AGRICULTURE COMMODITIES PRICE PREDICTION AND FORECASTING

GUIDE BY : R. DHAMOTHARAN,M.E.,

ASSISTANT PROFESSOR

PRESENTED BY

SURYA S(9922146016)

M.SC DATA SCIENCE

ABSTRACT

- In India, agriculture plays a pivotal role in the economy, yet farmers often lack support and information in regional languages.
- This study implements machine learning algorithms such as multi linear regression, Random Forest and Decision Tree regressor.
- Current methods for price prediction on short-term arrivals and historical data, but they fail to offer adequate recommendations for storage or sales decisions.
- This study aims to review existing research in this field, evaluate different models, and suggest avenues for improvement.

PROBLEM STATEMENT

- In India, agriculture is a key GDP contributor, but farmers lack regional language support.
- Current price forecasting relies on limited data, offering insufficient recommendations for farmer decisions.
- This study aims to review existing research, identify model pros and cons, and propose enhancements for more effective crop recommendations and price forecasting in Indian agriculture.

EXISTING SYSTEM DRAWBACK

- In the previous research, the dataset attributes is limited. In some cases, they preferred the soil types or in some cases, they focused on the temperature.
- The proposed technique does not guarantee perfect forecasts. The farmers will not get the proper yield or some time the farmers will not get the proper price for the commodities they grown.
- So, the immediate requirement is to help farmers go get the profit for the commodities they grown.

PROPOSED SYSTEM ADVANTANGES

- To measure or to learn the performance of the proposed model Random Forest and Support Vector Machine techniques are applied.
- Maximum Relevance Approach issued to improve the forecast accuracy.
- The datasets includes all attributes like soil. temperature, historical places and etc to gain more insight on the data.

SOFTWARE AND HARDWARE REQUIREMENTS

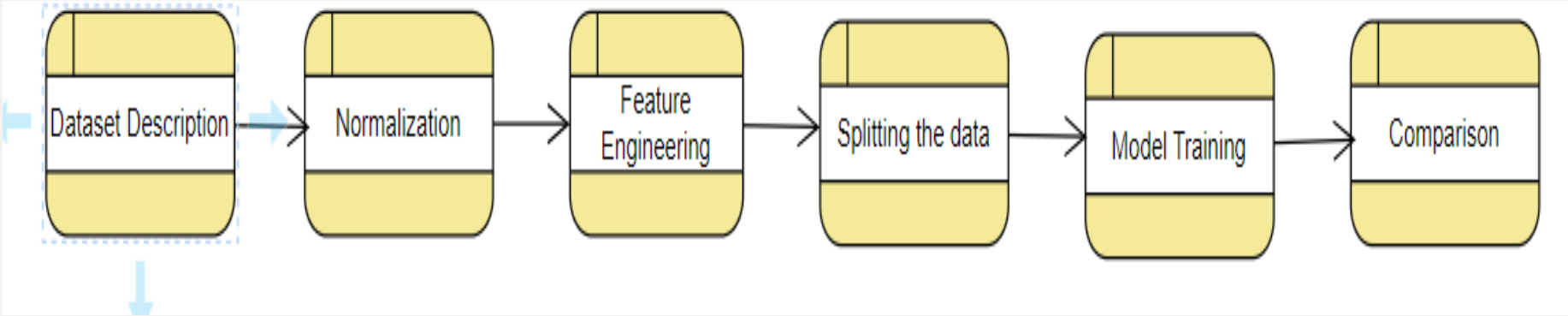
SOFTWARE REQUIREMENTS

- Operating System : Windows 11
- Coding Language : Python
- Working Platform : Google Colab

HARDWARE REQUIREMENTS

- System : Intel i5 processor
- Hard Disk : 500GB
- Monitor : 15’’’LED
- Input Devices : Keyboard , Mouse
- Ram : 8GB

DATA FLOW DIAGRAM



DATASET DESCRIPTION

- Source

  https://drive.google.com/file/d/10YmsnJoaEh0bNib6Je_JGwvqbgYIWktE/view?usp=drive_link

- Number of Data : 62429

- Number of attributes : 11

- Size : 4.72 MB

NORMALIZATION

- MinMaxScaler is used to scale features to a specified range, typically between 0 and 1. This scaling technique ensures that all features contribute equally to the model's learning process, regardless of their original scale or magnitude.
- This proposed normalization will give the relative relationships between different data points are maintained, allowing the model to learn from the data effectively without introducing any bias due to differences in scale.

