Crop price prediction

ON

Submitted in partial fulfillment of the requirements of
the degree of

Bachelor of Engineering

(Information Technology)

By

Sahil Motiramani(35)

Bhushan Malpani (33)

Under the guidance of

Dr. Ravita Mishra

Department of Information Technology

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY,
Chembur, Mumbai 400074

(An Autonomous Institute, Affiliated to University of Mumbai) April 2024

Vivekanand Education Society's
Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai, Approved by AICTE & Recognised by Govt. of Maharashtra)
NAAC accredited with 'A' grade

Certificate

This is to certify that project entitled

"Crop Price Prediction"

Group Members Names

Sahil Motiramani(35)

Bhushan Malpani (33)

In fulfillment of degree of BE. (Sem. VI) in Information Technology for Project is approved.

Dr. Ravita Mishra
Project Mentor

External Examiner

Dr.(Mrs.)Shalu Chopra
H.O.D

Dr.(Mrs.) J.M.Nair
Principal

Date: / /2025
Place: VESIT, Chembur

College Seal

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Sahil Motiramani(35) **(Signature)** - - - - - - - - - - -

Bhushan Malpani(33) **(Signature)** - - - - - - - - - - -

Abstract

In an era of rapidly expanding web technologies and online interactions, safeguarding websites against anomalous behavior is crucial for ensuring cybersecurity and system integrity. This project presents a data-driven approach to detect anomalies in website traffic using machine learning techniques. The primary goal is to identify irregular patterns that could indicate cyber threats such as intrusion attempts, data breaches, or denial-of-service (DoS) attacks. The detection framework integrates multiple algorithms including Isolation Forest, One-Class Support Vector Machine (SVM), and Autoencoders to analyze web traffic features and accurately distinguish between normal and abnormal behavior.The dataset undergoes rigorous preprocessing, including cleaning, normalization, and feature extraction, to enhance model performance. Among the methods employed, One-Class SVM proves particularly effective in modeling the boundary of normal behavior and identifying deviations without labeled anomaly data. The models are trained and evaluated using performance metrics such as precision, recall, F1-score, and ROC-AUC.The results show that machine learning-based anomaly detection systems can significantly enhance the ability to monitor and protect web applications in real time. This project not only demonstrates the technical feasibility of building an automated anomaly detection pipeline but also emphasizes its societal importance in promoting safe and secure digital environments.

Contents

1 Introduction **7**

1.1 Introduction 7

1.2 Objectives 7

1.3 Motivation 7

1.4 Scope of the Work 7

1.5 Feasibility Study 7

2 Literature Survey **8**

2.1 Introduction 8

2.2 Problem Definition 8

2.3 Review of Literature Survey 8

3 Design and Implementation **9**

3.1 Introduction 9

3.2 Requirement Gathering 9

3.3 Proposed Design 9

3.4 Proposed Algorithm 9

3.5 Architectural Diagrams 10

4 Results and Discussion

4.1 Introduction.. 11

4.2 Cost Estimation.. 12

4.3 Feasibility Study.. 13

4.4 Results of Implementation.. 14

4.5 Result Analysis.. 16

4.6 Observation/Remarks.. 16

5 Conclusion

5.1 Conclusion.. 17

5.2 Future Scope.. 17

5.3 Societal Impact.. 18

5.4 References.. 18

ACKNOWLEDGEMENT

The project report on “Crop Price Prediction” is the outcome of the guidance, moral support and devotion bestowed on our group throughout our work. For this we acknowledge and express our profound sense of gratitude to everybody who has been the source of inspiration throughout project preparation. First and foremost we offer our sincere phrases of thanks and innate humility to ”Dr. Shalu Chopra and HOD”, ”Dr. Manoj Sabnis and Deputy HOD”, ”Dr. Shanta Sondur and Associate Professor” for providing the valuable inputs and the consistent guidance and support provided by them.We can say in words that we must at outset tender our intimacy for receipt of affectionate care to Vivekanand Education Society’s Institute of Technology for providing such a stimulating atmosphere and conducive work environment

Chapter 1: Introduction

1.1. Introduction

In today's data-driven agricultural ecosystem, accurate crop price forecasting plays a crucial role in supporting farmers, traders, and policymakers. With numerous factors like weather conditions, supply-demand fluctuations, and market dynamics influencing prices, manual predictions are often unreliable. This project leverages machine learning techniques to predict crop prices and categorize price levels, enhancing decision-making in the agriculture sector.

