Crop Price Prediction System using LightGBM for markets in Madhya Pradesh State

Puneet L.S[*], Akash V[*], Tanvi B[*], Roshan P[**] Dr.Thiyagu Priyadarsan M.R[***]

[*] Student, Department of Electrical and Electronics Engineering, VIT Bhopal
[**] Student, Department of Computer Science Engineering, VIT Bhopal
[***] Assistant Professor, Department of Electrical and Electronics Engineering, VIT Bhopal

Abstract- The agricultural sector in Madhya Pradesh holds a position of prominence as the dominant economic activity. However, a prevailing issue persists whereby farmers lack access to accurate and up-to-date information regarding the optimal pricing of their crop. Machine learning, coupled with prediction models, has garnered widespread acclaim due to its impressive outcomes and performance. This work presents a system which uses machine learning algorithms along with a web portal for farmers to help them know the right prices of their crop in order to earn profits. The price is predicted just by entering the crop, market and district name. This way, the power of technology can be used to make their lives easier. The complete research comes up to a conclusion that LightGBM is the suitable technique for our project.

Index Terms- Decision Tree, Flask, LightGBM, Random Forest

I. INTRODUCTION

Optimizing profitability with limited land resources is the objective of agricultural planning in agro-centric countries. Farmers face immense challenges when their crops fetch low prices, jeopardizing their economic stability and sustainability. They frequently fall victim to the deceitful practices of middlemen and traders, who exploit their lack of market knowledge to offer unfair prices. There are multiple ways to increase the profits of farmers by helping them predict the price of a particular crop. Traditionally, farmers relied on their past experiences and knowledge of a particular crop to estimate the price they would receive. This method of price prediction was based on the assumption that historical trends and patterns would repeat themselves. This approach, although subjective and based on individual experiences, provided farmers with a rough idea of the potential returns on their agricultural produce. However, it is important to note that this method had limitations, as it did not account for broader market trends, price fluctuations, or external factors that could significantly impact crop prices. Keeping these things in mind, a system is developed which will help farmers predict the price of their crop based on the crop, market and district name. This will help them maximize their gain by selling the specific crop in the best market available. The price forecast relies on data acquired from governmental entities and employs advanced machine learning algorithms. The platform “eMandi” has the potential to be advantageous for other agriculture-dependent sectors by facilitating access to bulk markets where the particular crop is traded.

II. IDENTIFY, RESEARCH AND COLLECT IDEA

Our system development considered various approaches that have been employed to enhance the financial yield of agricultural produce. We examined and analyzed several notable systems as part of our research. [1] This paper proposes that due to drastic changes in climate in recent years, the crops prices have faced variations and hence farmers have suffered the most due to this. Hence this paper suggests price and profit prediction models using machine learning.For price prediction, Naïve Bayes algorithm is used whereas for profit prediction K Nearest Neighbour is used. [2] This paper proposes crop price prediction models for farmers using machine learning tools like Logistic Regression, Decision Trees, XGBoost, Neural Nets, and Clustering to identify the pattern among data and then process it. An additional feature used here is Exploratory data Analysis (EDA). It is an approach in which data is understood more keenly using scatter plots, bar graphs etc. This provides a better understanding of trends in the data. XGBoost algorithm is used as it has predicted better targets than all other algorithms. [3] This paper proposes prediction for farmers to plan their next crop to be grown and avoid hyperinflation. The dataset used has over 330 different crops altogether. The results have shown that Random Forest Regressor and Decision Tree Regressor have the best prediction model with 99% accuracy. [4] This paper basically focuses on predicting the yield of the crop using machine learning algorithms. Random Forest Algorithm has been used here which has parameters to predict values - District Name, Season, Crop, Area in Hectare. 10-Fold Cross Validation technique has been used to improve the accuracy of the model to 87%. An application called Smart Farm has been used to provide a user interface to farmers. [5] This paper proposes a system that integrates data that are obtained from past prediction and current prices of the crop and to give farmers an idea of the best crop to be grown. An additional feature of Rainfall consideration has been taken here. A python and Flask based web page is developed where data is stored with the help of MySQL. The model has used Random Forest and Decision Tree Regressor for forecasting prices.[6] This paper evaluates the accuracy of Light Gradient Boosted Machine (LightGBM) against the previously known Random Forest & Support Vector Machines(SVM) for classifying land use and land cover over a geographical area. It has been found that LightGBM is more accurate with 0.01 and 0.059 increase in the overall accuracy when compared to SVM and Random

Forest. Also, LightGBM was 25% quicker on average. The process time for LightGBM, SVM and Random Forest were 287s, 367s and 410s respectively. The space occupation was 8 CPU and 15 GB RAM for each of them.