FEATURE ENGINEERING

- Label Encoder is utilized to transform categorical features such as commodity, APMC, Month, district name, and state name into numerical representations.
- This encoding process assigns a unique integer to each category, allowing machine learning algorithms to interpret and process these features effectively.

SPLITTING THE DATA

- The dataset has been divided into two – Training and Testing
- The proportion of data allocated to each set is determined by the test_size parameter. In this case, 20% of the data is assigned to the testing set, while 80% is assigned to the training set.

MODEL TRAINING

- In the present investigation, an attempt has been made to explore efficient ML algorithms e.g., Decision Tree Regression, Random Forest (RF) and Multiple Linear Regression for forecasting wholesale price of crops in 33 major markets of Maharashtra.
- The superiority of the models is established by means of R2-score, and other accuracy measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Prediction Error (MAPE).

COMPARISON

Machine Learning	MAE	MSE	R2
Multi linear regression	0.0104	0.00030	0.46
Decision tree regression	0.0025	0.00010	0.84
Random forest regression	0.00076	0.00008128	0.88

OUTPUT

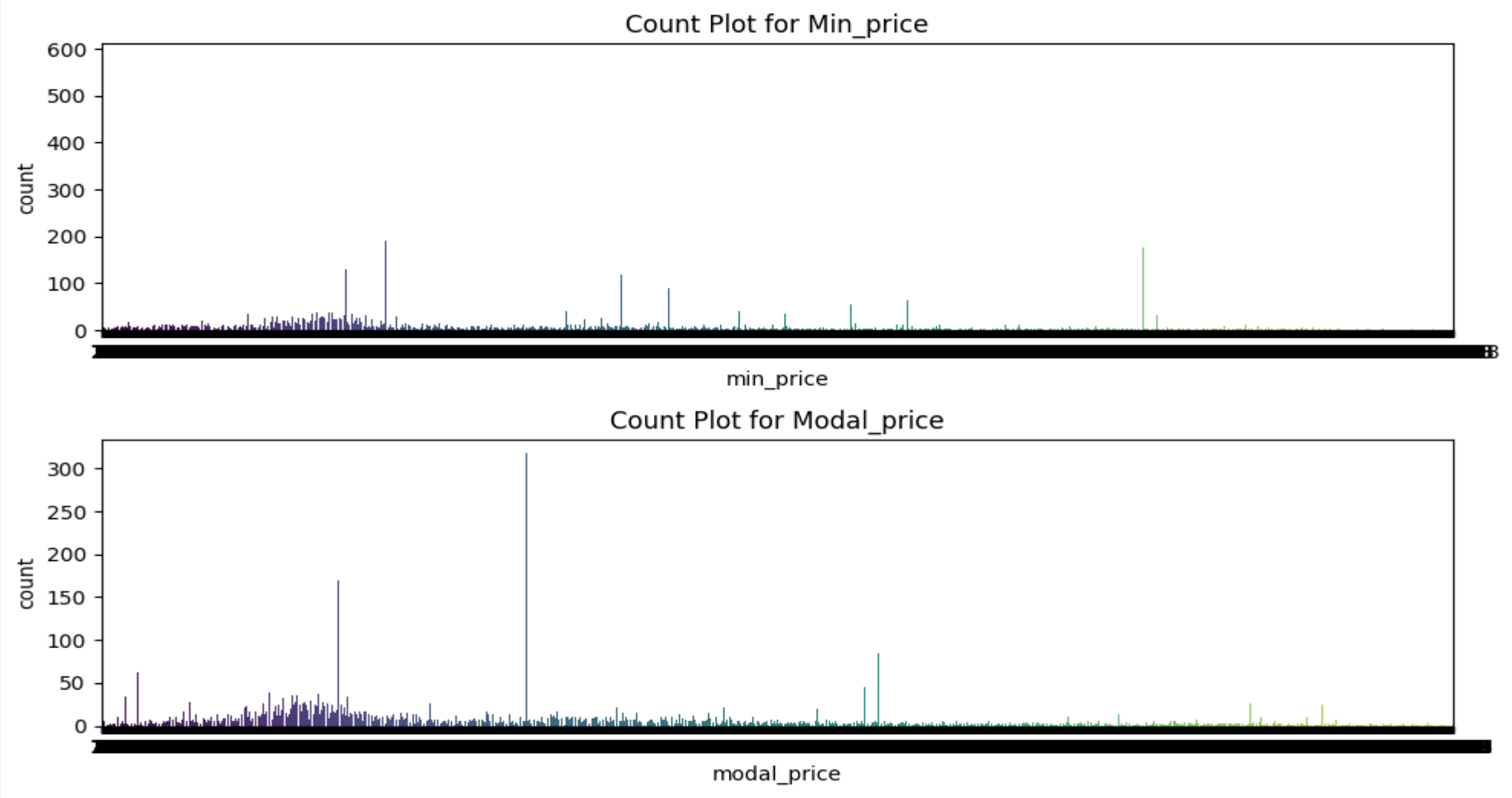


Fig.1 COUNT PLOT

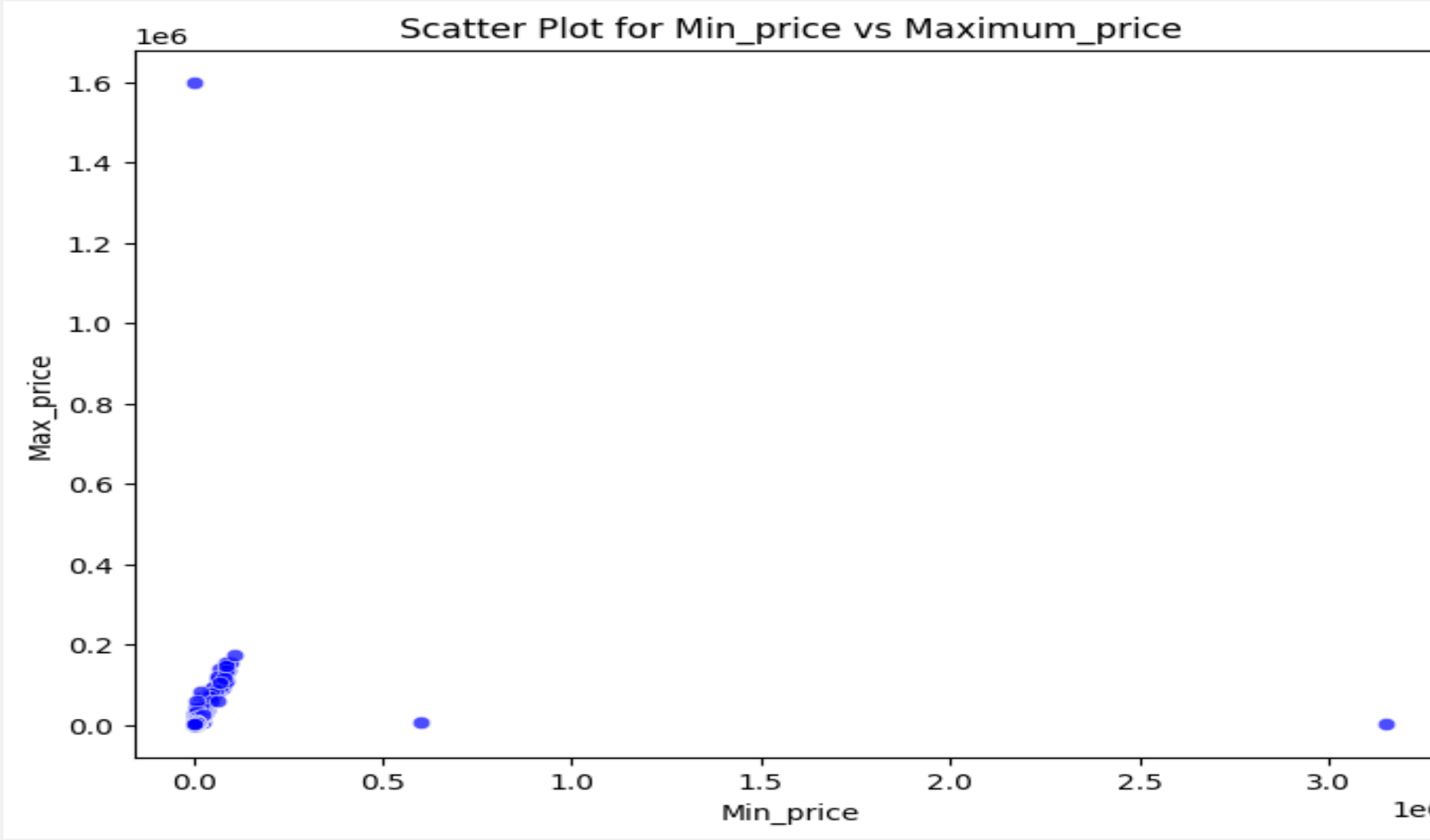


Fig.2 SCATTER PLOT

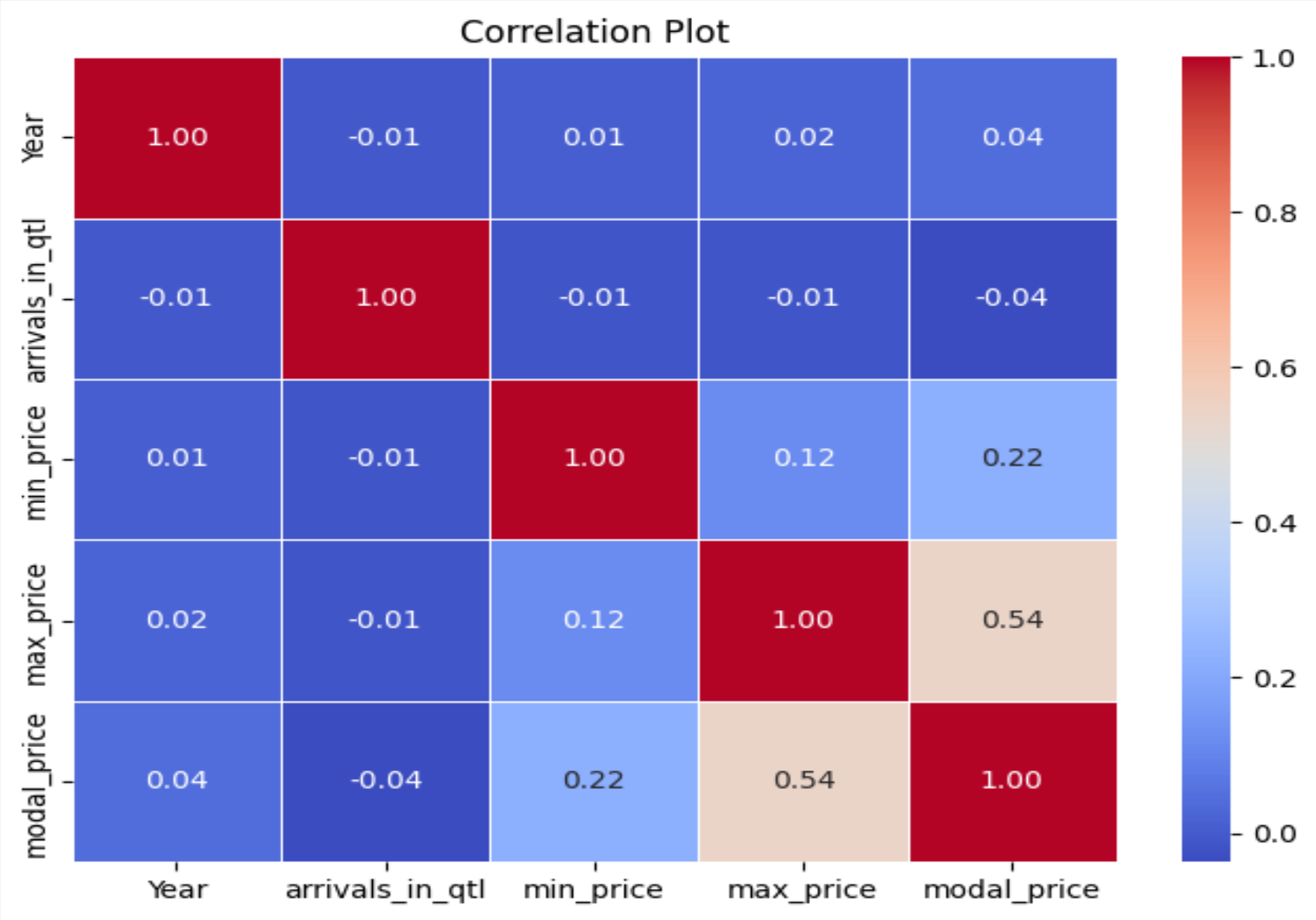


Fig.3 CORRELATION PLOT

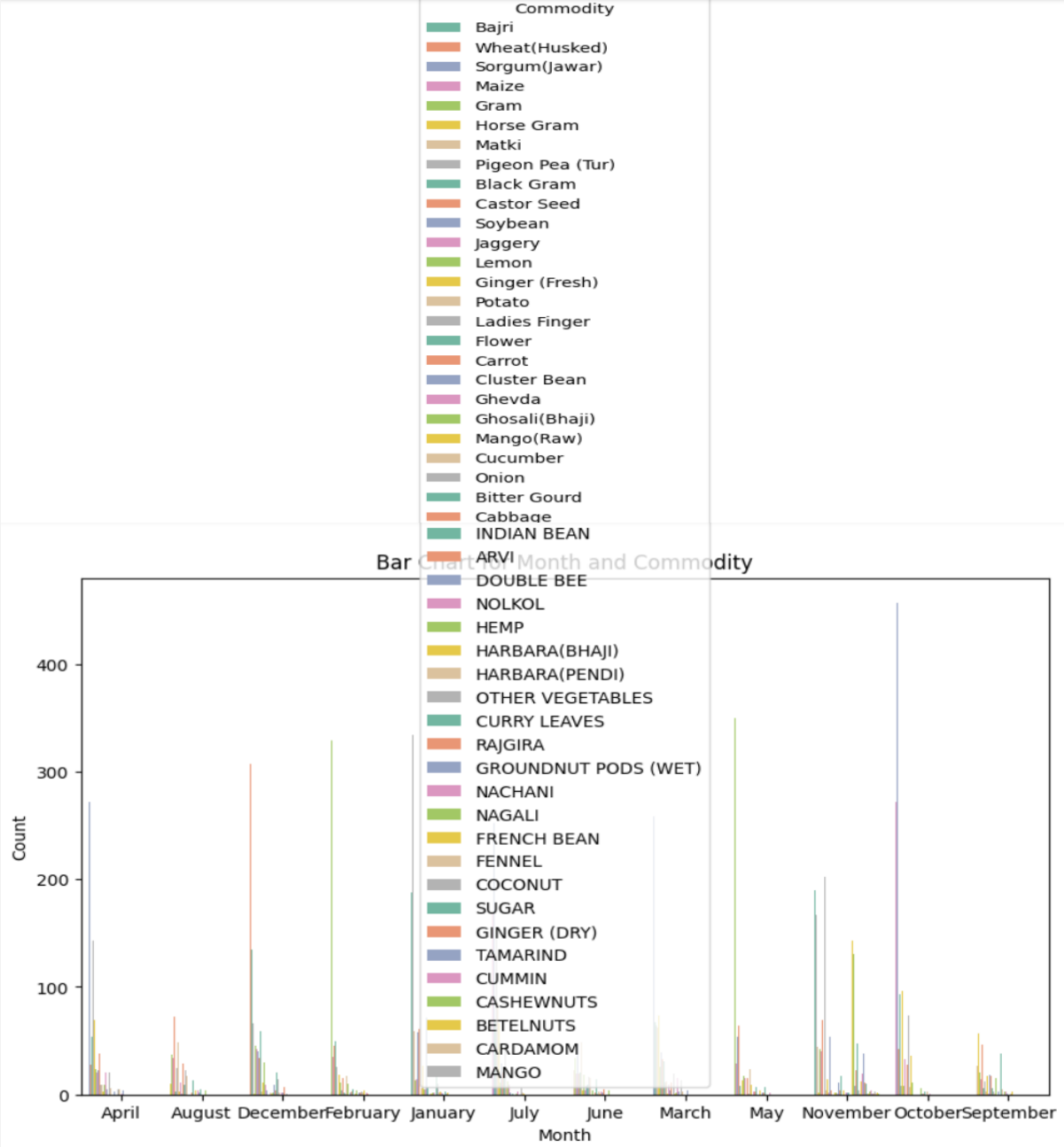


Fig.4 BAR for MONTH and COMMODITY

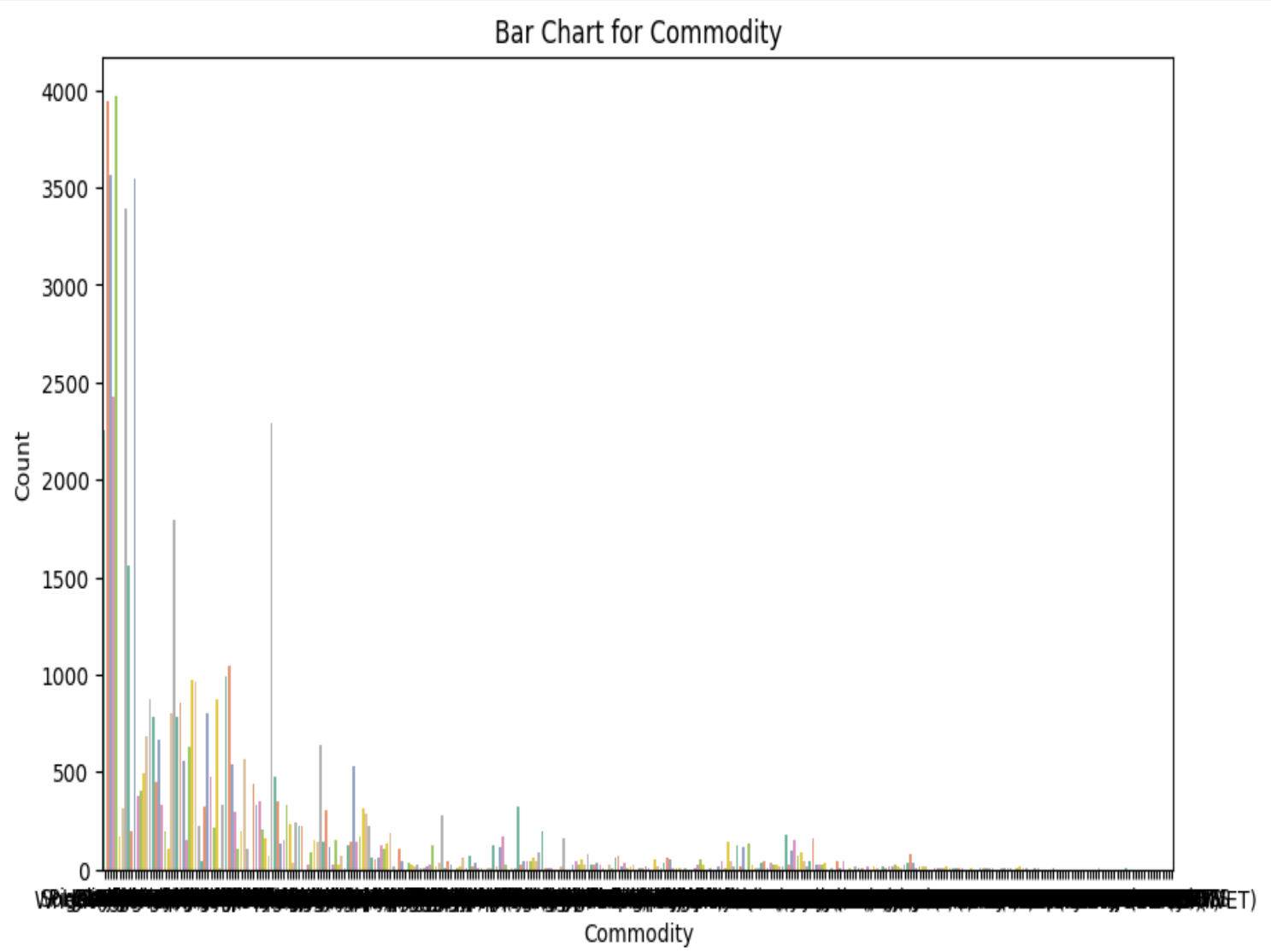


Fig.5 BAR CHART for COMMODITY

CONCLUSION

- In our country, India, majority of the people employed with agriculture.
- It is very important to give more importance for development in agriculture sector.
- So, the immediate requirement is to help formers go get the profit for the commodities they grown.
- This frameworks developed to predict and recommend the yield as well as price.
- But most of the cases only specific attributes are used either like, soil, temperature, historical price etc.

REFERENCE

- https://www.manage.gov.in/kcc/kcc.asp
- https://mkisan.gov.in
- S.Kavitha and N. Anandaraja, “Constraints and Suggestions as Percived by the Kisan Call Center Beneficiaries,”. International Journal of Pure and Applied Bioscience, July 2017.
- Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q. Vera Liao, Khai Truong, and Shwetak Patel, “FarmChat: A Conversational Agent to Answer Farmer Queries,” Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 2, No. 4, Article 170, 2018.
- Hai Jin, Yi Luo, Chenjing Gao, Xunzhu Tang, and Pingpeng Yuan, “ComQA: Question Answering Over Knowledge Base via Semantic Matching,” IEEE Access, vol. 7, 2019.