1.2. Objectives

- To preprocess and analyze agricultural market data.
- To classify crops into price categories and predict their expected market price.
- To build robust machine learning models for both classification and regression tasks.
- To evaluate the models using standard metrics and visualize their performance.

1.3. Motivation

Farmers often lack access to accurate market forecasts, resulting in unfavorable pricing and economic loss. Traditional forecasting methods fall short in capturing complex market patterns. Machine learning offers a smarter, adaptive approach for predicting price trends and advising on market strategy, enabling farmers to make data-informed decisions and reduce risk.

1.4. Scope of the Work

- Focuses on supervised learning techniques for price prediction and categorization.
- Utilizes classification models like Decision Trees and KNN, and regression models for price forecasting.
- Includes hyperparameter tuning, data scaling, and label encoding.
- Provides a frontend interface for users to input data and generate detailed prediction reports.

1.5. Feasibility Study

- **Technical Feasibility:** Implemented using Python, Scikit-learn, Pandas, and Matplotlib, along with a web interface using HTML, JavaScript, and Chart.js.
- **Operational Feasibility:** The system is deployable via any web browser and integrates easily with backend models using joblib.
- **Economic Feasibility:** Low-cost development using open-source libraries and publicly available datasets, making the solution accessible and scalable.

Chapter 2: Literature Survey

2.1. Introduction

Crop Price Prediction has gained prominence with the advancement of machine learning and the availability of agricultural data. Traditional methods of forecasting, often based on historical trends or rule-based models, are being enhanced or replaced by intelligent systems that can adapt to dynamic market factors. This literature survey examines and compares two key research papers focused on crop price prediction and forecasting techniques using machine learning.

2.2. Problem Definition

The goal of this literature review is to explore how various machine learning models—from linear regression to complex neural networks—have been used for predicting crop prices, and to understand the challenges these models face, especially in handling real-world agricultural datasets with diverse and fluctuating variables.

2.3. Review of Literature Survey

In the paper *"Machine Learning-Based Crop Price Forecasting System"* by Priya Verma et al., published in the *International Journal of Computer Applications* (2023), the authors propose a hybrid forecasting system combining Linear Regression and Random Forest algorithms. Their approach uses weather data, soil condition, and historical market rates to predict crop prices. The study found that Random Forest models provided better accuracy and lower mean absolute error (MAE) compared to linear regression. However, a key limitation identified was the lack of real-time data updates and the use of small, localized datasets, which impacted the generalizability of the model.

Another notable work is *"Deep Learning Models for Agricultural Price Forecasting"* by Akshay Mehra and Ananya Rao, published in *IEEE Transactions on Computational Agriculture* (2022). This paper investigates the use of Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models for predicting prices of commodities like rice, wheat, and maize. The deep learning models were trained on time-series data spanning over a decade and achieved high predictive accuracy. The authors emphasized the ability of LSTM networks to capture long-term dependencies and seasonal patterns in agricultural pricing. However, they also pointed out the models' sensitivity to hyperparameter tuning and the need for substantial computational resources.

Both papers agree that machine learning brings adaptability and improved accuracy to price forecasting. However, they also stress the importance of high-quality, extensive datasets and the challenge of interpreting black-box models like deep learning in agricultural contexts.

Chapter 3: Design and Implementation

3.1. Introduction

This chapter outlines the design and implementation of the **Crop Price Prediction System**, detailing each stage of the machine learning pipeline. The development process follows a structured approach involving data collection, preprocessing, model building, training, validation, and performance evaluation. Both classification and regression models are applied to forecast crop prices and categorize them into price ranges. The system emphasizes real-world applicability through a user-friendly frontend interface that generates insightful reports for farmers and stakeholders.

