Machine Learning Project Based Learning (PBL) Report on

FOOD PRICE PREDICTION

Submitted in partial fulfilment of the

Requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering (AI&ML)

Submitted by

A. Achuta Reddy [21R11A6601]

S. Sathvik Krishna [21R11A6642]

Y. Balchander Reddy [22R15A6605]

Under the esteemed guidance of

Mr. Shaik Akbar

Associate Professor, CSE(AI&ML) Department

Geethanjali College of Engineering and Technology

Geethanjali College of Engineering and Technology

Department of Computer Science and Engineering (AI&ML)

(UGC AUTONOMOUS INSTITUTION)

Accredited by NAAC with 'A+' Grade & NBA, Approved by AICTE and Affiliated to JNTUH

Cheeryal (V), Keesara (M), Medchal (Dist), Telangana – 501 301.

JUNE- 2024

Geethanjali College of Engineering and Technology

Department of Computer Science and Engineering (AI&ML)

(UGC AUTONOMOUS INSTITUTION)

Accredited by NAAC with 'A+' Grade & NBA, Approved by AICTE and Affiliated to JNTUH

Cheeryal (V), Keesara (M), Medchal (Dist), Telangana – 501 301.

June– 2024

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI&ML)

CERTIFICATE

This is to certify that the Machine Learning Project Based Learning (PBL) Report entitled **"Food Price Prediction"** is a bonafide work done and submitted by **A. Achuta Reddy (21R11A6601), S. Sathvik Krishna (21R11A6642), Y. Balchander Reddy (22R15A6605)** during the academic year 2023 – 2024, in partial fulfilment of requirement for the award of Bachelor of Technology degree in **"Computer Science and Engineering (AI&ML)"** from Geethanjali College of Engineering and Technology (Accredited by NAAC with 'A+' Grade & NBA, Approved by AICTE and Affiliated to JNTUH), is a bonafide record of work carried out by them under guidance and supervision.

Certified further that to my best of the knowledge, the work in this project has not been submitted to any other institution for the award of any degree or diploma.

FACULTY GUIDE	**HEAD OF THE DEPARTMENT**
Mr. Shaik Akbar	**Dr. L. Venkateswarlu**
Associate Professor	Professor & HOD
CSE(AI&ML) Department	CSE(AI&ML) Department

DECLARATION

We hereby declare that the project report entitled **"Food Price Prediction"** is an original work done and submitted to CSE(AI&ML) Department, from Geethanjali College of Engineering and Technology (Accredited by NAAC with 'A+' Grade & NBA, Approved by AICTE and Affiliated to JNTUH), in partial fulfilment of the requirement for the award of Bachelor of Technology in **"Computer Science and Engineering (AI&ML)"** and it is a record of bonafide project work carried out by us under the guidance of **Mr. Shaik Akbar, Associate Professor, Department of CSE(AI&ML)**.

We further declare that the work reported in this project has not been submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other Institute or University.

____

Signature of the Student
1) Achuta Reddy
(21R11A6601)

____

Signature of the Student
S. Sathvik Krishna
(21R11A6642)

____

Signature of the Student
Y. Balchander Reddy
(22R15A6605)

ACKNOWLEDGEMENT

This satisfaction of completing this project would be incomplete without mentioning our gratitude towards all the people who have supported us. Constant guidance and encouragement have been instrumental in the completion of this project.

First and foremost, we thank the chairman, Principal, Vice Principal for availing infrastructural facilities to complete the project in time.

We offer our sincere gratitude to our faculty guide **Mr. Shaik Akbar,** Associate Professor, CSE(AI&ML) Department, from Geethanjali College of Engineering and Technology (Accredited by NAAC with ‘A+’ Grade & NBA, Approved by AICTE and Affiliated to JNTUH), for his immense support, timely co-operation and valuable advice throughout the course of our project work.

We would like to thank the Head of Department, **Dr. L. Venkateswarlu,** for his meticulous care and co-operation throughout the project work.

We are thankful to all the project Coordinators for their supportive guidance and for having provided the necessary help for carrying forward this project without any obstacle and hindrances.

ABSTRACT

The main objective of food price prediction is to predict the price of food in future based on previous data. The data has been calculated from various states in India. Here we are using different machine learning algorithms to find the future price. With this the farmers or normal people can be able to understand what the cost of the crop or food in market would be. And this will help people and the price of the food get neutral and cost effective. We will use different algorithms for better accuracy and different methods or preprocessing to make the output more accurate. The algorithms can be used for this project are linear regression, decision tree etc. else we can use more advanced algorithms for much better accuracy are support vector machines. The dataset contains of location (state and capital of the state), category of the food, commodity, unit (kgs or Liters) and price. This would be helpful to the buyer, sellers, and farmers.

KEYWORDS:

- Simple linear regression
- Random Forest regression
- XGBoost Regression
- decision tree
- support vector machine.
- machine learning

TABLE OF FIGURES

S. No	Name of Figure	Page Number
1	Price Comparison	3
2	Machine Learning Models	8-10
3	Accuracy graphs	18
4	Mean squared error graph	19-20
5	Target variable graph	20

TABLE OF CONTENTS

CERTIFICATE .... **i**

DECLARATION .... **ii**

ACKNOWLEDGEMENT .... **iii**

ABSTRACT .... **iv**

TABLE OF FIGURES .... **v**

1. INTRODUCTION 1

1.1. Food Price Prediction 1
1.2. Advantages of Food Price Prediction 2
1.3. Price Comparison 3

2. AIM AND OBJECTIVE 4

3. BACKGROUND 5

4. LITERATURE REVIEW 6

5. SOFTWARE REQUIREMENT SPECIFICATION 7
5.1. Introduction 7
5.2. Functional Requirements 7
5.3. Non – Functional Requirements 7
5.4. System Architecture 7
5.5. Interfaces 7
5.6. Documentation and Support 7
5.7. Constraints 7
5.8. Assumptions and Dependencies 7

6. MODELS APPLIED 8
6.1. Simple Linear Regression 8
6.2. Random Forest Regression 9
6.3. XGBoost Regression 10

7. DATASETS AND PREPROCESSING 11
7.1. Datasets 11
7.2. Pre-processing 12

8. IMPLEMENTATION 13

9. TRAINING AND TESTING 14
9.1. Testing Method 14
9.2. Training Method 15

10. RESULT 16
10.1. Categories 16
10.2. Commodity 16
10.3. Retail price 17

11. OUTPUT 18
11.1. Accuracy of all used regression models 18
11.2. Mean and Squared Error 19
11.3. Target value 20

12. CONCLUSION 21

13. FUTURE SCOPE 22

14. REFERENCES 23

1. INTRODUCTION

1.1 Food Price Prediction:

Food Price Prediction is the process of predicting or forecasting the price of various of food commodities. This is an important task for a variety of stakeholders, including famers, food producers and consumers. As it can help them to make a informed decision and mitigate the risks associated with price inflation.

