Market-Driven Crop Planning Using Machine Learning Techniques

Neha Sawant1, Pradnya Mehta2, Anuradha Yenkikar3

Department of Computer Science and Engineering (Artificial Intelligence),

Vishwakarma Institute of Information Technology, Kondhwa, Pune.

[1nehasawant2801@gmail.com](mailto:1nehasawant2801@gmail.com), [2pradnya.mehta@viit.ac.in](mailto:2pradnya.mehta@viit.ac.in), 3anuradha.yenkikar@viit.ac.in

***Abstract*—Even in this day of advanced technology, Indian farmers continue to practice age-old traditional agricultural techniques. Numerous problems plague Indian farmers in their line of work. Despite the presence of numerous urgent problems, it is essential to focus on market-driven crop planning. The tragic news of farmer suicides hitting headlines makes it obvious that money plays a big role. A lot of farmers are struggling because of the unpredictable weather and shifting market prices and many are deeply in debt and unable to repay their debts. Even though nature is beyond human’s control, human can utilize technology, such as machine learning, to forecast crop prices by using historical data. To do this, the proposed system has proposed a machine learning method that uses the Random Forest Algorithm, Decision Tree Regressor, XGBoost and Support Vector Machine. The suggested method incorporates data for the years 2021, 2022 and 2023 and uses monthly minimum and maximum prices from the Pune city market for the 14 fruits. The following methods have R2 scores: Support Vector Machine (SVM) 0.87, XGBoost 0.87, Random Forest 0.89, and Decision Tree Regressor 0.93. Therefore, it has been determined that the decision tree classifier is the most appropriate method for market crop price prediction after analysing a number of characteristics. Using a trained decision tree model, market prices for every fruit for 2024 have been projected month by month.**

***Keywords*—Decision Tree, market price, cultivation, machine learning, Indian farming**

I. INTRODUCTION

Technology is advancing in every industry and facet of life today. However, not all industries in India have embraced technological revolutions, including smart agriculture that makes use of AI or machine learning. For many years, farming in India has not been successful. This could be due to a number of factors, including unreliable water supplies, a lack of compensation, fragmented land holdings, difficulty obtaining formal credit and insurance, loss of agricultural land, a lack of infrastructure, soil erosion, insufficient controls, redundant seeds, poor quality control, irrigation issues and the topic of this paper agricultural marketing challenges. After analysing the aforementioned issues, it is clear that none of the other issues are under human control. For example, although technology makes an effort to forecast weather, rainfall and natural disasters, farming is entirely dependent on the elements, so even with technological predictions, humans can only partially mitigate the effects of these uncontrollable events. However, the focus is on agricultural products that have been successfully gathered and crop production completed, in case of pricing. A lot of work has gone into growing and harvesting the crop by farmers. They have had backing from even nature, yet the market lets them down. Therefore, the suggested methodology concentrates on methods to recommend the most lucrative crop selection for growing utilizing agricultural economy.

A few websites that the government has created to assist farmers with their problems are not fully operational. While a great deal of research is being conducted in the areas of weather forecasting, soil quality prediction and crop yield prediction, relatively little research is being done in the area of market-driven crop planning. It is simple to access historical or previous share price performance on trading apps or websites; however, agricultural items are not included in this type of price performance. Therefore, the strategy put out in this research aims to close this gap by creating a trustworthy model that would assist farmers in selecting the crop to grow in order to maximize profits. In order to provide farmers with a report that looks like a graph and includes pricing information in tabular form, the suggested approach would forecast the prices of 14 fruits for the Pune city market for the full year.

In order to achieve this goal, real-time market price data has been gathered from trade and official websites. This includes the lowest and highest prices of 14 fruits over a period of three years, broken down by month. Using this dataset, the suggested method examined four distinct algorithm outcomes based on various factors and conducted a comparative performance analysis to determine which algorithm is most appropriate for agricultural market pricing data. The suggested approach determines the optimal outcomes, provides a model and suggests it for market price forecasting. With the use of a graph and table, the trained model will assist farmers in seeing the prices of agricultural items over the next few months.

