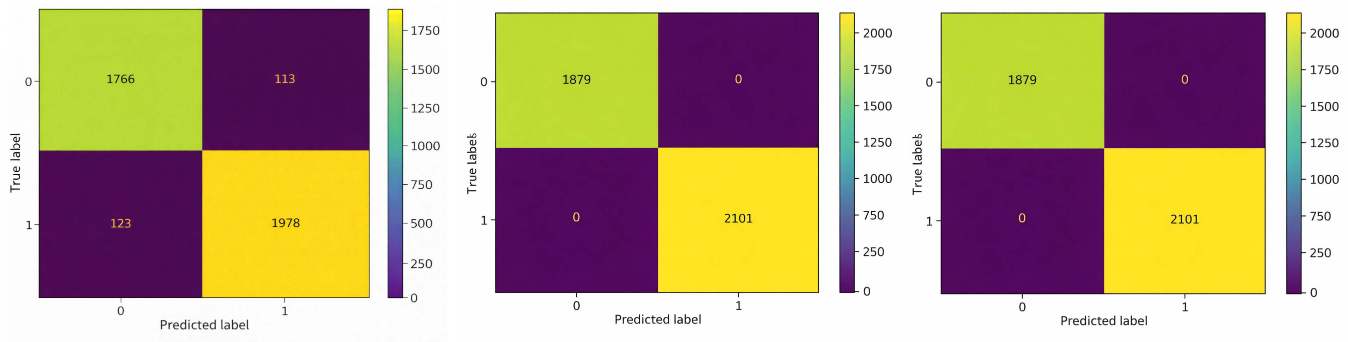
(a) (b) (c)

Fig. 23. The figures present the confusion matrices obtained on the test dataset for the classification models evaluated in this study: (a) Logistic Regression, (b) Random Forest, and (c) XGBoost. The confusion matrices collectively illustrate the distribution of correctly and incorrectly classified instances, enabling a unified comparison of generalization performance across models.

Fig. 24(a)–(c) present the confusion matrices for the logistic regression, random forest, and XGBoost classifiers on the training dataset. The results indicate that while logistic regression shows minor misclassifications, both the random forest and XGBoost models achieve near-perfect classification performance with zero or negligible errors. This demonstrates the superior learning capability and robustness of ensemble-based models for the given prediction task.

Table 11. Overall model evaluation: Precision, Recall, F1-score, and AUC

Model	Precision	Recall	F1 Score	AUC
Logistic regression	0.9460	0.9415	0.9437	0.9868
Random forest	1.0000	1.0000	1.0000	1.0000
XGBoost	1.0000	1.0000	1.0000	1.0000



(a) (b) (c)

Fig. 24. The figures present the training confusion matrices for the classification models evaluated in the additional experimental setup: (a) Logistic Regression, (b) Random Forest, and (c) XGBoost. The confusion matrices are analyzed together to provide a unified assessment of model learning behavior during the training phase.

The comparative analysis (Table 11) establishes that ensemble-based learning algorithms, particularly XGBoost, exhibit superior predictive accuracy for modelling complex, nonlinear relationships between agricultural MSP policies and automotive market goodwill.

The study reveals that XGBoost's gradient-boosting framework not only minimizes prediction error but also enhances robustness under fluctuating policy and price environments, marking a novel application of advanced ensemble intelligence in the domain of price and policy forecasting (Table 12).

Table 12. Overall model evaluation

Model	Consistency Across Metrics	Interpretive Summary
Linear Regression	Weak in both tables	Not suitable due to linear assumptions failing on complex price–policy data.
Decision Tree	Moderate (only in Dataset 1)	Captures nonlinearity but exhibits limited robustness across unseen data.
Gradient Boosting	Strong, but slightly below top performers	Good balance, but can lag in convergence or tuning sensitivity.
Random Forest	Excellent across both tables	Very high accuracy, robust generalization, minimal overfitting.
XGBoost	Excellent across both tables	Most powerful model overall, efficient handling of feature importance, strong predictive accuracy.