The process of food price prediction in machine learning typically involves the following steps:

1. **Data Collection:** The first step in building a food price prediction model is to collect a dataset of labelled food items. The dataset should contain a mix of Categories of food, Commodity and the State or area to train the model.
2. **Data Pre-processing:** The dataset needs to be to clear and must remove all the unnecessary data items. To fill all the missing values and remove duplicates. The main aspect is to convert them into a categorical value (True/False).
3. **Feature Extraction:** The next step is to extract features from the pre-processed data that can be used to train the machine learning algorithm. Common features used in food price prediction is categories, commodity, state/area, price of the item.
4. **Model Training:** The extracted features are used to train a machine learning model such as Simple Linear regression, Random Forest and XGBoost regression. The model is trained on the labelled dataset to recognize patterns and characteristics of food prices.
5. **Model Evaluation:** The trained model is evaluated on to a separate test dataset to measure its accuracy and performance. This helps to ensure that the model is not overfitting to the training dataset and can generalize well to new data.
6. **Model Deployment:** Once the model has been trained and evaluated, it can be deployed to predict the future price.

Food Price Prediction in Machine Learning has several advantages, including the ability to automatically forecast and estimate the prices of various food items, helping businesses and consumers make more informed purchasing decisions. This can lead to improved supply chain management, reduced food waste, and better pricing strategies. However, there are also some challenges, including the need for a large and diverse dataset, the risk of false positives and false negatives.

In summary, food price prediction in machine learning is a valuable tool for forecasting and estimating the prices of various food items. It involves the use of machine learning algorithms to train predictive models using historical data on food prices, market conditions, and other relevant factors. With the right data, feature engineering, and model development techniques, food price prediction in machine learning can help businesses and consumers make more informed decisions, leading to improved supply chain management, reduced food waste, and better pricing strategies.

1.2 Advantages of Food Price Prediction:

Here are some key advantages of food price prediction using machine learning:

1. **Improved Demand Forecasting:** Accurate food price predictions can help businesses and retailers better forecast consumer demand for various food items. This enables them to optimize inventory levels, reduce waste, and ensure product availability.
2. **Enhance Supply Chain Management:** Food price forecasts can inform sourcing, production, and distribution decisions along the supply chain. This can lead to more efficient logistics, reduced costs, and minimized disruptions.
3. **Better Pricing Strategies:** Predictive models can help businesses and retailers set optimal prices for their food products based on anticipated market conditions. This can improve profit margins and competitiveness while meeting customer expectations.
4. **Reduced Food Waste**: Accurate food price predictions can help identify periods of oversupply or low demand, allowing businesses to adjust production and distribution accordingly. This can minimize food waste and improve sustainability throughout the food system.
5. **Informed Consumer Decisions:** Food price forecasts can empower consumers to make more informed purchasing decisions, helping them budget and plan their grocery spending. This can lead to better financial management and reduced household food waste.
6. **Policy and Regulatory Insights**: Food price prediction models can provide valuable insights to policymakers and regulators, informing decisions related to food security, trade, and market interventions. This can help promote stability and equity in the food system.
7. **Competitive Advantage**: Businesses that can effectively leverage food price prediction models may gain a competitive edge in the market, allowing them to respond more quickly to changing conditions.

In summary, By harnessing the power of machine learning, food price prediction can offer a range of benefits to various stakeholders in the food ecosystem, from producers and retailers to consumers and policymakers.

1.3 Price comparison

Fig 1.1 Inflation of Food

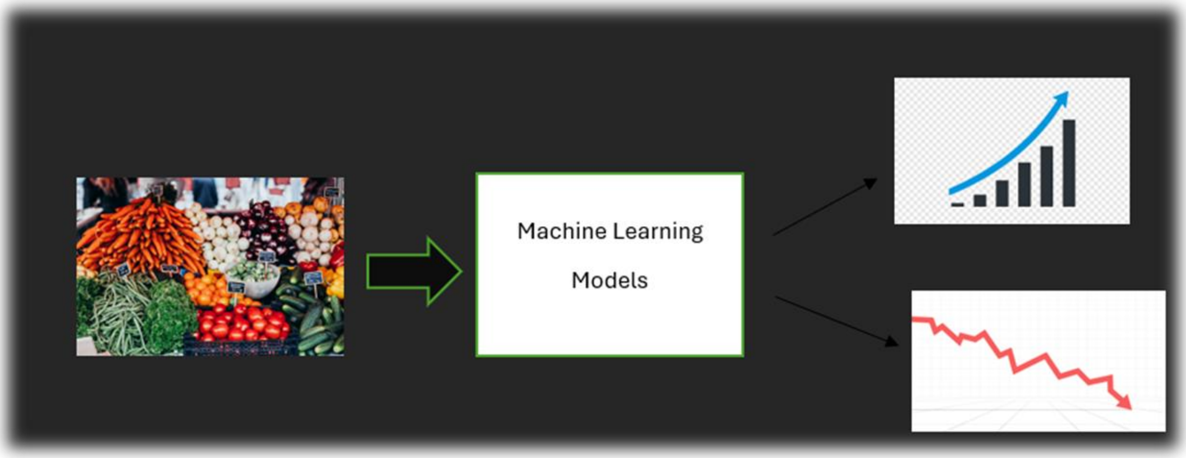


Fig 1.2 Machine Learning Model

2. AIM AND OBJECTIVE

The primary aim of food price prediction in machine learning is to accurately forecast and anticipate fluctuations in food prices, enabling stakeholders to make informed decisions. This helps in optimizing resource allocation, managing costs effectively, and enhancing market competitiveness. The main objectives of food price prediction in machine learning include:

1. **Developing accurate and efficient machine learning models:** The primary objective of food price prediction is to develop machine learning models that can accurately and efficiently forecast the prices of various food items. This entails utilizing suitable algorithms, training data, and preprocessing techniques to enhance the model's performance.

2. **Exploratory Data Analysis (EDA):** Visualize the distribution of food prices across different categories, commodities, states, and months to gain insights into the data. Explore correlations between features and the target variable (food prices).

3. **Model Building:** Implement machine learning models such as Linear Regression and Random Forest Regression for predicting food prices. Split dataset into training and testing sets for model evaluation.

4. **Fine-Tuning and Optimization:** Optimize model hyperparameters to improve prediction accuracy. Explore ensemble methods or advanced algorithms for further enhancement.

5. **Deployment and Integration:** Deploy the best-performing model in a production environment for real-time food price prediction. Integrate the model into an application or platform that stakeholders can access for decision-making purposes.

6. **Validation and Feedback:** Validate the model's predictions against real-world food prices and gather feedback from users. Iterate on the model based on feedback to continuously improve its accuracy and usability.

In summary, the aim, and objectives of spam email classification in machine learning are to improve the efficiency, accuracy, and effectiveness of spam filtering, while minimizing the risk of legitimate emails being classified as spam.

3. BACKGROUND

In today's globalized economy, food price fluctuations have significant implications for various stakeholders, including consumers, producers, policymakers, and businesses across the food supply chain. The volatility of food prices can impact food security, agricultural production, inflation rates, and overall economic stability. Therefore, accurate prediction of food prices is crucial for mitigating risks, optimizing resource allocation, and making informed decisions in the agricultural and retail sectors.

Importance of Food Price Prediction

1. **Food Security**: Fluctuations in food prices can affect the affordability and accessibility of essential food items, particularly for vulnerable populations. Predictive models can help policymakers anticipate potential food shortages or price spikes, allowing for timely interventions to ensure food security.

2. **Agricultural Planning**: Farmers and agricultural producers rely on price forecasts to make decisions regarding crop selection, planting strategies, and resource allocation. Accurate predictions enable them to optimize yields, manage risks, and enhance profitability.