# II.LITERATURE SURVEY

Literature survey of 10 previously available literatures is as follows:

Authors in [1] have explored traditional forecasting methods, intelligent techniques and combinational models. Their finding is, growing trend of using combined models, for accurate price forecasting, integrating structured and unstructured data and emphasizing both value accuracy and trend precision.

Authors have used 3 different methods for prediction in [2]. Among those, Random Forest Algorithm has given 95% accuracy to them.

In [3], authors have implemented Linear Support Vector Machine (LSVM) to optimize approach, which have given them 91% of accuracy. Authors have tried to enhance the classification accuracy of crop yield prediction by these methods.

CNN-Bidirectional LSTM model is used for forecasting agricultural commodity prices in Gujrat in paper [4]. It outperforms traditional CNN and LSTM models, achieving lower MAE and RMSE and an R2 closer to1.

Approach of Time Series Forecasting with focus of Auto Regressive Integrated Moving Average (ARIMA) modeling to predict agricultural market trends is implemented in paper [5]. Authors have gathered historical monthly data on crop prices and have used ARIMA model to analyze and make predictions. To ensure accuracy, they have checked for stationarity in the data using the Augmented Dickey-Fuller Test (ADF) and have selected the best ARIMA model parameters based on the Akaike Information Criterion (AIC).

The authors of paper [6] have implemented a Machine Learning (ML) based approach for predicting agricultural prices, they have focused on crops like rice, wheat, milk and fruits such as mangoes. They have utilized a 5-year dataset from the NAFHA source and have applied clustering algorithms along with Naïve Bayes for improved accuracy. The Random Forest Algorithm has been chosen by authors for its ability to handle dynamic data, including environmental and climate factors.

Utilization of Decision Tree Algorithm to analyze various parameters such as season, location, seed sales and market trends can be seen in paper [7]. Their model is giving 80% accuracy in price prediction and authors are expecting future improvements to increase accuracy up to 95% by including additional parameters like petroleum and investment costs.

System AGRI-PRO developed by authors in paper [8] recommends crops, fertilizers and marketplaces to farmers. Authors have used Random Forest, Decision Tree, K-nearest Neighbors (KNN) and Support Vector Machine (SVM) to analyze data and make predictions. Their proposed system considers factors like soil nutrients, weather conditions and market trends to suggest the most suitable crops and profit. According to their findings, Random Forest Algorithm showed the highest accuracy in predicting the best crops and fertilizers.

Implementation of Decision Trees and Neuro-evolutionary algorithms can be seen in paper [9] to predict crop prices and improve crop productivity. Decision Tree technique is employed for price and profit prediction considering factors like rainfall and production costs. Neuro-evolutionary algorithm optimized predictions by simulating neural evolution processes.

Authors have conducted comparative study on data mining techniques for predicting agricultural crop prices in [10]. They have utilized various methods such as regression analysis, tracking patterns, cluster analysis and visualization techniques to analyze data.

|  |  |
| --- | --- |
| Reference no | Algorithms used |
| [11] | Random Forest Algorithm, Convolutional Neural Network (CNN), K-Neast Neighbour(KNN) |
| [12] | Convolutional Neural Network (CNN) |
| [13] | K-Nearest Neighbor (KNN) |
| [14] | Random Forest, Decision Tree, XGBoost |
| [15] | Support Vector Machine, Random Forest |

III.METHODOLOGY

Proposed system has employed a comprehensive data-driven approach to predict the minimum and maximum prices of various fruits over different months and years. The dataset, considered of 540 entries, each detailing the fruit name, month, year and corresponding price range. Dataset is prepared after gathering real time data from Pune city market website.

**Data preprocessing** initially, proposed method conducted data preprocessing to ensure the quality and integrity of the used dataset. This involved:

* Data Cleaning: Removal of any missing values to maintain consistency.
* Feature Encoding: Conversion of categorical variables, such as ‘Fruit name’ and ‘month’ (fig. 1), into numerical representations (fig. 2) to facilitate computational analysis.



fig. 1. Attributes & the features of the price prediction system before encoding

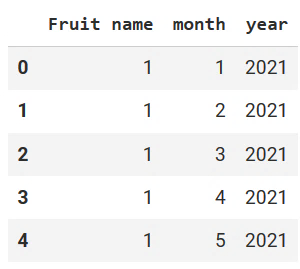


fig. 2. Attributes & features after encoding

**Model Selection and Training**, proposed method selected a suite of regression models to capture the nuances of the prediction task.