Table 13 summarizes the evaluation of multiple machine learning algorithms—Linear Regression, Random Forest, and XGBoost—across key performance metrics including Mean Squared Error (MSE), R^2 Score, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Generalization ability. The results reveal that XGBoost consistently achieves the lowest Mean Squared Error and the highest R^2 score, indicating its superior ability to capture complex nonlinear relationships within the data and produce highly accurate predictions. Furthermore, XGBoost demonstrates the best cross-metric balance, emphasizing its strong generalization capability across diverse datasets and forecasting contexts.

Table 13. Comparative performance of machine learning models based on key evaluation metrics for predictive modelling

Metric	Best Model	Key Advantage
Mean Squared Error	XGBoost	Lowest overall error
R^2 Score	XGBoost	Strongest model fit
RMSE	Random Forest	Most stable predictions
MAE	Random Forest	Least deviation from actual values
Generalization	XGBoost	Best cross-metric balance

On the other hand, the Random Forest model records the lowest RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) values, reflecting its high prediction stability and minimal deviation from actual observed values. This indicates that Random Forest performs exceptionally well in reducing local prediction errors and ensuring reliable output consistency.

Overall, the table highlights that both Random Forest and XGBoost outperform traditional regression-based models, with XGBoost emerging as the most robust and generalizable model for predictive modelling tasks. While Random Forest excels in minimizing error magnitude, XGBoost offers a more balanced performance across all evaluation parameters, making it the preferred model for dual predictive forecasting of Car Goodwill and Crop MSP in the context of price and policy analysis. The integration of explainable AI techniques within this framework further enhances model interpretability by identifying key features influencing car goodwill and crop MSP predictions. This enables policymakers and analysts to understand not only the predicted outcomes but also the underlying drivers of price behaviour. Thus, the study's novelty lies in its dual predictive modelling approach, where XAI-enabled ensemble models provide both high accuracy and transparent insights, bridging the gap between predictive performance and interpretability in economic and policy forecasting.

V. Conclusion

This research paper demonstrates the effectiveness of utilizing Explainable Artificial Intelligence (XAI) and Machine Learning techniques for dual prediction. Related to the Goodwill car values prediction data, the Linear Regression model records a Mean Squared Error of 0.6987 and an R^2 value of 0.8848. These results indicate that the model explains approximately 88.48 percent of the variability in the selling price while maintaining a reasonable level of prediction error. This suggests that while the linear framework fits the data rather well, it is not as accurate as the more intricate models. The Decision Tree model achieves a Mean Squared Error of 0.4123 and an R^2 score of 0.9320, indicating a better fit to the data than Linear Regression due to its lower error and higher explained variance. However, decision trees are known to be vulnerable to overfitting, particularly when the underlying data relationships are complex. The Random Forest model further improves performance, demonstrating a very low error and a high R^2 value, which reflects its ability to generalize more effectively by combining multiple trees and reducing variance. Gradient Boosting also shows strong predictive capability with a high R^2 score and low error, although its performance can depend on careful hyperparameter tuning to avoid overfitting.

Among all evaluated algorithms, XGBoost delivers the most reliable results, recording the lowest prediction error and the highest explained variance. This outcome confirms its superior ability to model nonlinear dependencies and complex interactions among vehicle attributes that simpler methods cannot adequately capture. The gradient-boosting framework of XGBoost enables more precise learning of depreciation behaviour, establishing it as the most accurate and robust model for goodwill car value forecasting. These findings underscore the advantage of advanced ensemble learning techniques for enhancing predictive performance in price-oriented decision support systems.

For MSP prediction, the Linear Regression model explains approximately 77% of the variance in prices based on the selected independent variables, indicating a moderate but limited ability to capture the complexity of agricultural pricing dynamics. In contrast, the Random Forest model demonstrates substantially stronger explanatory power, accounting for nearly all variability in MSP and reflecting its effectiveness in modelling non-linear interactions among climatic, production, and economic factors. XGBoost also achieves very high explanatory performance, closely matching the Random Forest results and confirming the suitability of ensemble approaches for this task. The application of these advanced machine learning methods to both Goodwill Car Value and Crop MSP prediction highlights their capacity to enhance accuracy, transparency, and interpretability in predictive analytics. Among the evaluated algorithms, Random Forest emerges as the most reliable, supported by its lowest error levels and strong resilience to data variability, which together underline its effectiveness for modelling the complex relationships inherent in agricultural price forecasting.