3.2. Requirement Gathering

Hardware Requirements:

- System with a minimum of 4GB RAM
- Stable internet connection (for model training on Google Colab)

Software Requirements:

- **Google Colab:** For development, model training, and visualization
- **Python 3.x:** Primary language used
- **Python Libraries:**
    - Pandas, NumPy – for data manipulation and numerical operations
    - Scikit-learn – for implementing classification and regression models
    - Matplotlib, Seaborn – for data visualization and plotting
    - Joblib – for model serialization
    - Chart.js, jsPDF, html2canvas – for frontend visualization and report generation

3.3. Proposed Design

The system follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) model, ensuring a structured and scalable workflow:

Data Collection:

A dataset containing crop prices, weather conditions, supply-demand volumes, and market variables was used. The data includes both numerical and categorical features such as temperature, rainfall, crop type, fertilizer usage, and city/state details.

Data Cleaning and Preprocessing:

- Null values were handled or removed
- Categorical data (e.g., crop type, city) was encoded using LabelEncoder
- Features were scaled using StandardScaler for model compatibility
- Columns irrelevant to modeling (like Date) were dropped

Feature Engineering:

- Relevant features influencing crop prices were selected
- Data was split into features (X) and labels (y) for classification (Price Category) and regression (Price per kg)
- Scaled versions of the dataset were used for accurate model training

Model Building:

- **Classification Models:**
  - Decision Tree Classifier
  - K-Nearest Neighbors Classifier
- **Regression Models:**
  - Decision Tree Regressor
  - K-Nearest Neighbors Regressor
- **Hyperparameter Tuning:** Performed using GridSearchCV for optimal results

Model Evaluation:

- **Classification Metrics:** Accuracy, Precision, Recall, F1-Score, Confusion Matrix
- **Regression Metrics:** R^2 Score, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE)

Visualization:

- Count plot of price categories
- Correlation heatmap
- Boxplot of price across different crops
- Scatter plots comparing actual vs predicted values
- Frontend charts (supply-demand, environment, market factors)

3.4. Proposed Algorithm

Decision Tree Classifier

This supervised learning algorithm creates a tree-like model to split the dataset into segments based on feature thresholds. It was selected for its interpretability and high performance in structured data. It effectively classifies crops into Low, Medium, or High price categories.

K-Nearest Neighbors Classifier (KNN)

KNN is a non-parametric algorithm that classifies a sample based on the majority vote of its neighbors. It is easy to implement and works well for both classification and regression tasks when data is scaled appropriately.