A **decision tree** is like a flowchart where each step (node) is a question, the branches are the answers to those questions, and the final endpoints (leaves) give the outcome. The decision tree is trained using the data in X (features: 'Fruit name', 'month', 'year') and Y (target variables: 'minimum price', 'maximum price').

**XGBoost**(Extreme Gradient Boosting) is a more advanced algorithm that works by creating multiple trees and perfecting the model with each round. It's designed for speed and high performance. Here, the XGBoost regressor is being used to predict fruit prices. Two XGBoost models are trained, one to forecast the "minimum price" and the other for "maximum price."

Both methods are used to predict fruit prices in the dataset. Decision trees provide a simple, easy-to-understand model, while XGBoost offers a more advanced and powerful model that is likely to make better predictions.

**Random Forest** is ensemble of decision trees. Random forest operates by constructing multiple decision trees during training & outputting the average prediction of the individual trees. It’s like having a team of experts(trees) & going with the majority vote. By averaging multiple trees, it reduces the risk of overfitting, which is common in single decision trees. It can handle categorical & numerical data, as seen with the ‘Fruit name’, ‘month’ and ‘year’ features.

Whereas, **SVM** aims to find the best boundary(hyperplane) that separates classes of data. In this case, it tries to distinguish between different price levels.It maximizes the margin between the data points and the hyperplane, which helps in making robust predictions. The Radial Basis Function (RBF) kernel allows SVM to handle non-linear data by mapping it to a higher-dimensional space where it is linearly separable.

**Model Evaluation** to assess the performance of proposed model, all the algorithm results have been evaluated using different matrix.

**R2 score**, also known as the coefficient of determination, tells us how well the independent variables(features) explain the friction in the dependent variable(target), in this case, fruit prices. It is calculated using formula (01). The score ranges from 0 to 1, where 1 indicates a perfect fit. In simpler terms, a higher R² score suggests that the model is better at explaining the factors that influence fruit prices. A score close to 1 signifies that the model is effective in predicting fruit prices.

**= ……….** 01)

**MSE** measures the average squared difference between the factual and predicted values. In the formula (02), yi refers to ith observed value and yrefers to the corresponding predicted value. (n=No. of observations). It provides a relative measure of the model's quality. Lower values are better. MSE tells on average, how far off model predictions are from the factual fruit prices. Lower MSE values mean model is better at predicting fruit prices

**MSE = )2 ……**02)

**MAE** measures the average absolute difference between the factual and predicted values (Formula 03). Like MSE, it provides a relative measure of the model's quality. Lower values are better. MAE tells on average; how important predictions deviate from the factual fruit prices. Lower MAE values mean model is better at predicting fruit prices.

**MAE = ……**03)

**MAPE** measures the normal of the absolute chance differences between the factual and predicted values. Formula (04) is used to calculate MAPE. It's expressed as a chance and provides a relative measure of the model's quality.

**MAPE=**

**…**04)

**Mean Percentage Error (MPE)** calculates the average of the percentage errors between actual and predicted values as shown in the formula (05). It indicates the model’s tendency to overestimate or underestimate the data. (n=No. of observations)

**MPE =**  **……**05)

**Huber Loss** is a combination of squared error loss for small errors and linear loss for large errors. It’s less sensitive to outliers than mean squared error.

* For small errors (when the absolute difference between the actual value(ytrue) and the predicted value(ypred) is less than or equal to δ. The parameter δ is a threshold that determines the point at which the loss function changes from a quadratic function to a linear function as shown in the formula (06):

**…**06)

* For large errors (when the absolute difference between the actual value and the predicted value exceeds δ) formula (07) is used in this case.

**.…..**07)

**Mean Squared Logarithmic Error (MSLG)** computes the mean of the squared logarithmic differences between actual and predicted values, penalizing underestimates more than overestimates.