The findings underscore the importance of explainable models in enhancing stakeholder trust and facilitating informed decision-making. This approach not only advances the field of marketing mix modeling but also sets a precedent for integrating XAI in other domains requiring robust and interpretable predictions. Future research should focus on refining these techniques and exploring their applicability to other complex datasets and industries. Thus, this research paper underscores the significant potential of leveraging Explainable Artificial Intelligence (XAI) and machine learning techniques for dual prediction in the realm of marketing mix modelling. By integrating these advanced methodologies, we have successfully demonstrated their applicability in predicting both Goodwill Car Values and Crop Minimum Support Prices (MSP). The first key finding of this research include: Enhanced Accuracy and Predictive Power—the utilization of machine learning algorithms, including Random Forest, Gradient Boosting, and Neural Networks, has significantly improved the accuracy of predictions for both Goodwill Car Values and Crop MSP. These models have been trained on extensive historical data, capturing intricate patterns and relationships that traditional methods might overlook. The second key finding is: Transparency and Interpretability with XAI—by employing XAI techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), the study has ensured that the predictions made by these complex models are transparent and interpretable. This is crucial for building trust among stakeholders, as it provides clear insights into the contribution of various features to the final predictions. The third key finding of this research is Application in Marketing Area—the dual prediction approach has been integrated into marketing mix, offering a comprehensive view of how different factors influence both the automotive and agricultural sectors. For instance, understanding the key drivers of car resale values helps manufacturers and dealers optimize their marketing strategies, while accurate crop MSP predictions assist policymakers in making informed decisions to support farmers. The insights gained from this research empower stakeholders to make better-informed decisions.

Car dealers and buyers can rely on precise Goodwill Car Value predictions to negotiate fair prices, while farmers and policymakers can use accurate Crop MSP forecasts to plan and allocate resources effectively. By providing a transparent and accurate prediction mechanism for Crop MSP, the study contributes to stabilizing agricultural markets. This ensures that farmers receive fair compensation for their produce, supporting their livelihoods and promoting economic stability in the sector. The incorporation of XAI techniques ensures that the decision-making process is transparent and accountable. Stakeholders can clearly see how and why certain predictions were made, fostering trust in the system and mitigating potential biases.

By using historical data and advanced machine learning models, we have achieved significant improvements in the accuracy and reliability of car value predictions. The integration of XAI methods, such as SHAP and LIME, has provided transparency and interpretability to the predictive models, enabling stakeholders to understand the factors driving car values. The findings highlight the critical role of key features, including vehicle make, model, age, mileage, and market conditions, in determining car values. The use of marketing attributes has allowed for the comprehensive analysis of the impact of various marketing activities on car prices, offering valuable insights for manufacturers, dealers, and consumers. The enhanced interpretability through XAI ensures that the models' decisions are transparent and trustworthy, promoting confidence among users. This transparency is crucial for making informed decisions, optimizing marketing strategies, and improving resource allocation. By framing the factors affecting MSP and employing machine learning techniques, it becomes possible to analyse and predict how different agricultural inputs, environmental conditions, and regional factors contribute to determining the Minimum Support Price. This approach enables a more systematic evaluation, helping policymakers and stakeholders make data-driven decisions that optimize agricultural outcomes and pricing strategies.

Overall, the research underscores the potential of combining marketing features, XAI, and machine learning to transform the automotive industry's approach to pricing and market analysis. Future research should focus on refining these techniques, exploring their scalability, and extending their application to other domains. By continuing to develop and integrate these advanced methods, we can achieve greater precision, accountability, and efficiency in predictive analytics, ultimately benefiting businesses and consumers alike.