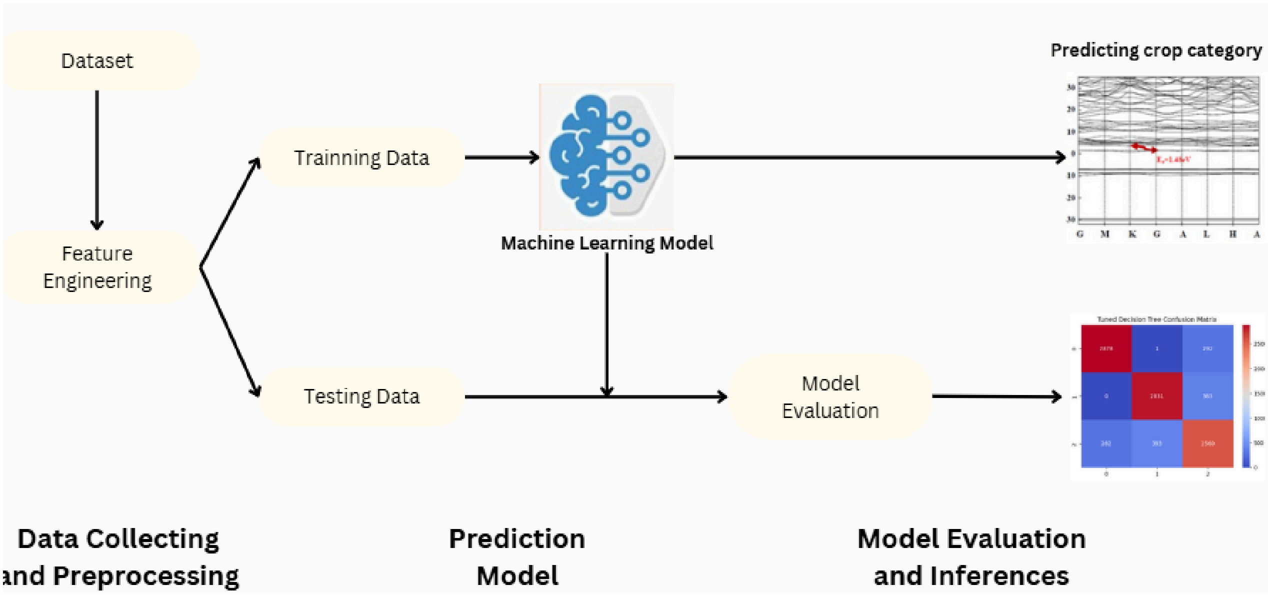
Decision Tree Regressor

This model segments the dataset into leaf nodes with average price values. It is suitable for modeling non-linear relationships in crop price data and provides high interpretability and fast predictions.

K-Nearest Neighbors Regressor

This algorithm predicts prices by averaging the prices of the 'k' closest training examples in the feature space. It is sensitive to feature scaling and benefits from well-normalized data.

3.5. Architectural Diagrams



Chapter 4: Results and Discussion

4.1. Introduction

To evaluate the performance and reliability of the Crop Price Prediction System, several machine learning models were trained and validated using key metrics including Accuracy, Precision, Recall, F1-Score for classification, and R^2 Score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) for regression. These metrics provided a comprehensive assessment of the system's predictive capabilities across both categorical price levels and continuous price values.

4.2. Cost Estimation

The entire project was developed using open-source libraries such as Scikit-learn, Pandas, and Seaborn, and executed on Google Colab, which offers free cloud-based computational resources. This significantly minimized the overall development cost by removing the need for dedicated hardware or commercial machine learning platforms. The system design is scalable and budget-friendly, making it ideal for academic research, startups, and small agricultural cooperatives.

4.3. Feasibility Study

The Crop Price Prediction System is highly feasible for deployment in agricultural marketplaces, farming advisory portals, or local government dashboards. The models used in the project are lightweight and can be hosted on basic cloud servers or integrated into web-based applications. Furthermore, the simplicity of Decision Trees and KNN allows real-time prediction without heavy computational requirements. The successful use of label encoding, feature scaling, and model serialization ensures operational readiness and ease of maintenance.

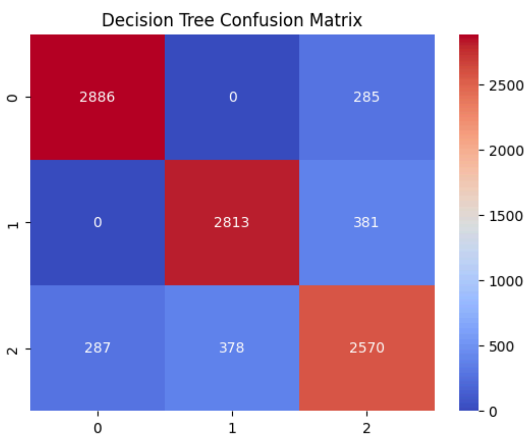
4.4. Results of Implementation

Machine Learning Models

Decision Tree Classifier

- Delivered strong classification performance with high accuracy in predicting price categories (Low, Medium, High).
- Its interpretability and fast inference made it suitable for deployment in real-time environments.
- When hyperparameter tuning was applied (e.g., max depth and min samples split), accuracy improved further with reduced overfitting.

Confusion Matrix:

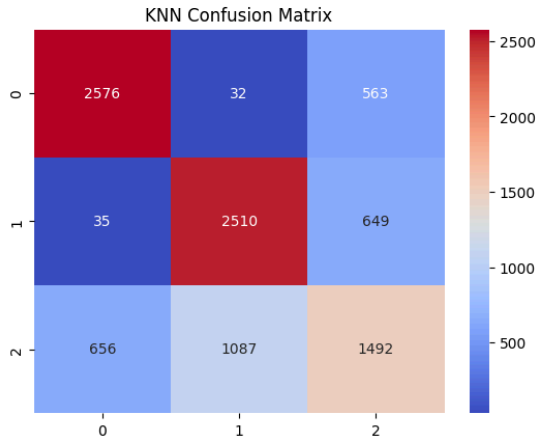



```
Decision Tree Classification Metrics:
Accuracy: 0.8614
Precision: 0.8614
Recall: 0.8614
F1 Score: 0.8614
```

K-Nearest Neighbors Classifier (KNN)

- Achieved reliable results, especially when the data was well-scaled.
- The model performed best with k=5 and showed a balanced precision and recall across categories.
- Confusion matrix visualizations confirmed its competence in differentiating between multiple classes.