**…..**08)

Here in formula (08),

(n) represents the total number of observations

(yi) is the actual target valuefor the ith observation

( is the predicted value for the ith observation.

**Theil’s U Statistic** measures the accuracy of a forecast by comparing the forecast error to the error of a naive model that simply predicts the mean of the actual values.

Theil’s U=Naive Model ErrorMean Absolute Percentage Error​

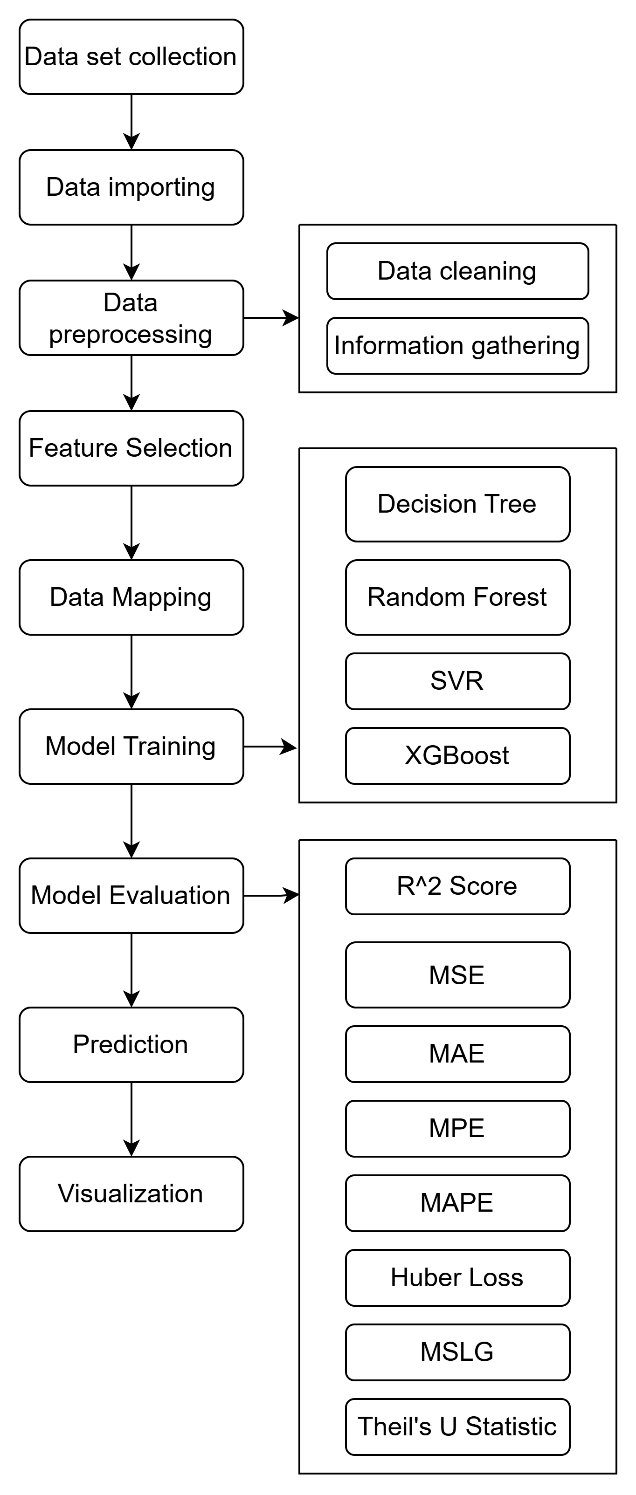


fig.3. Flow diagram

**Prediction and Visualization:** With the trained models, proposed system predicted the price ranges for the year 2024, creating a comprehensive forecast that included all fruit names and months. Proposed method further visualized the predicted prices using line graphs to illustrate the trends and fluctuations throughout the year.