III. METHODOLOGY

The objective of our project is to address the crop price prediction problem in a more effective manner, ensuring the benefit of the impoverished farmers. To accomplish this, we have utilized Python programming language within the Jupyter Notebook environment to implement the machine learning model, while Flask has been utilized as the backend to facilitate website deployment.

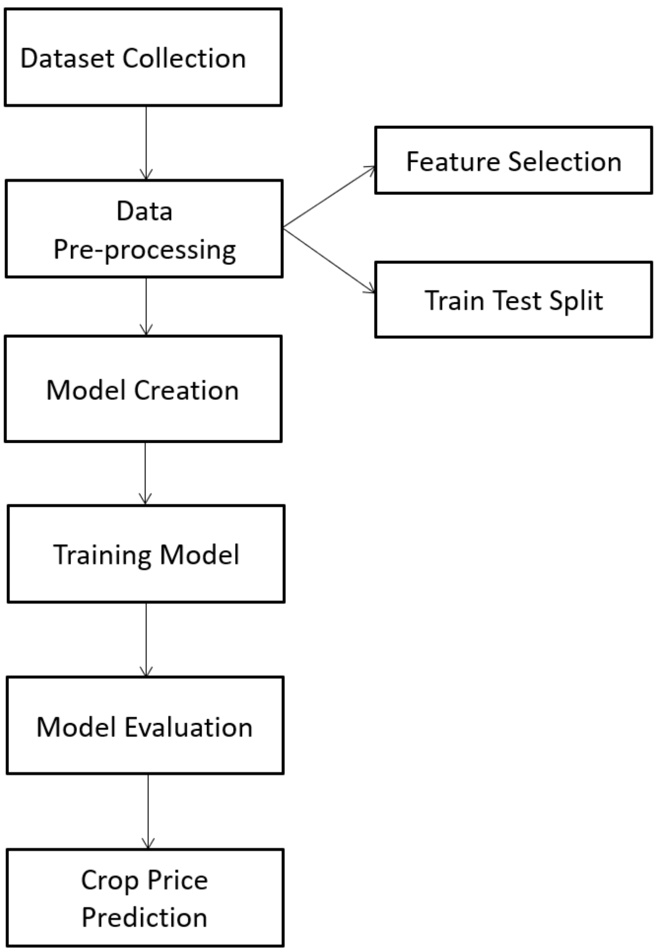

Fig 1. Block Diagram of the Methodology

Data Collection: To gather the requisite data for our project, we have collated the dataset of crops spanning from the year 2020 to 2022, which has been obtained from armarknet.gov.in. This platform is maintained by the Directorate of Marketing & Inspection (DMI), Ministry of Agriculture and Farmers Welfare and is an initiative of the Government of India. We have chosen to work with only major crops, such as Arhar, Cotton, Kodo, Lentils, Linseed, Maize, Mustard, Peas, and Wheat, to train the machine learning model.

Data preprocessing: Data preprocessing involves converting raw data into a practical and effective format, and has been employed as a method for data mining. The initial dataset contained various feature columns, including 'Sl no.', 'District Name', 'Market Name', 'Commodity', 'Variety', 'Grade', 'Min Price (Rs./Quintal)', 'Max Price (Rs./Quintal)', 'Modal Price (Rs./Quintal)', and 'Price Date'. However, only 'Market Name', 'Commodity', 'Modal Price (Rs./Quintal)', and 'Price Date' were deemed relevant as features to train the model with.

As the machine learning model only accepts numerical values, certain features, such as 'Commodity' and 'Market Name', have been encoded using Scikit-learn's LabelEncoder Function. Further, the 'Modal Price' feature has been normalized utilizing Scikit-learn's MinMaxScaler Function to ensure that the price values lie within the range of (0,1). Additionally, the 'Date', which was initially in the YYYY/MM/DD format, has been converted to Unix format to enable easy interpretation by the model.

Finally, these features have been stacked column-wise in a numpy array, which has been inputted into the train_test_split function of the Scikit-learn library. This function has been utilized to split the dataset into the training set and validation set in an 80:20 ratio, respectively.