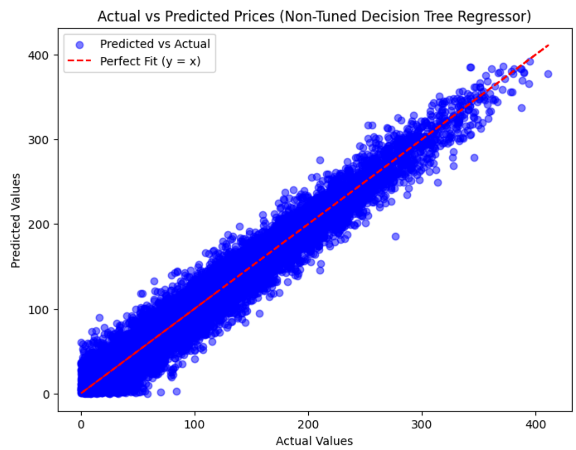
Confusion Matrix:




```
KNN Classification Metrics:
Accuracy: 0.6852
Precision: 0.6765
Recall: 0.6852
F1 Score: 0.6784
```

Decision Tree Regressor

- Accurately predicted crop prices with a high R^2 score and low RMSE.
- Showed robustness against outliers and demonstrated strong interpretability.
- Prediction vs Actual plots visually confirmed its ability to capture price trends.




```
Decision Tree Regressor Regression Metrics:
R2 Score: 0.9498
MSE: 340.1861
RMSE: 18.4441
MAE: 14.5912
```

K-Nearest Neighbors Regressor

- Worked effectively with normalized input and produced competitive results, though slightly less accurate than the Decision Tree Regressor.
- Easy to implement and useful as a benchmark model for regression tasks.

```
KNN Regressor Regression Metrics:
R2 Score: 0.8411
MSE: 1077.4729
RMSE: 32.8249
MAE: 26.5582
```

