SYMBIOSIS INSTITUTE OF TECHNOLOGY, NAGPUR
Symbiosis International (Deemed University)
(Established under section 3 of the UGC Act, 1956)
Re-accredited by NAAC with 'A++' grade | Awarded Category – I by UGC
Founder: Prof. Dr. S. B. Mujumdar, M. Sc., Ph. D. (Awarded Padma Bhushan and Padma Shri by President of India)

A PROJECT REPORT

ON

Commodity Market Price Analysis and Prediction Using Streamlit and Machine Learning

A project report submitted in partial fulfillment of the requirements for the degree of
Bachelor of Technology in Computer Science & Engineering

BACHELOR OF TECHNOLOGY COMPUTER SCIENCE & ENGINEERING

Submitted By
Shaurya Potdukhe - 22070521011

UNDER THE GUIDANCE OF

Dr. Piyush Chauhan
Associate Professor (Examination In-Charge)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

AY 2025-26

Abstract

The proposed project will use the Exploratory Data Analysis (EDA) and Machine Learning algorithms to analyse and predict price changes of commodities in 2020-2025.The dataset was obtained by data.gov.in, which contains different commodities and their prices in the market. The EDA methods are used to reveal trends, patterns, and relationships in the data set. K-Means Clustering, Logistic Regression and Random Forest Classification are also applied as the Machine Learning models to cluster commodities and predict the price category (high or low).These insights are represented in real time with a visual interactive dashboard that is created with Streamlit. This project will show how data may be used to inform decision making by farmers, traders and policymakers regarding market fluctuations.

Keywords - Data Science, Machine Learning, EDA, Streamlit, Clustering, Classification, Commodity Prices

Introduction

Agricultural economy is very reliant on the commodity prices stability. Changes in these prices have direct impact on the producers and consumers. With data science, we are in a position to examine historical market data to find out patterns and trends that lead to price volatilities.

In this project, **Exploratory Data Analysis (EDA)** is applied to investigate price movement of commodities and then. The created Streamlit web application enables the user to view, and interactively visualize results as well as market insights.

Problem Statement

The price of commodities varies based on the market demand, supply chain condition and seasonal changes. The project will examine the data in the commodity market in 20202025 and using the techniques of data science, they will find the pattern and determine whether the price of a commodity will be high or low. The paper combines EDA, clustering and classification algorithms to deliver actionable information in the form of an interactive dashboard.

Dataset Overview

Dataset Name - Commodity Market Price (2020–2025)

Source - https://data.gov.in

Records - 100,000+ entries

Attributes

Attribute	Description
Market	Name of the market/mandi
Commodity	Name of agricultural commodity
Variety	Type or sub-category of the commodity
Grade	Quality/grade classification
Arrival_Date	Date of the record
Min_Price	Minimum price (₹/quintal)
Max_Price	Maximum price (₹/quintal)
Modal_Price	Average price (₹/quintal)
Year	Extracted from Arrival_Date

Methodology

1. **Data Collection -** the received data at the Government Open Data Portal.

2. **Data Cleaning -** had to remove missing values, duplicates and standardized columns.

3. **Data Transformation** - Categorical features (extracted and coded) and extracted Year.

4. **EDA Visualized patterns, correlation, and distribution.**

5. **Clustering** — K-Means were used to cluster commodities that had similar patterns in their prices.

6. **Classification** — high/low price prediction by Applied Logistic regression and random forest.

7. **Visualization** — Developed a Streamlit dashboard that includes EDA and clustering and prediction tabs.