# IV.RESULTS

Comparison table for Decision Tree Classifier, Random Forest Algorithm, SVM and XGBoost results on R2 score, MSE, MAE, MPE, MAPE, Huber Loss, MSLG and Theil’s Statistics.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Validation Method | Decision Tree | Random Forest Algorithm | SVM | | XGBoost | |
| Min & max | Min & max | min | max | min | max |
| R2 Score | 0.93 | 0.89 | -0.24 | -0.09 | 0.87 | 0.99 |
| MSE | 455643.51 | 2031434.23 | 6259702.13 | 47897795.93 | 605177.53 | 361097.88 |
| MAE | 134.53 | 396.12 | 1414.81 | 4364.62 | 151.20 | 285.53 |
| MPE | -1.26 | -4.12 | 1.81 | -35.22 | -2.42 | -0.76 |
| MAPE | 2.95 | 8.90 | 53.28 | 83.89 | 6.68 | 5.00 |
| Huber Loss | 134.48 | 395.67 | 1414.33 | 4364.14 | 150.70 | 285.03 |
| MSLG | 0.012 | 0.025 | 0.747 | 0.669 | 0.021 | 0.007 |
| Theil’s U Statistics | 0.016 | 0.048 | 0.45 | o.69 | 0.05 | 0.04 |

Table 1. Comparison table

**R2 Score:** Decision Tree (0.93) and XGBoost for maximum price (0.99) have higher R2 scores. SVM has a negative R2 score which indicates poor performance as shown in fig. 4.

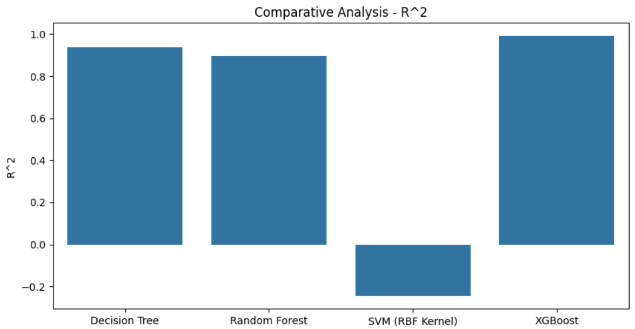


Fig.4. Comparative analysis- R^2 Score

**MSE:** It is visible in fig. 5 that, decision tree has the lowest MSE (455643.51), Random Forest algorithm and XGBoost for maximum price follow closely and SVM for maximum price has highest MSE (47897795.93).

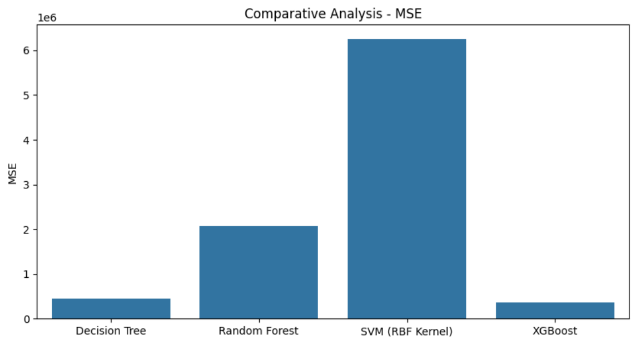


Fig.5. Comparative analysis- MSE

**MAE:** Decision Tree performs well (134.53) Random Forest and XGBoost have higher MAE, SVM have the highest MAE values (fig. 6).

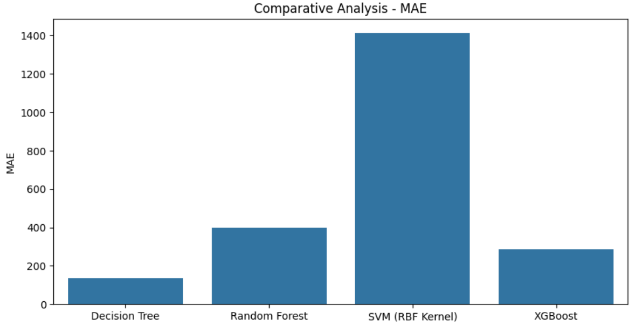


Fig.6. Comparative analysis- MAE

**MPE and MAPE:** From fig. 7 and fig. 8, decision Tree, Random Forest Algorithm and XGBoost have low MPE and MAPE. SVM has positive MPE and highest MAPE (83.89).