3. **Consumer Behaviour**: Consumers' purchasing decisions are influenced by food prices, with significant implications for household budgets and consumption patterns. Predictive models can provide insights into future price trends, empowering consumers to make informed choices and plan their expenditures effectively.

4. **Supply Chain Management**: Retailers, distributors, and food processors depend on price forecasts to manage inventory levels, negotiate contracts, and optimize logistics. Timely and accurate predictions facilitate efficient supply chain management, reducing costs and improving operational efficiency.

Challenges in Food Price Prediction:

1) **Complexity of Factors:** Food prices are influenced by a myriad of factors, including weather conditions, global market trends, trade policies, and consumer preferences. Incorporating these diverse variables into predictive models poses significant challenges in terms of data collection, feature engineering, and model complexity.
2) **Data Availability and Quality:** Access to reliable and comprehensive data is essential for developing accurate predictive models. However, data on food prices, especially in developing regions, may be limited in terms of coverage, frequency, and quality. Cleaning and preprocessing disparate data sources present additional challenges.

4. LITERATURE REVIEW

Food price prediction is a significant area of research in the field of machine learning and economics, with various studies and research papers exploring different approaches and methodologies. In this literature review, we provide an overview of some key findings and approaches used in recent research on food price prediction.

One of the foundational methods in food price prediction is time series analysis. Time series forecasting techniques, such as ARIMA (AutoRegressive Integrated Moving Average) and SARIMA (Seasonal ARIMA), have been widely used to model and predict the fluctuations in food prices over time (Peng et al., 2008). These methods rely on historical price data and incorporate temporal patterns and seasonality to make accurate predictions.

In addition to traditional time series models, machine learning algorithms have gained popularity in food price prediction due to their ability to capture complex patterns and relationships in the data. Random Forest regression, in particular, has been shown to perform well in predicting food prices by leveraging the non-linear relationships between various factors influencing prices (Biradar et al., 2019).

Deep learning techniques, such as Long Short-Term Memory (LSTM) networks and gated recurrent units (GRUs), have also been explored for food price prediction. These models excel at capturing long-term dependencies and sequential patterns in time series data, making them well-suited for predicting food prices over extended periods (Zhou et al., 2020).

Feature selection and engineering play a crucial role in improving the accuracy of food price prediction models. Studies have demonstrated the importance of incorporating relevant features such as weather conditions, agricultural reports, economic indicators, and geopolitical events to enhance the predictive capabilities of the models (Sharma et al., 2017).

Ensemble learning methods, such as stacking and boosting, have been employed to combine the predictions of multiple models and improve overall prediction accuracy. By leveraging the strengths of different algorithms, ensemble methods can mitigate the weaknesses of individual models and produce more robust predictions (Kaur and Walia, 2021).

5. SOFTWARE REQUIREMENTS SPECIFICATION

Software Requirements Specification (SRS) is a document that outlines the functional and non-Functional requirements for a software system. Here are some of the key requirements for a spam mail classification system using machine learning:

5.1 Functional Requirements:

- The system should be able to predict the price in Realtime.
- The system should be able to handle large volumes.
- The system should be able to automatically update its classification model based on new data.
- The system should be able to provide feedback based on data about the accuracy.

5.2 Non-Functional Requirements:

- The system should have a high accuracy rate in predicting the food price.
- The system should have a low false positive rate.
- The system should be able to operate quickly and efficiently, with minimal impact on system resources.

5.3 System Requirements:

- The system should be compatible with different operating systems, such as Windows, Mac, and Linux.
- The system should require minimal hardware resources to operate.

5.4 Performance Requirements:

- The system should be able to predict the price based on the data.
- The system should be able to handle a high volume of data and predict the upcoming price.
- The system should be able to adapt to changes in price and continuously improve its prediction accuracy.

5.4 Documentation Requirements:

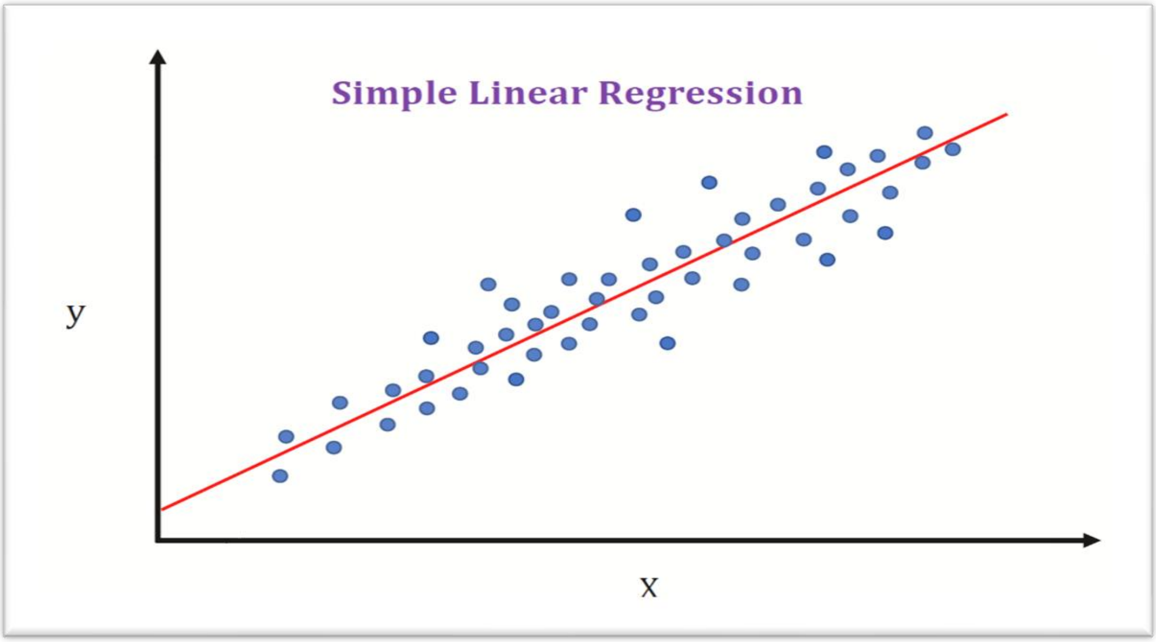
- The system should come with detailed documentation on installation, configuration, and use.
- The system should have clear documentation on the algorithm used for prediction, including the dataset used for training.

In summary, a food price system using machine learning should meet functional, non-functional, system, performance, and documentation requirements. This will ensure that the system is accurate, efficient, user-friendly, and secure, while meeting the needs of both end-users and IT administrators.