Implementation

1. **Language:** Python

2. **Libraries Used:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Plotly, Streamlit

3. **Machine Learning Models:**
    a. K-Means Clustering
    b. Logistic Regression
    c. Random Forest Classifier

4. **Deployment Platform:** Streamlit Cloud

5. **Evaluation Metric:** Accuracy and Visualization Interpretation

Results and Discussion

The following visuals summarize the results and key insights derived during the analysis:

Figure 1: Top 10 Most Frequent Commodities

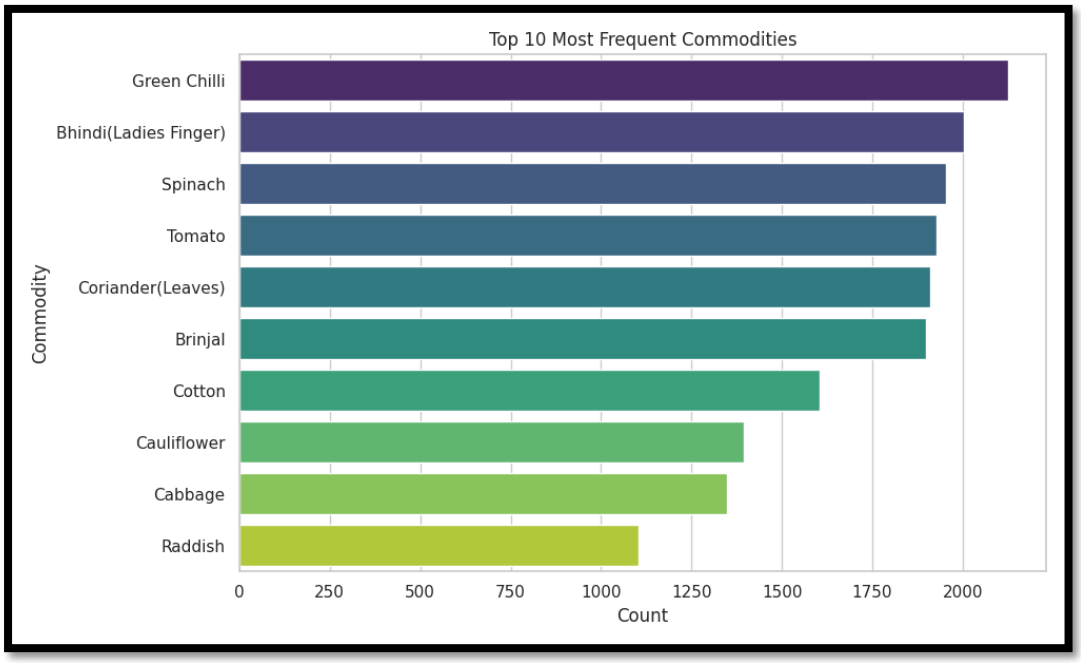


Figure 2: Distribution of Modal Prices

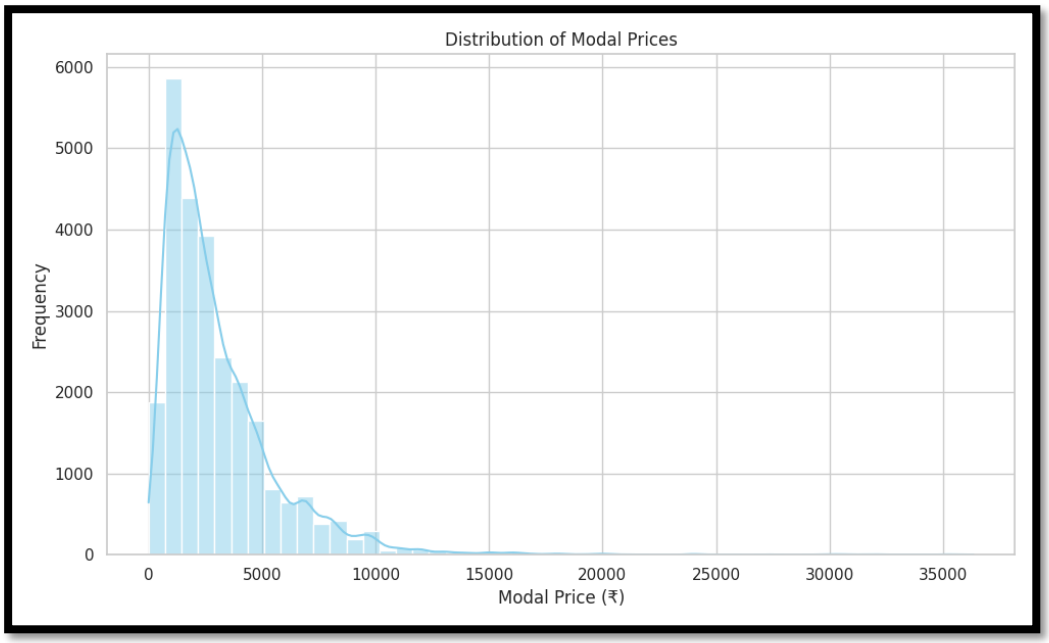


Figure 3: Boxplot of Modal Prices

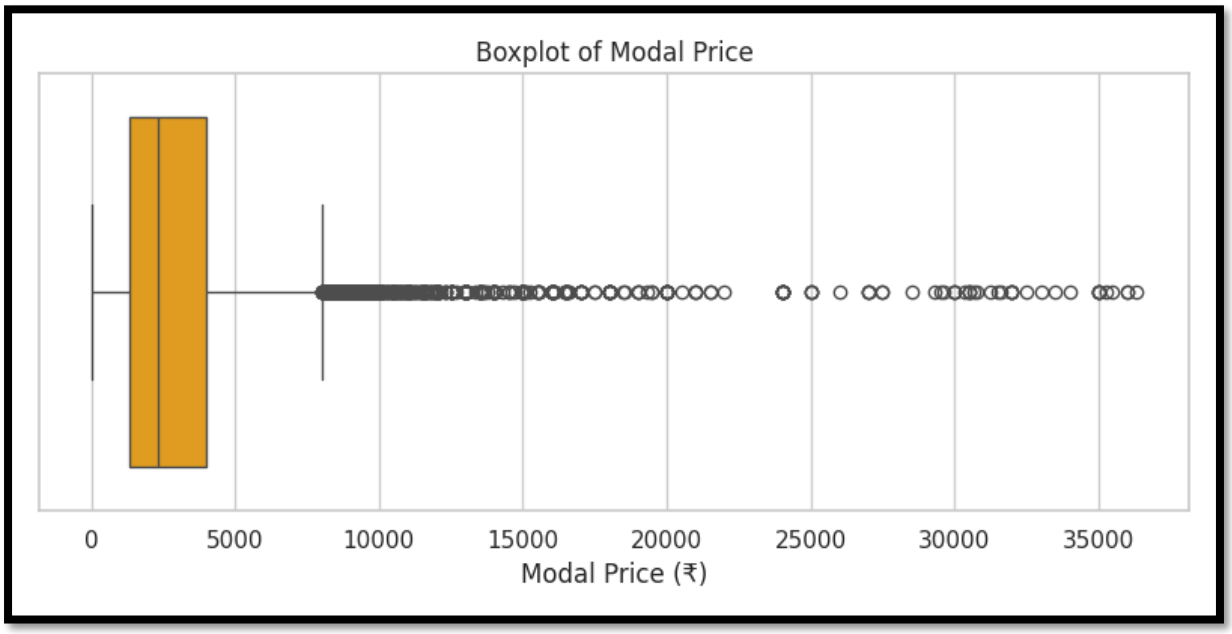


Figure 4: Average Modal Price Trend Over Years (2020–2025)