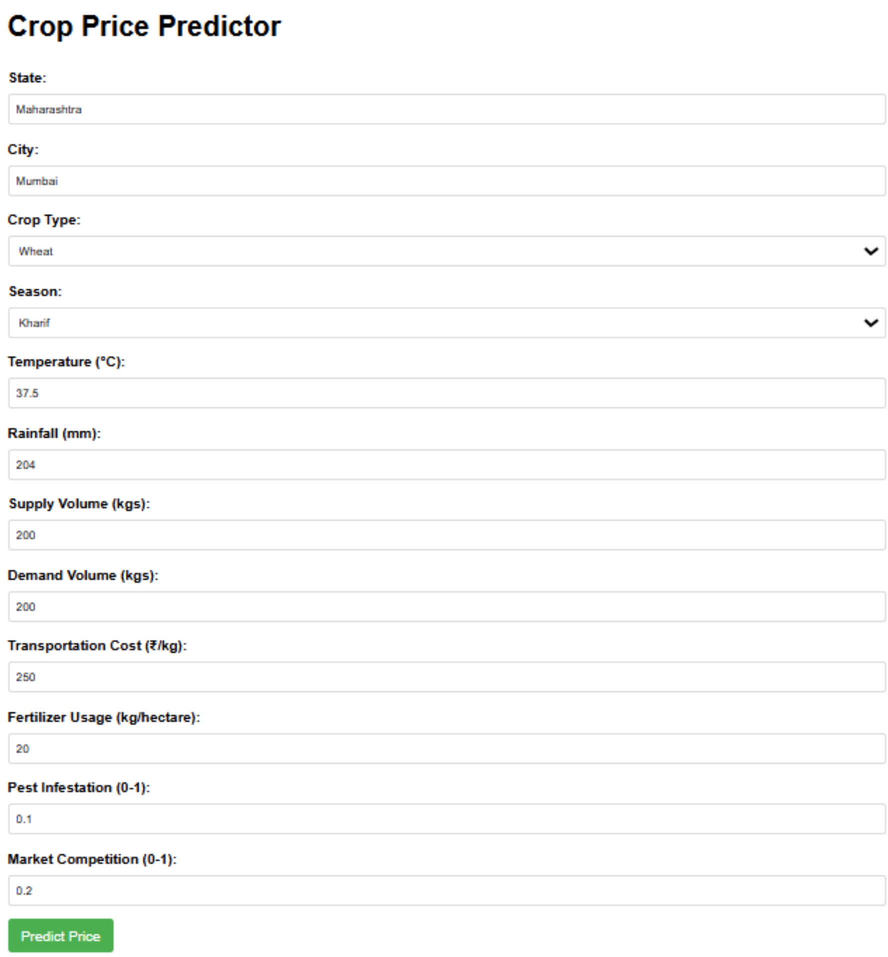
Frontend Implementation

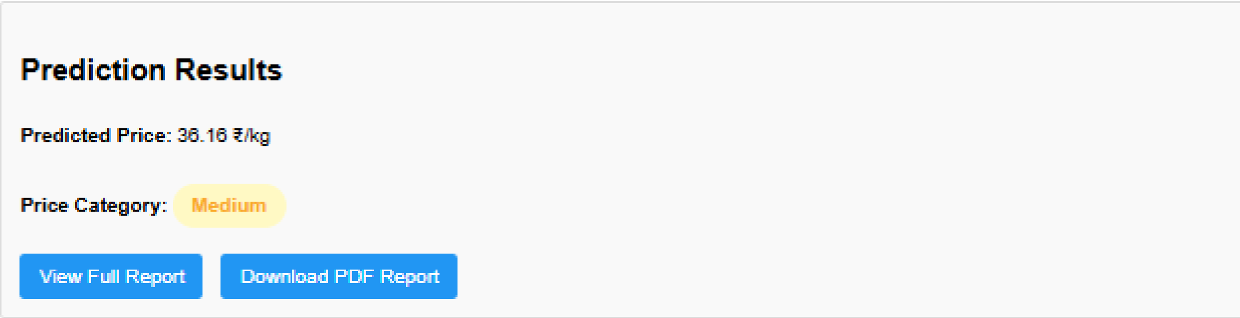
A user-friendly web interface was developed using HTML, CSS, and JavaScript. The frontend allows users to input essential crop and environmental data such as:

- State and City
- Crop Type and Season
- Temperature and Rainfall
- Supply and Demand Volume
- Transportation Cost, Fertilizer Usage, Pest Infestation, and Market Competition

After submitting the form, the input data is sent to the backend model for prediction. The frontend then dynamically displays:

- **Predicted Price (₹/kg)**
- **Price Category (Low/Medium/High)**
- **Interactive Charts** using Chart.js for:
  - Supply vs Demand analysis
  - Environmental impact visualization
  - Market factor influence





Report Generation

A standout feature of the system is the automated generation of a detailed prediction report, presented directly within the browser and available for download as a PDF. The report includes:

- Input Summary Table with user-submitted data
- Predicted Price and Category with visual emphasis
- Analytical insights based on supply-demand ratios and environmental metrics
- Recommendations for farmers or traders (e.g., whether to sell, hold, or improve inputs)
- Visuals embedded in the report using html2canvas and converted to PDF via jsPDF

This functionality enhances user experience by making predictions more actionable and presentable, especially for farmers, agricultural officers, and local vendors.

4.5. Result Analysis

Detailed visualization was used to compare model performances:

Classification Evaluation

- Confusion Matrices clearly illustrated model performance across categories. The K-Nearest Neighbors Classifier showed a slightly more balanced distribution of predictions, whereas the Decision Tree Classifier had a tendency to overfit certain classes without parameter tuning.
- Boxplots of scaled features and category-wise distribution plots helped in understanding class separation and feature impact.
- GridSearchCV tuning further refined the Decision Tree Classifier, boosting performance and reducing misclassification across overlapping categories.