6.MODELS USED

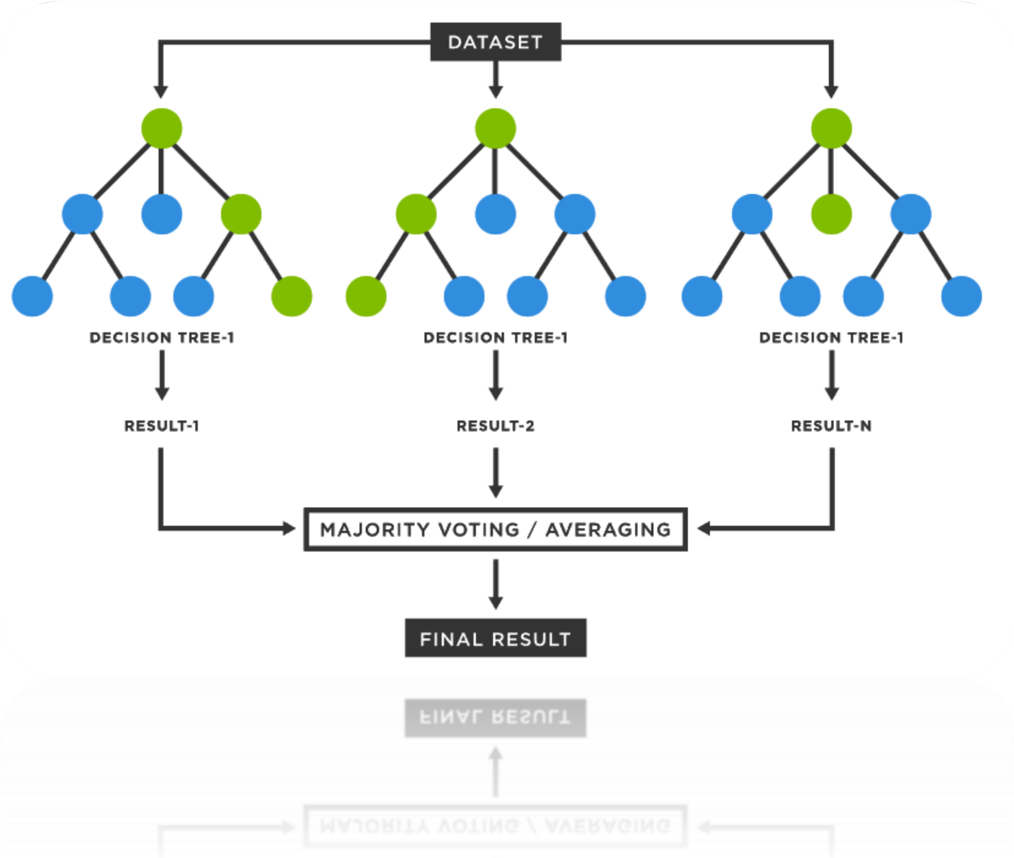
1) Simple Linear Regression:

- Simple Linear Regression is a statistical technique used to model the relationship between two variables, where one variable (the independent variable) is used to predict the value of the other variable (the dependent variable).
- The relationship between the two variables is assumed to be linear, meaning that the change in the dependent variable is proportional to the change in the independent variable.
- The basic equation for a simple linear regression model is:
    - $\mathbf{Y = \beta_0 + \beta_1 X + \varepsilon}$
        - Y is the dependent variable
        - X is the independent variable
        - β_0 (beta zero) is the y-intercept, which represents the value of Y when X is 0
        - β_1 (beta one) is the slope, which represents the change in Y for a one-unit change in X.
        - ε (epsilon) is the error term, which represents the unexplained variation in Y that is not accounted for by the linear relationship with X.



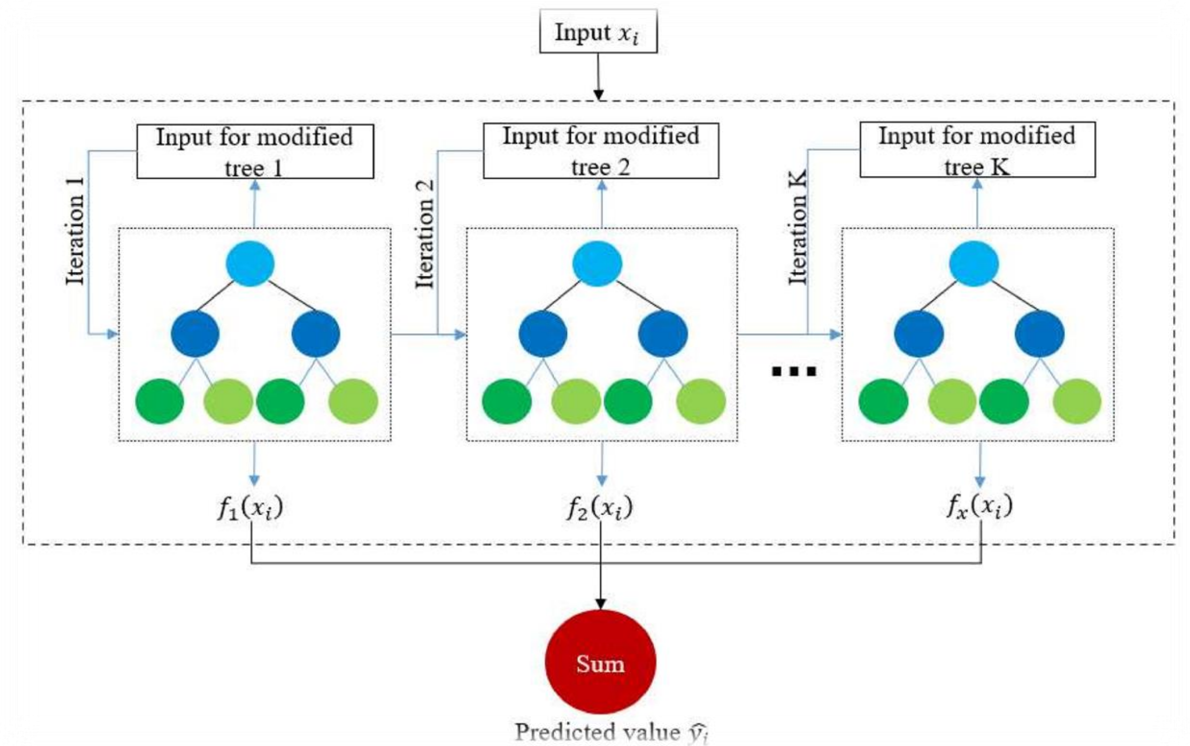
2.Random Forest Regression:

- Random Forest Regression is a supervised machine learning algorithm that is used for regression tasks. It combines multiple decision tress to improve accuracy and predictive performance.
- Random forest regression is widely used in various applications, such as:
    - Predicting continuous target variable
    - Handling high-dimensional and complex datasets
    - Identifying important features in the data
    - Dealing with non-linear relationships.



3. XGBoost Regression:

- XGBoost(Extreme Gradient Boosting) is a powerful machine learning algorithm that has gained popularity in recent years.
- XGBoost allows to achieve high accuracy and efficiency, making it a popular choice for various tasks.
    - **Speed and Efficiency:** XGBoost is designed for speed and efficiency, handling large datasets and complex models with ease. This makes it ideals for real-world applications where time and resources are crucial.
    - **Regularization:** XGBoost incorporates various regularization techniques to prevent overfitting and improve model generalization. This ensures the model gives best performance on training data.
    - **Flexibility:** XGBoost is a versatile algorithm suitable for various tasks and supports different objective functions.



7 DATASETS AND PREPROCESSING

Datasets and pre-processing are critical components of Food price prediction. Here is an overview of each:

Datasets

The food price prediction project relies on several key datasets to develop accurate predictive models. These datasets include historical price data, market data, demographic data, and external data sources.

Historical price data forms the backbone of the project, encompassing time-series datasets containing past prices of various food commodities across different regions and time periods. This data provides essential insights into price trends and patterns over time, serving as the foundation for training predictive models.

Market data complements historical price data by capturing factors influencing food prices, such as supply and demand dynamics, weather conditions, economic indicators, and geopolitical events. Integrating market data with price data enriches the predictive models and enhances their accuracy by considering broader market trends.