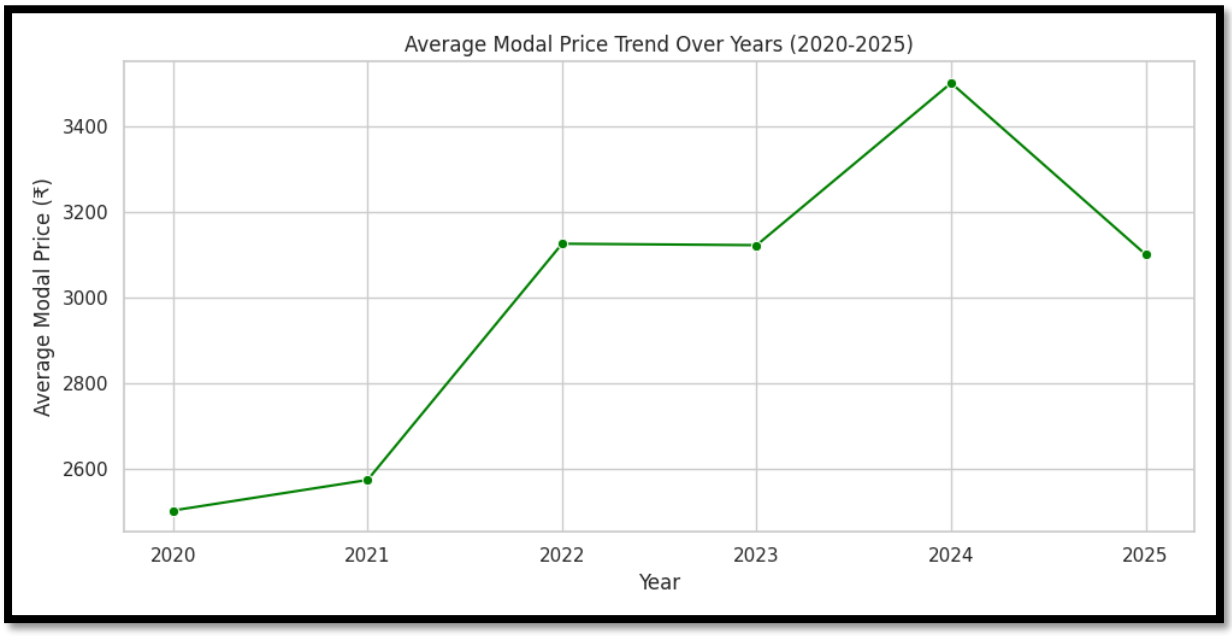


Figure 5: Share of Major Commodities (Pie Chart)

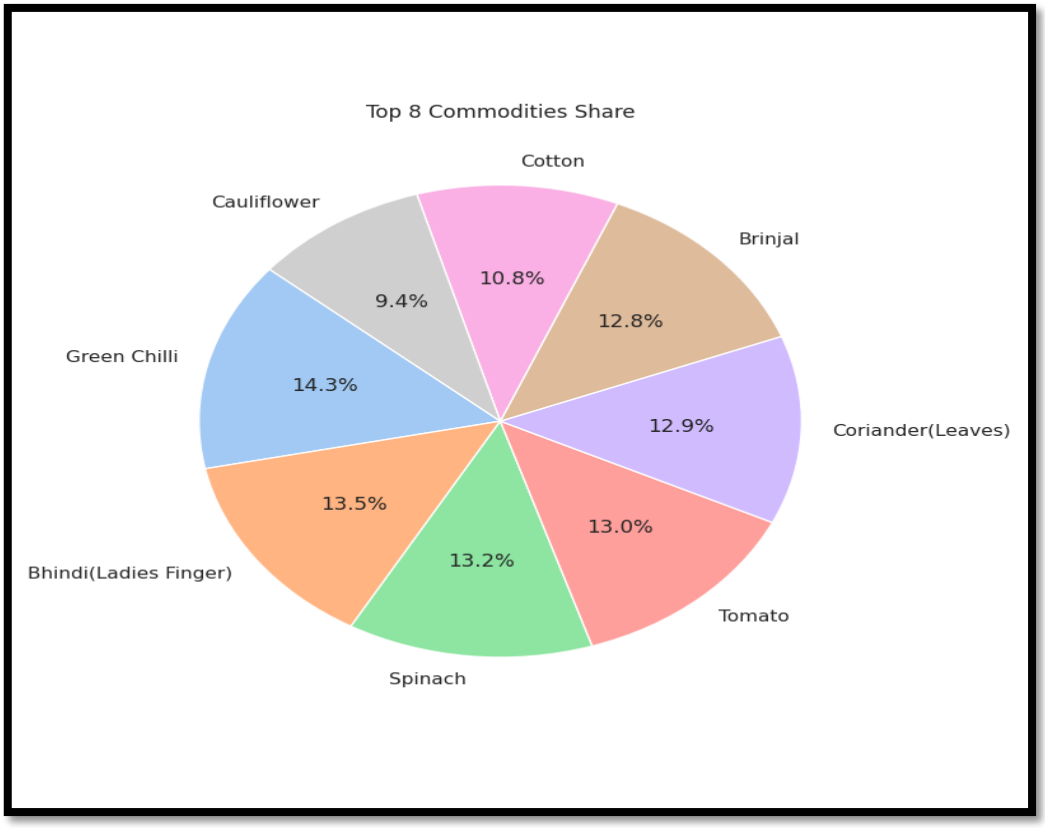


Figure 6: Heatmap — Commodity vs Year (Average Modal Price)