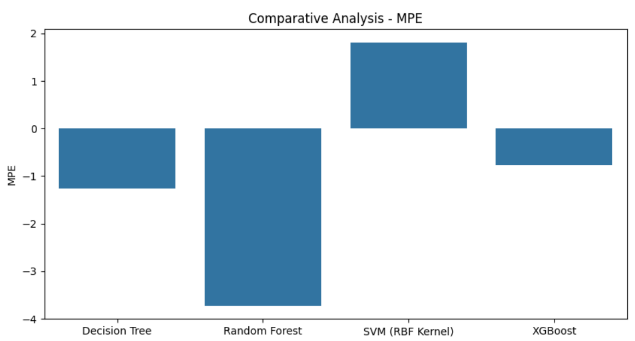


Fig.7. Comparative analysis- MPE

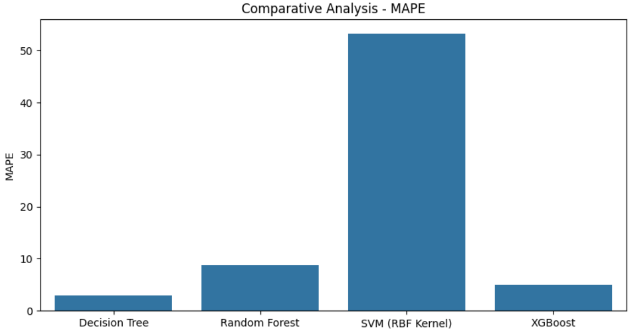


Fig.8. Comparative analysis- MAPE

**Huber Loss:** In fig. 9, decision Tree performed best, then XGBoost and Random Forest. SVM have higher Huber Loss.

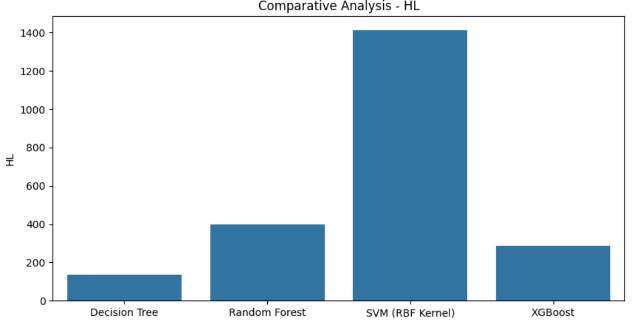


Fig.9. Comparative analysis- Huber Loss(HL)

MSLG and Theil’s Statistics: XGBoost, Decision Tree and Random Forest have excelled in both. SVM has lag behind (fig 10, fig 11).