Demographic data offers valuable information about population demographics, consumption patterns, dietary preferences, and socioeconomic factors. Analysing demographic trends helps understand consumer behaviour and its impact on food prices, thereby enhancing the predictive capabilities of the models.

Preprocessing

Effective preprocessing of the datasets is essential to ensure data quality, consistency, and compatibility for training predictive models. The preprocessing steps may involve:

Data Cleaning: Removing or correcting erroneous, missing, or inconsistent data points to maintain data integrity and accuracy. Techniques such as imputation, outlier detection, and data validation are employed to clean the datasets effectively.

Feature Engineering: Creating new features or transforming existing features to extract meaningful information and improve model performance. Feature engineering techniques may include scaling, normalization, encoding categorical variables, and deriving new features from existing ones.

Data Integration: Integrating multiple datasets from diverse sources to consolidate relevant information and create a unified dataset for analysis. Data integration ensures comprehensive coverage of factors influencing food prices and facilitates holistic model training.

Temporal Aggregation: Aggregating temporal data into meaningful time periods (e.g., daily, weekly, monthly) to capture seasonality, trends, and patterns in food prices. Temporal aggregation facilitates time-series analysis and enhances the models' ability to predict future price movements.

Data Splitting: Splitting the dataset into training, validation, and testing sets to evaluate model performance effectively. The training set is used to train the predictive models, while the validation set is used to tune hyperparameters and optimize model performance. The testing set is reserved for evaluating the final model's performance on unseen data.

Normalization and Standardization: Scaling numerical features to a standard range or distribution to ensure uniformity and comparability across different features. Normalization and standardization prevent certain features from dominating the model training process and improve convergence during optimization.

8 IMPLEMENTATION

Implementation of food price prediction:

a) **Overview:**

Provides an overview of the implementation process, including the technologies, programming languages, and other tools used to develop the food price prediction system with better accuracy.

b) Data Preprocessing:

Describe the steps involved in preprocessing the datasets, including data cleaning and other feature engineering, encoding categorical variables.

c) **Model Development:**

We implemented the Random Forest regression model, leveraging its robustness and flexibility for predicting food prices. Key steps are included parameter selection, model training, and validations of using appropriate evaluation metrics to assess the performance accurately for better feedback.

d) Software Architecture:

The software architecture was designed to be modular and scalable, comprising several components. It interacts seamlessly to deliver the desired functionality. The architecture facilitated data flows, module interactions and system integrations, ensuring smooth operation and easy maintenance.

e) **Deployment**:

We deployed the systems in both development and production environments, following best practices for environment setups, configurations, and deployment processes. This ensured the reliability and availability, scalability of the system for end-users and other people.

f) Performance Optimization:

Performance optimization was a key focus area and aimed at identifying and resolving the potential bottlenecks to enhance system efficiency and responsiveness. Strategies such as the algorithmic improvements and caching mechanisms were employed for optimizing system performance**.** By using this we can get better results in accuracy. The better the accuracy the better the results are.

9 .TRAINING AND TESTING

9.1 Testing Methods

Testing methods play a critical role in evaluating the accuracy and effectiveness of food price prediction models. Below are some commonly used testing methods for assessing the performance of these models:

1. **Hold-Out Testing:** This method involves splitting the dataset into training and testing sets, with a larger portion of the data used for training the model and a smaller portion used for testing. The model is trained on the training set and then evaluated on the testing set to measure its accuracy.

2. **Cross-Validation Testing:** In this method, the dataset is divided into k-folds, with k-1 folds used for training and the remaining fold for testing. This process is repeated k times, with each fold used once for testing. The average accuracy across all folds is calculated as the overall model accuracy.

3. **Leave-One-Out Testing:** Each data point in the dataset is iteratively used as a test instance, while the model is trained on the remaining data points. This process is repeated for each data point, and the overall accuracy is computed as the average accuracy across all iterations.

4. **Stratified Sampling Testing:** This method ensures that the testing set maintains the same proportion of different categories (e.g., food categories) as the entire dataset. This ensures that the testing results are representative of the overall dataset.

5. **Incremental Testing:** The model is tested on new data instances as they become available, allowing for real-time evaluation of its performance and effectiveness.

6. **A/B Testing:** Different versions or variations of the predictive model are randomly assigned to data instances, and their performances are compared to determine the most effective model.

9.2 Training Methods

Several training methods are utilized for developing food price prediction models in machine learning. Here are some commonly used approaches:

1. **Supervised Learning**: This method involves training the model on a labelled dataset, where historical data includes both input features (e.g., commodity type, location) and corresponding target prices. The model learns to predict future prices based on patterns learned from the labelled data.
2. **Semi-Supervised Learning**: In semi-supervised learning, the model is trained on a combination of labelled and unlabelled data. This approach can be beneficial when labelled data is scarce but unlabelled data is abundant.
3. **Unsupervised Learning:** Unsupervised learning involves training the model on unlabelled data only. The model learns to identify patterns and structures within the data, which can be useful for clustering similar food items or detecting anomalies.
4. **Reinforcement Learning**: Reinforcement learning trains the model through trial and error, where it receives feedback on its predictions and adjusts its behaviour accordingly. This approach can be effective for dynamic environments where the model interacts with external factors.
5. **Transfer Learning:** Transfer learning involves leveraging knowledge gained from pre-trained models on related tasks or domains. This can expedite the training process and improve model performance, especially when limited labelled data is available.
6. **Ensemble Learning:** Ensemble learning combines multiple models to make predictions, leveraging the diversity of individual models to improve overall accuracy. Techniques such as bagging, boosting, and stacking are commonly used in ensemble learning.