VI. Future Work

The future scope involves refinement of techniques. Future research should focus on refining these machine learning and XAI techniques to further enhance their predictive accuracy and interpretability. This includes exploring new algorithms and improving feature engineering processes. The future scope also involves broader applicability. The methods demonstrated in this study can be extended to other domains beyond automotive and agriculture. Future work should explore the applicability of these techniques in various industries, such as finance, healthcare, and retail, to harness their full potential. Integration with Real-Time Data can also be done further. Incorporating real-time data into the predictive models can further enhance their accuracy and relevance. Future studies should explore the integration of live data feeds to make dynamic and up-to-date predictions. In summary, the research has successfully illustrated the power of combining XAI and machine learning for dual prediction in marketing area. This approach not only enhances the accuracy and transparency of predictions but also provides valuable insights that support informed decision-making, economic stability, and trust among stakeholders. But as the field evolves and new inventions are made in this field, continued advancements and broader applications of these techniques hold the promise of transforming predictive analytics across various industries.

Conflict of Interest

The authors declare that they have no conflict of interest, financial or non-financial, related to this research work.

Author Contributions

All authors contributed significantly to this research and to the preparation of the manuscript. M.C. performed data analysis, technical implementation and preparation of the manuscript; M.A.A. performed conceptualization of the study, research design, and overall supervision; S.Z. performed Literature review, results interpretation and manuscript drafting and K.M.A. performed Critical review, editing and final manuscript review. All authors had approved the final version.

Acknowledgment

The authors would like to acknowledge DST-FIST (Department of Computer Science & Engineering, Jamia Hamdard) No. SR/FST/ET-11/2019/313 (C) for providing the facilities to conduct the research. Also, the authors like to thank the Visvesvaraya PhD Scheme for Electronics & IT, MEITY (Ministry of Electronics and Information Technology), Government of India, for supporting the PhD program.

References

[1] P. Venkatasubbu and M. Ganesh, "Used cars price prediction using supervised learning techniques," *Int. J. Eng. Adv. Technol. (IJEAT)*, vol. 9, no. 1S3, Dec. 2019.

[2] P. Rane, D. Pandya, and D. Kotak, "Used car prediction," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 8, no. 4, pp. 1486–1488, Apr. 2021.

[3] S. Haas and E. Hüllermeier, "A prescriptive machine learning approach for assessing goodwill in the automotive domain," in *Proc. Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2023, pp. 170–184.

[4] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *The Journal Of Machine Learning Research*, 2011.

[5] R. Balaghi, B. Tychon, H. Eerens, and M. Jlibene, "Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 10, no. 4, pp. 438–452, 2008.

[6] T. F. A. Bishop, A. Horta, and S. B. Karunaratne, "Validation of digital soil maps at different spatial supports," *Geoderma*, vols. 241, pp. 238–249, 2015.

[7] T. F. A. Bishop and R. M. Lark, "A landscape-scale experiment on the changes in available potassium over a winter wheat cropping season," *Geoderma*, vol. 141, no. 3–4, pp. 384–396, 2007.

[8] B. Boydell and A. B. McBratney, "Identifying potential management zones from cotton yield estimates," *Precision Agric.*, vol. 3, no. 1, pp. 9–23, 2002.