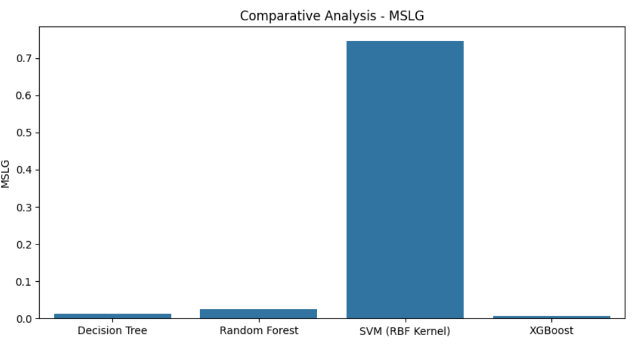


Fig.10. Comparative analysis- MSLG

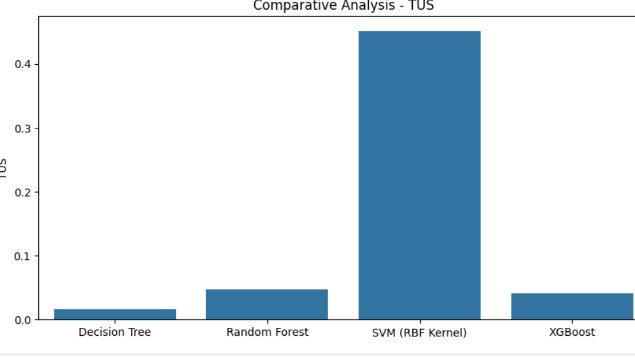
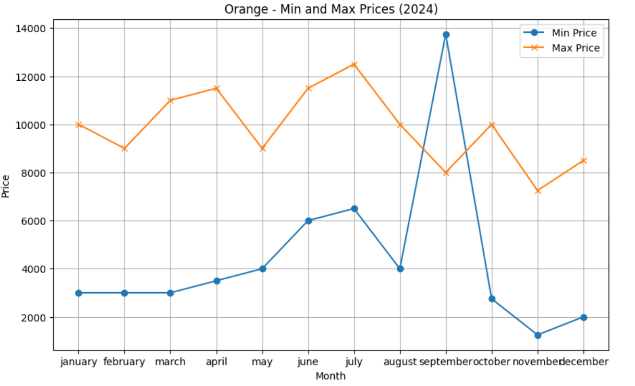
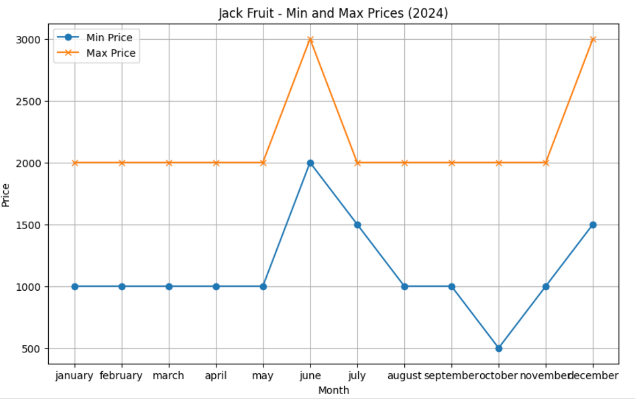


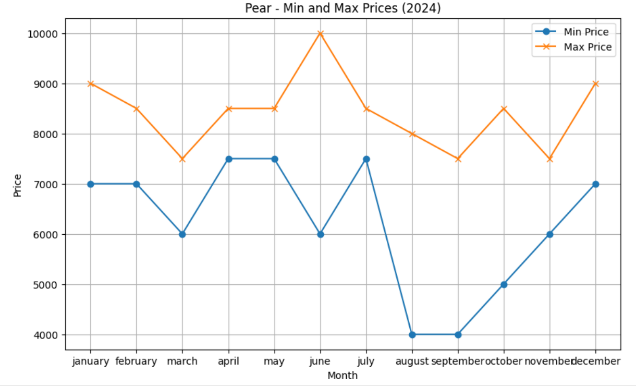
Fig.11. Comparative analysis- Theil’s U Statistics(TUS)

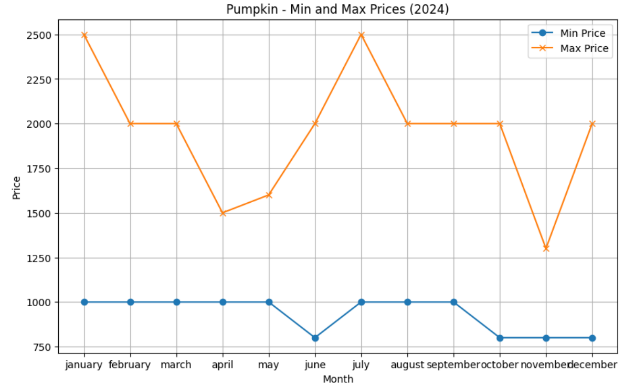
Recommendation: considering overall performance, Decision Tree and Random Forest Algorithm strikes a good balance across most metrics.

The following figures (fig. 12, fig. 13, fig. 14, fig. 15) are minimum and maximum price prediction graphs for the year 2024 for different fruits month-wise by the trained Decision Tree model.