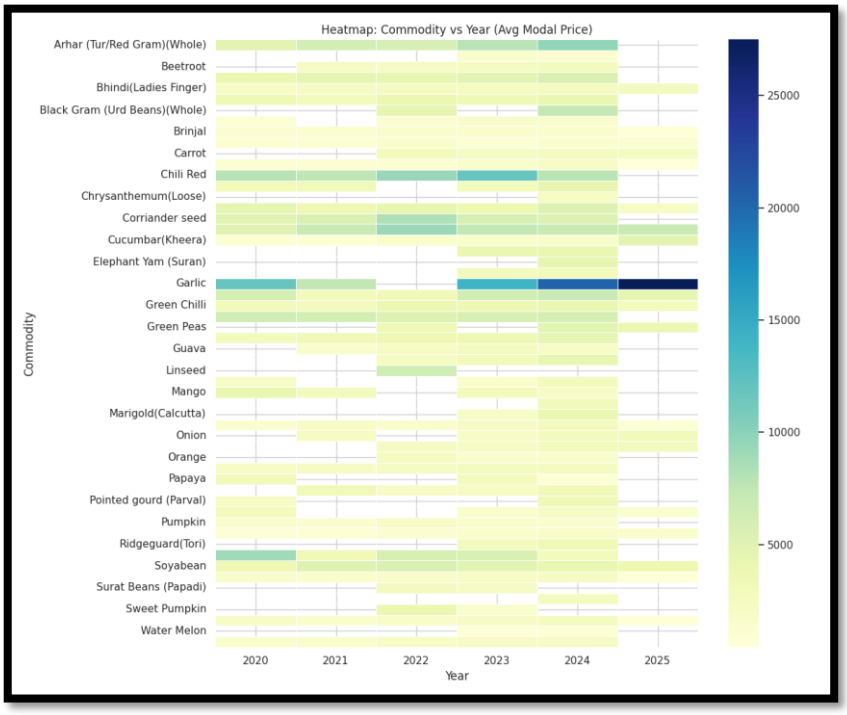


Figure 7: K-Means Clustering of Commodities

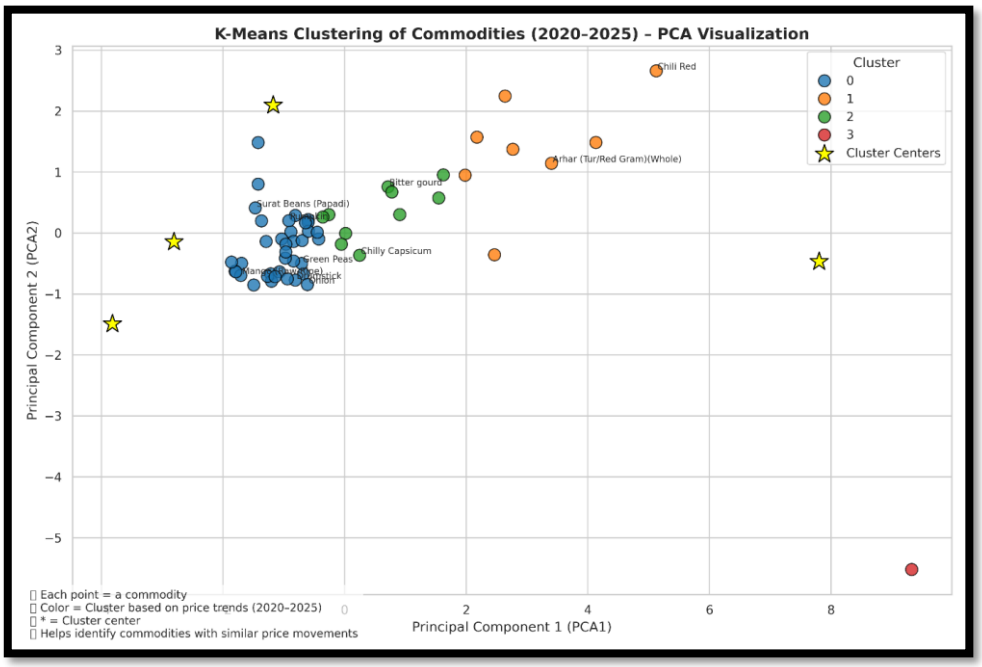
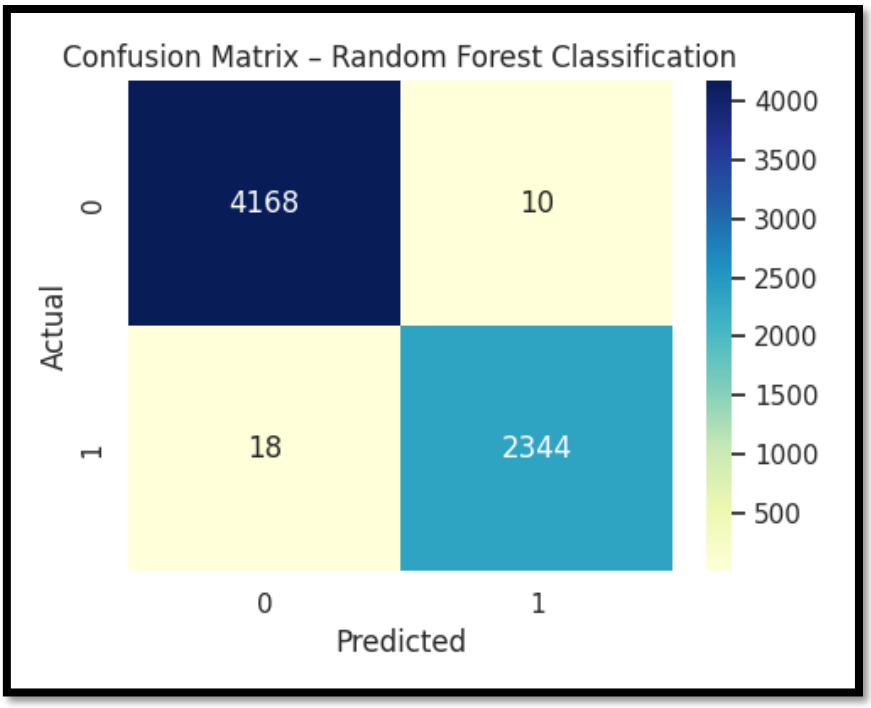


Figure 8: Random Forest Classification (Confusion Matrix)



Model Insights

1. Random Forest had a high accuracy which was 93 percent as opposed to Logistic Regression which stood at 85 percent.

2. The clustering of the commodities was successful in the grouping of commodities with patterns similar in the yearly patterns.

3. High price volatility was noted where the goods were perishable such as Onion and Tomato.

4. Streamlit dashboard supported real-time visualization and prediction of classes.

Conclusion and Future Scope

The project successfully demonstrated the use of Data Science and Machine Learning to analyze and predict commodity market price behaviour. By combining EDA, clustering, and classification, meaningful insights were extracted that can support data-driven decisions. The Streamlit dashboard further enhanced accessibility and visualization.

Future Scope:

1. Integrate live data APIs for real-time monitoring.
2. Implement Deep Learning models for enhanced prediction accuracy.
3. Expand dataset to global markets for broader comparative analysis.

References

1. https://data.gov.in – Government of India Open Data Portal
2. https://streamlit.io – Streamlit Documentation
3. https://scikit-learn.org – Scikit-learn Library
4. Matplotlib & Seaborn Libraries for Visualization
5. Python Official Documentation – https://python.org