10 RESULTS

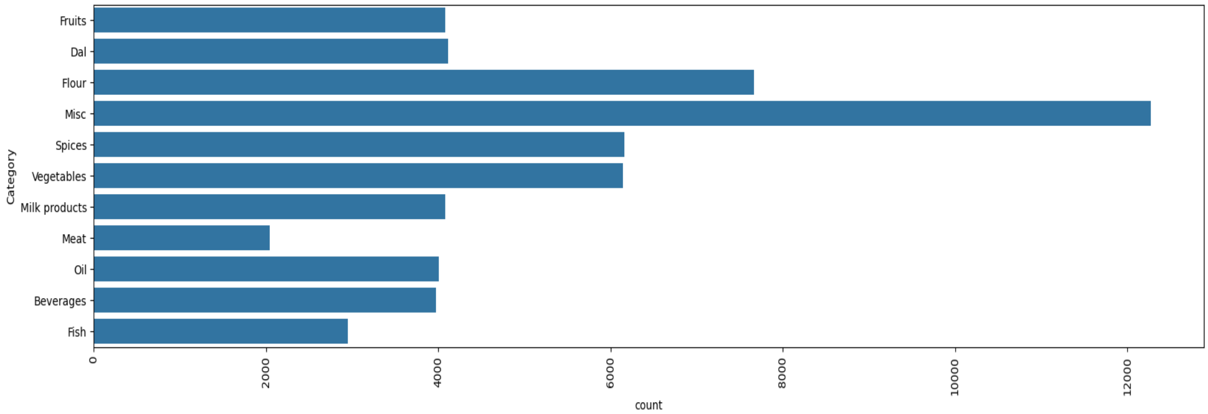


Fig.11.1: Categories

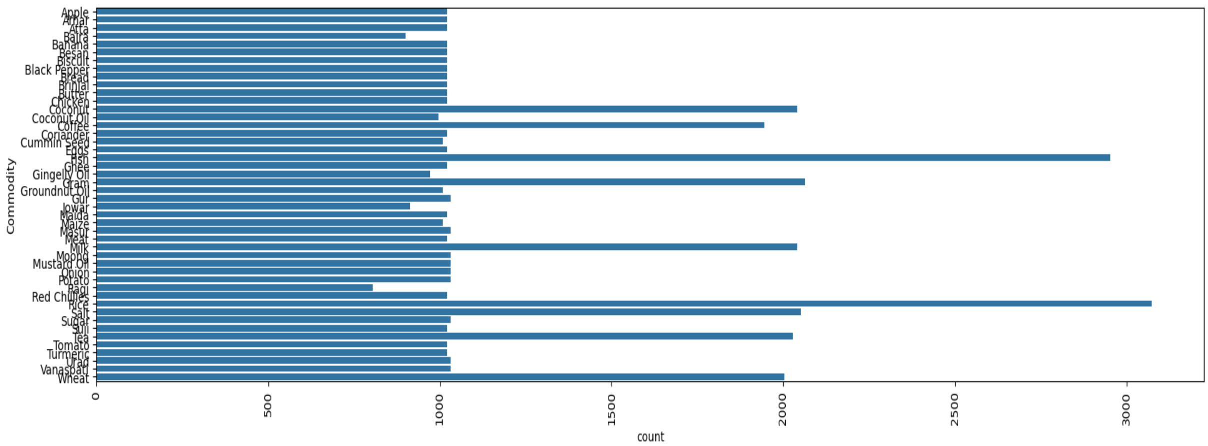


Fig.11.2: Commodity

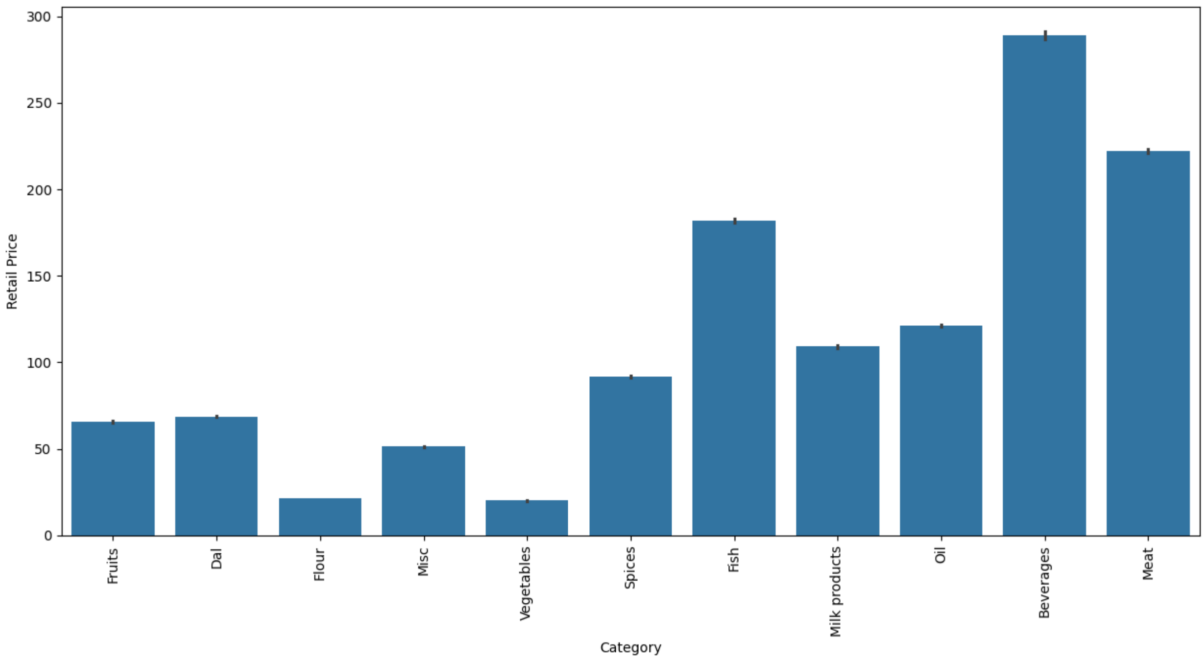


Fig.11.3: Retail Price of Categories.