fig.12. price prediction of orange for year 2024

fig.13. price prediction of Jack Fruit for year 2024

fig.14. price prediction of Pear for year 2024

fig.15. price prediction of Pumpkin for year 2024

# V.CONCLUSION

In conclusion, using existing historical pricing data, market prices for agricultural products can be tracked and predicted. Farmers can choose which crop to plant in which season based on the climate in their area and the conditions that work best for that particular crop. Farmers will be able to plan strategically for higher profitability by having a comprehensive understanding of market prices and price fluctuations. After evaluating and contrasting the outcomes of these algorithms on metrics like R2 score, MSE, MAE, MPE, MAPE, Huber Loss, MSLG and Theil's Statistics, it has been determined that the Decision Tree Algorithm among the four implemented methods is the best appropriate algorithm. As a result, Decision Tree classifiers can be used to estimate market prices through machine learning. The minimum and maximum prices of 14 fruits in the Pune market are predicted for the year 2024 using a Decision Tree Classifier model that was trained on three years of market pricing data from 2021, 2022 and 2023.For future research, while this method is only being utilized for 14 fruits in the Pune market, it can be used and implemented for market price forecasting for all other agricultural commodities pricing in future research. To maximize profit, forecasts using this method can also take into account the import and export prices of agricultural items. In order to predict the most accurate outcomes, market price forecasting can also be integrated with weather forecasts, soil qualities, irrigation supplies, the effect of area festivals on pricing and the impact of foreign affairs on prices.

# VI.REFERENCES

|  |  |
| --- | --- |
| [1] | C. S. R. M. M. B. and P. M. , "Price Prediction Model of fruits, Vegetables and Pulses according to Weather," in *13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2023. |
| [2] | . S. R. S. K. A. J. A. S. and R. K. M. , "Commodities Price Prediction using Various ML Techniques," in *2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 2022. |
| [3] | K. R. P. V. B. D. P. S. s. and K. , "An Enhanced Approach For Crop Yield Prediction System Using Linear Support Vector Machine Model," in *International Conference on Communication, Computing and Internet of Things (IC3IoT)*, 2022. |
| [4] | A. M. J. and S. P. , "A CNN – Bidirectional LSTM Approach for Price Forecasting of Agriculture commodities in Gujarat," in *Proceedings of the International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2022)*, 2022. |
| [5] | . V. T. . K. D. and D. S. , "Predictive Analysis of Market trends in Agriculture using ML/AI Techniques," in *IEEE Applied Sensing Conference (APSCON)*, 2024. |
| [6] | S. S. A. D. S. R. N. V. K. and S. R. , "Enhanced Crop Price Prediction & Forecasting System," in *International Conference on Computer, Power and Communications (ICCPC)*, 2022. |
| [7] | . M. M. . M. . M. S. and M. , "Demand Prediction of Agricultural Crops using Artificial Intelligence," in *International Conference on Automation and Computation (AUTOCOM)*, 2024. |
| [8] | Y. S. . M. D. M. S. . P. A. C. P. and D. S. S. , "AGRI-PRO: Crop, Fertilizer and Market Place Recommender for Farmers using Machine Learning Algorithms," in *7th International Conference on Trends in Electronics and Informatics (ICOEI)*, 2023. |
| [9] | G. T. and . M. G. , "A Methodology for Crop Price Prediction Using A Methodology for Crop Price Prediction Using," in *IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, 2022. |
| [10] | . J. H. and . D. V. , "A Comparative study of Data Mining Techniques for Agriculture Crop Price Prediction," in *IEEE 7th International conference for Convergence in Technology (I2CT)*, 2022. |

[11] S. . S. M. P. and G. C. , "Smart Cultivation and Prediction System for Agriculture," International Journal of Grid and Distributed Computing, vol. 13, January 2020.

[12] M. P. . S. S. . L. T. . K. C. and P. Y. , "Recognition of Diseases on Horticultural Leaves Using Machine Learning Approach," In *3rd Asian Conference on Innovation in Technology (ASIANCON)*, August 2023.

[13] M.S.R. .A.S. .N.V.S.R. and D.U.A., "Crop prediction using machine learning," In *Journal of Phisics: Conference Series,* 2022.

[14] P.S. .P.D. .N.A. and S.A. , "Predicting agriculture yields based on machine learning using regression and deep learning," In IEEE Access (volume 11), October 2023.

[15] M.A. .I.K. .A.A. .M.U.T. .H.K. and A.G. , "Accurate wheat yield prediction using machine lerning and climate-NDVI data fusion," In *IEEE Access, vol. 12, pp. 40947-40961, 2024, doi: 10. 1109/ACCESS.2024.3376735*, March 2024.