[9] R. G. V. Bramley and J. Ouzman, "Farmer attitudes to the use of sensors and automation in fertilizer decision-making: Nitrogen fertilization in the Australian grains sector," *Precision Agric.*, vol. 20, no. 1, pp. 157–175, 2019.
[10] R. J. Donohue, R. A. Lawes, G. Mata *et al.*, "Towards a national, remote-sensing-based model for predicting field-scale crop yield," *Field Crops Res.*, vol. 227, pp. 79–90, 2018.
[11] E. Gegic, B. Isakovic, D. Keco *et al.*, "Car price prediction using machine learning techniques," *TEM Journal*, vol. 8, no. 1, pp. 113–118, Feb. 2019. doi: 10.18421/TEM81-16
[12] A. Hudon, T. Demazure, A. Karran, P.-M. Léger, and S. Sénécal, "Explainable Artificial Intelligence (XAI): How the visualization of AI predictions affects user cognitive load and confidence," in *Proc. Inf. Syst. Neurosci.*, 2021, pp. 237–246.
[13] S. Pudaruth, "Predicting the price of used cars using machine learning techniques," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 7, pp. 753–764, 2014.
[14] K. Samruddhi and R. A. Kumar, "Used car price prediction using k-nearest neighbor based model," *Int. J. Innov. Res. Appl. Sci. Eng. (IJIRASE)*, vol. 4, no. 3, pp. 686–689, Sept. 2020. doi: 10.29027/IJIRASE.v4.i3.2020.686-689
[15] N. Pal, D. Sundararaman, P. Arora *et al.*, "How much is my car worth? A methodology for predicting used cars prices using random forest," in *Proc. Future of Information and Communication Conference*, 2018, pp. 413–422.
[16] K. L. Chong, K. D. Kanniah, C. Pohl, and K. P. Tan, "A review of remote sensing applications for oil palm studies," *Geo-Spatial Inf. Sci.*, vol. 20, no. 2, pp. 184–200, Apr. 2017.
[17] L. J. Young, "Agricultural crop forecasting for large geographical areas," *Annual Review of Statistics and its Application*, vol. 6, no. 1, pp. 173–196, Mar. 2019.
[18] T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and Electronics in Agriculture*, vol. 177, Oct. 2020.
[19] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Computers and Electronics in Agriculture*, vol. 151, pp. 61–69, Aug. 2018.
[20] D. Elavarasan, D. R. Vincent, V. Sharma *et al.*, "Forecasting yield by integrating agrarian factors and machine learning models: A survey," *Computers and Electronics in Agriculture*, vol. 155, pp. 257–282, Dec. 2018.
[21] L. S. Woittiez, M. T. V. Wijk, M. Slingerland *et al.*, "Yield gaps in oil palm: A quantitative review of contributing factors," *Eur. J. Agronomy*, vol. 83, pp. 57–77, Feb. 2017.
[22] K. G. Liakos, P. Busato, D. Moshou *et al.*, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, 2674, Aug. 2018.
[23] B. Li, J. Lecourt, and G. Bishop, "Advances in non-destructive early assessment of fruit ripeness towards defining optimal time of harvest and yield prediction—A review," *Plants*, vol. 7, no. 1, Jan. 2018.
[24] R. E. Freeman and J. McVea, "A stakeholder approach to strategic management," *The Blackwell Handbook of Strategic Management*, pp. 183–201, 2005.
[25] J. Gerlings, M. S. Jensen, and A. Shollo, "Explainable AI, but explainable to whom? An exploratory case study of xAI in healthcare," in *Handbook of Artificial Intelligence in Healthcare*, vol. 2, Switzerland: Springer International Publishing, 2022, pp. 169–198.
[26] M. Saarela and V. Podgorelec, "Recent applications of explainable AI (XAI): A systematic literature review," *Appl. Sci.*, vol. 14, no. 19, 2017.
[27] J. Nahar, M. S. Hossain, M. M. Rahman *et al.*, "Advanced predictive analytics for comprehensive risk assessment in financial markets: Strategic applications and sector-wide implications," *Global Mainstream J. Bus. Econ. Dev. Project Manag.*, vol. 3, no. 4, pp. 39–53, 2024.
[28] O. C. Ofodile, A. T. Oyewole, C. E. Ugochukwu *et al.*, "Predictive analytics in climate finance: Assessing risks and opportunities for investors," *GSC Adv. Res. Rev.*, vol. 18, no. 2, pp. 423–433, 2024.
[29] S. Pathak, V. D. Vani, M. Gupta *et al.*, "Making strategic choices in the financial markets: Including deep learning models," in *Proc. Int. Conf. Commun., Comput. Sci. Eng. (IC3SE)*, 2024, pp. 1463–1467.
[30] N. Rane, S. Choudhary, and J. Rane, "Leading-edge artificial intelligence powered financial forecasting for shaping the future of investment strategies," *SSRN*, 2023.