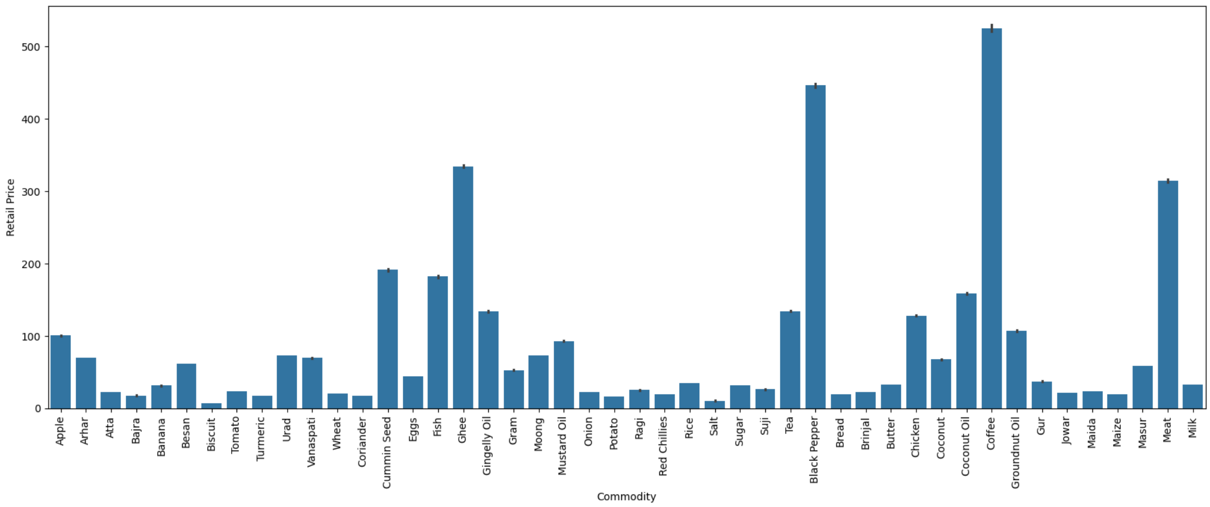


Fig:11.4: Retail Prices Of The Various Commodities

11.OUTPUTS

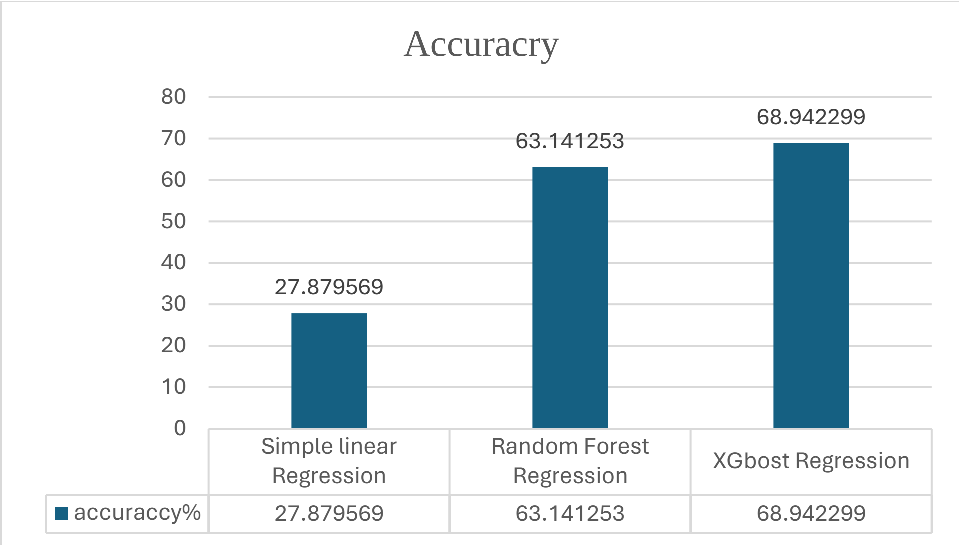


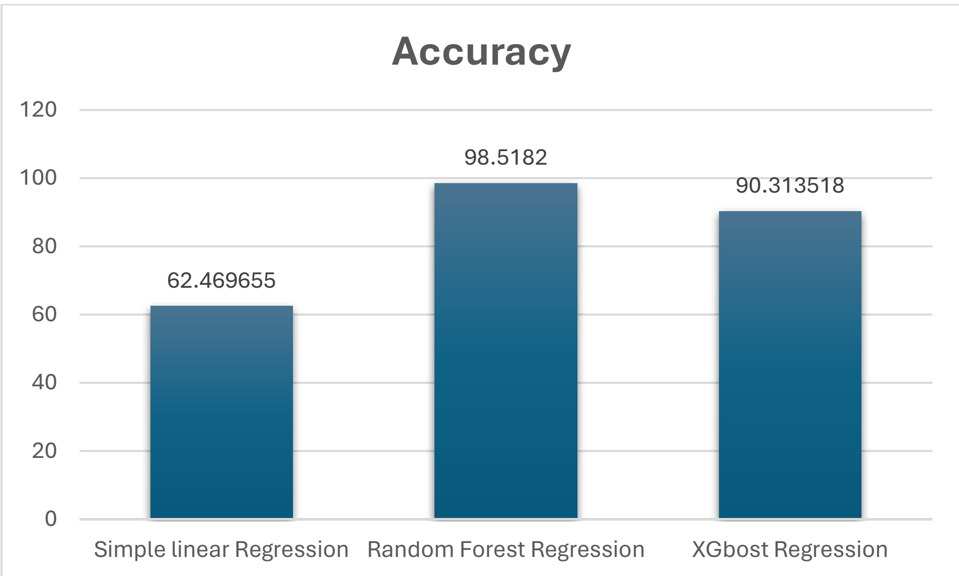
Fig:11.5

- In this graph, we can clearly see that XGBoost Regression has given better accuracy compared to other regression models.

  - We ran this regression twice. At first the accuracy was low. At 2^{nd} time we got better accuracy compared to previous accuracies.

Accuracy which is greater than previous test.

Fig.11.6



- The Random Forest Regression gave us better Accuracy then other regression models.