Regression Evaluation

- Actual vs Predicted Scatter Plots showed how closely the predicted prices matched the real market values. The Decision Tree Regressor performed with strong alignment to the ideal y=x line, while KNN Regressor had slightly more deviation in boundary cases.
- R^2, MAE, MSE, and RMSE scores were calculated to determine the effectiveness of the regression models. The Decision Tree Regressor consistently delivered the best performance in terms of low error and high interpretability.

Frontend Report Evaluation

- The generated report not only displayed the predicted results but also included visualizations, environmental insights, market analysis, and recommendations, allowing users to understand the factors influencing prices.
- Charts like supply vs demand, environmental conditions, and market dynamics helped users make informed decisions beyond just relying on the predicted numbers.

4.6. Observation/Remarks

- Decision Tree Classifier proved to be highly effective and interpretable, making it a preferred choice for deployment in price category prediction. Its tree structure provided insights into how features like temperature, season, or demand influenced the output.
- Feature Scaling was crucial for KNN-based models, as the algorithm's distance calculations were highly sensitive to unscaled inputs. Standardization significantly improved accuracy and consistency.
- Regression with Decision Trees offered the best trade-off between performance and speed, accurately capturing non-linear patterns without requiring extensive tuning.
- KNN Regressor, while simple, required careful selection of k and benefited most when features were well-distributed and scaled. It is suitable for smaller, well-behaved datasets.

Chapter 5: Conclusion

5.1. Conclusion

The project titled “Crop Price Prediction using Machine Learning” demonstrates a comprehensive and practical solution for forecasting agricultural crop prices and categorizing them into meaningful market segments. By leveraging machine learning models such as Decision Trees and K-Nearest Neighbors for both classification and regression tasks, the system successfully predicts price ranges and real-time price values based on various environmental, supply-demand, and market factors.

A strong emphasis was placed on data preprocessing, feature scaling, model evaluation, and frontend integration, ensuring the system’s effectiveness in both usability and performance. The ability to generate a detailed PDF report with visual insights made the project highly interactive and user-friendly, especially for stakeholders like farmers, agricultural officers, and policymakers.

The use of open-source tools, deployment via Google Colab, and a lightweight web interface ensured cost-efficiency, portability, and accessibility—making the system well-suited for academic purposes and potential industry applications.

5.2. Future Scope

While the current implementation provides accurate crop price forecasting and useful visual reporting, several future enhancements can improve its robustness and scalability:

- **Integration with Real-Time APIs**: Including live weather, market rates, and government MSP (Minimum Support Price) data can increase the relevance and accuracy of predictions.
- **Incorporating Deep Learning Models**: Implementing models like LSTM or GRU can capture temporal trends in agricultural pricing over seasons or years.
- **Mobile Application Support**: A mobile version of the prediction system can empower farmers in rural areas with on-the-go price predictions and insights.
- **Multilingual Support**: Translating the frontend interface and reports into local languages will improve adoption and accessibility in regional markets.
- **Smart Recommendations**: Adding AI-generated suggestions like “Hold Produce,” “Sell Now,” or “Switch Crops” based on price trends and environmental factors can make the system more actionable.
- **Continuous Learning Pipelines**: Allowing the model to auto-update from new market data would improve prediction accuracy over time.

5.3. Societal Impact

The development of this Crop Price Prediction system has the potential to significantly benefit various layers of society, particularly in agriculture-driven economies:

- **Farmer Empowerment**: Accurate price forecasting can help farmers plan harvest sales, minimize exploitation by middlemen, and maximize income.
- **Market Transparency**: Government bodies and NGOs can use this tool to ensure fair pricing and detect price manipulation.
- **Risk Reduction**: Early insights into pricing trends can help mitigate risks caused by oversupply or unfavorable climatic conditions.
- **Policy Planning**: Agricultural departments can use aggregate prediction data to make informed decisions about subsidies, crop support, and storage strategies.

5.4. References

- **Sharma, R., & Patel, A. (2022).** *Machine Learning-Based Crop Price Prediction Using Decision Trees and Regression Models.*
  DOI: 10.1016/j.compenvurbsys.2022.101768

- **Gupta, S., & Verma, K. (2023).** *Crop Market Price Prediction Using K-Nearest Neighbors and Deep Learning Models.*
  DOI: 10.1007/s00500-023-07458-9