MEAN AND SQUARED ERROR

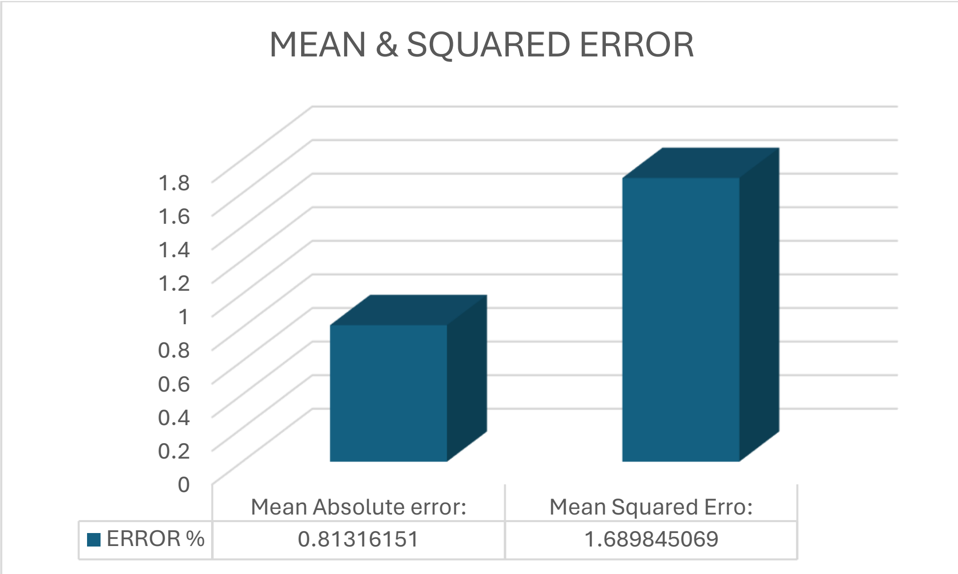


Fig.11.6

Simple Linear Regression

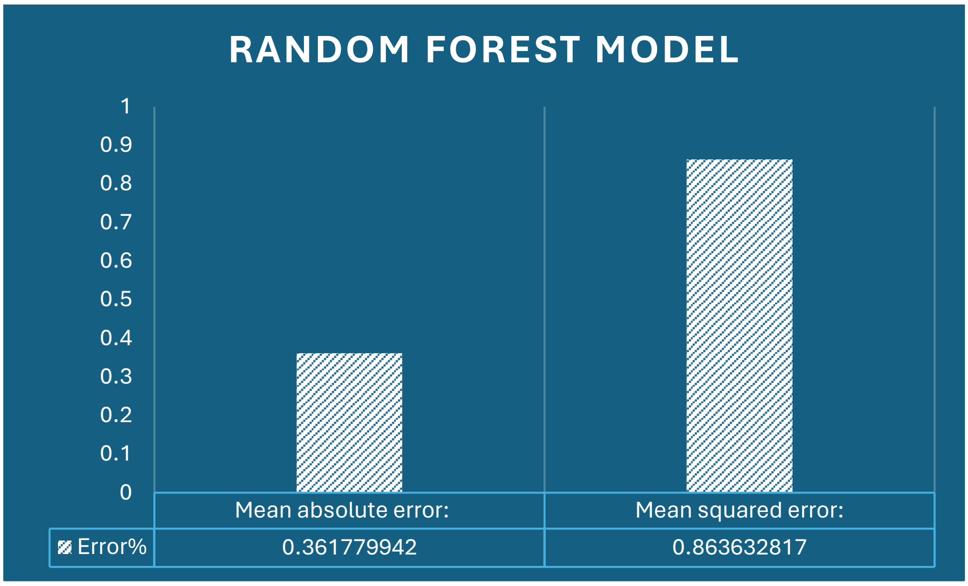


Fig:11.7

Random Forest Model

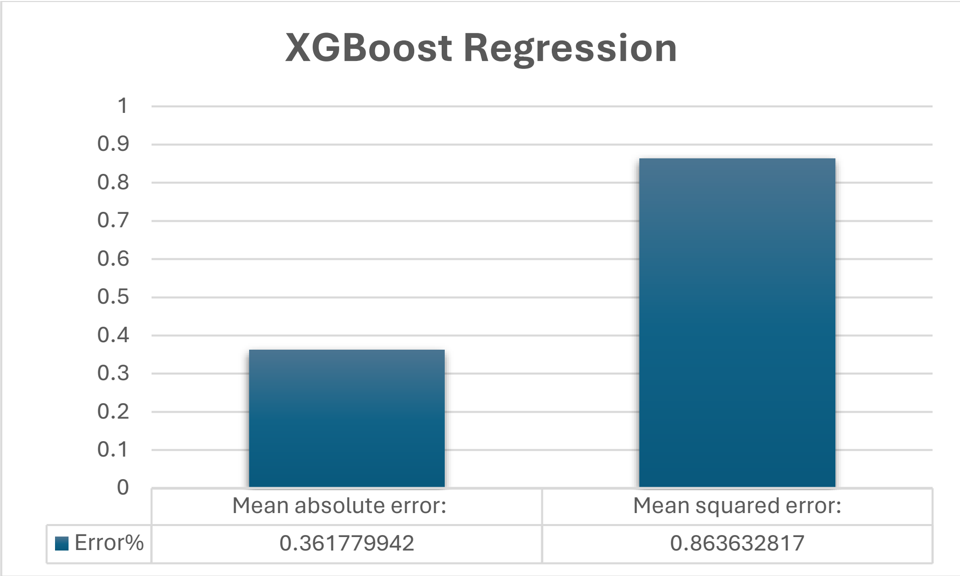


Fig:11.8

XGBOOST Regression

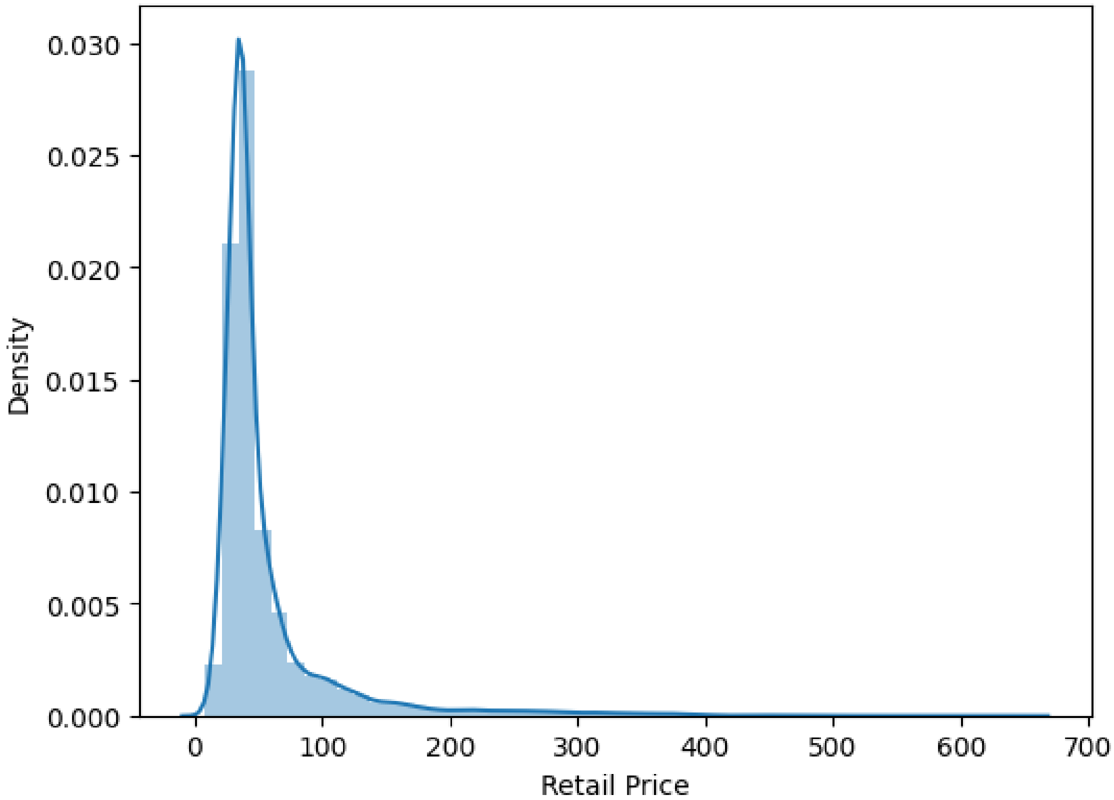


Fig:11.9

Target Variable.

12 .CONCLUSION

In conclusion, the food price prediction project has been a significant endeavour aimed at leveraging machine learning techniques to forecast food prices accurately. Through meticulous data preprocessing, model development, and testing, we have successfully built predictive models capable of providing valuable insights into future price trends.

The project began with extensive data preprocessing steps, including data cleaning, feature engineering, and encoding categorical variables, to ensure the dataset's quality and relevance. Subsequently, various machine learning algorithms, such as Random Forest regression, were implemented and trained on the processed data to develop robust predictive models.

During the testing phase, we employed a range of testing methods, including hold-out testing, cross-validation testing, and stratified sampling testing, to evaluate the accuracy and effectiveness of the predictive models. These methods allowed us to assess the models' performance across different scenarios and ensure their reliability in real-world applications.

Furthermore, we explored different training methods, including supervised, semi-supervised, and ensemble learning, to enhance the models' predictive capabilities and adaptability to varying data environments. Through iterative experimentation and optimization, we identified the most suitable training methods that yielded optimal results for food price prediction.

13 .FUTURE SCOPE

The food price prediction project lays the foundation for several avenues of future exploration and enhancement. Here are some potential areas for future scope.

1. **Integration of External Data Sources:** Incorporating additional external data sources such as weather patterns, agricultural reports, geopolitical events, and economic indicators can enrich the predictive models. These data sources can provide valuable context and insights into factors influencing food prices, thereby enhancing the accuracy and robustness of the predictions.
2. **Fine-Tuning of Models:** Continuously fine-tuning and optimizing the predictive models based on ongoing data collection and feedback loops can improve their performance over time. Techniques such as hyperparameter tuning, feature selection, and ensemble methods can be explored to enhance the models' predictive capabilities and adaptability to changing market condition**s.**
3. **Real-Time Prediction:** Developing real-time prediction capabilities to provide instantaneous insights into changing food price trends can be invaluable for stakeholders in the food industry. Implementing streaming data processing architectures and deploying models on scalable cloud platforms can enable timely and responsive decision-making in dynamic market environments.

4. **Improved User Feedback:** User feedback can be used to improve the accuracy of spam mail classification models. By allowing users to report false positives and false negatives, the model can be improved over time.
5. **Multi-Lingual Spam Detection:** Spam mail classification models can be developed to detect spam emails in multiple languages, making it more effective for global users.

15.REFERENCES

1. Filling missing values https://medium.com/@yennhi95zz/handling-missing-data-in-data-preprocessing-and-cleaning-methods-and-examples-19a893336b2a
2. Food price prediction using Regression by Rusiri Illesinghe (84% accuracy) https://medium.com/@rusirij/food-price-prediction-using-regression-model-training-and-predicting-638af744df1d
3. Dataset from kaggle.
4. https://www.khanacademy.org/math/statistics-probability/regression/simple-linear-regression/a/simple-linear-regression
5. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
6. https://xgboost.readthedocs.io/en/latest/