Building and training models:

Random Forest Regressor: The Random Forest algorithm is a machine learning technique that operates on the principle that an ensemble of classifiers generated through bootstrap aggregation outperforms a single classifier. The algorithm creates multiple decision trees, each of which is trained on a randomly selected subset of observations from the training dataset. After conducting a thorough literature survey, it was noted that Decision Tree Regression and Random Forest Regression were the only models used for crop price prediction problems. However, it was found that the Random Forest model was more efficient and accurate than other models used. Consequently, we attempted to reproduce the Random Forest Model and incorporated the dataset gathered by us.

Final Parameters:

```
params: {
        max_depth: 500,
        random_state=2
        }
```

LightGBM Regressor: The Gradient Boosting Machine (GBM) enhances the performance of a set of basic decision trees by optimizing a loss function to produce more precise predictions. These decision trees, which are relatively simple and not highly optimized, are incrementally constructed to form an additive model. To achieve this goal, we constructed a LightGBM Model and utilized the same dataset that was employed in the Random Forest Model.

Final Parameters:

```
params: {
num_leaves: 50,
n_estimators: 2000,
objective: regression,
learning_rate: 0.05,
}
```

RESULTS: Upon comparison of the two models, the r2_score of the LightGBM model was found to be 98.10, whereas the Random Forest Model had an r2_score of 97.66.

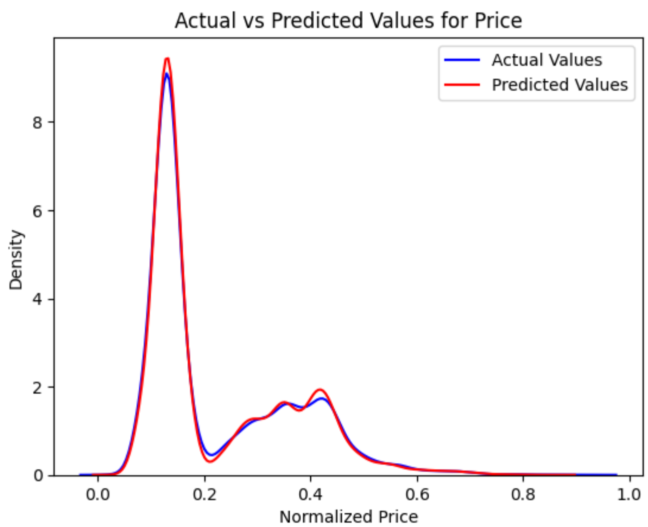

Fig 2. Actual vs. Predicted Values Graph for Random Forest Model

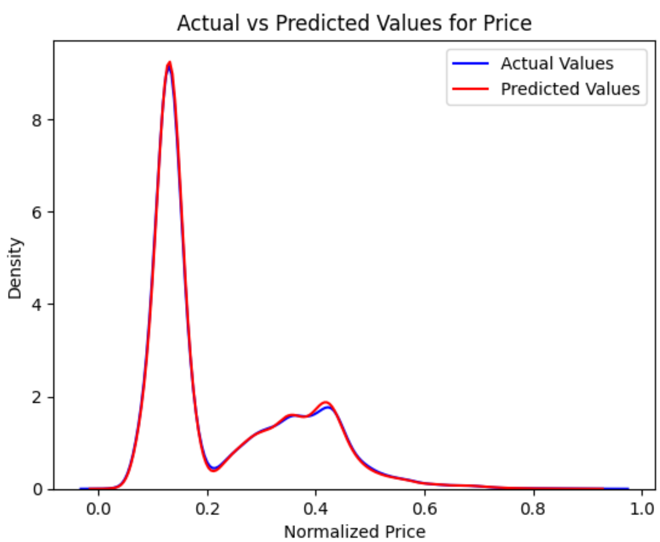

Fig 3. Actual vs. Predicted Values Graph for LightGBM Model

While the accuracies of both models were relatively similar, the training time for each varied greatly. The Random Forest Model required 32 seconds for training, while the LightGBM model was trained in just 7.5 seconds.

DEPLOYMENT: Our team has developed a user-friendly website called Emandi that aims to provide easy access to crop price predictions for farmers in India. This website is designed to be simple and intuitive, with static files for CSS, templates for HTML pages, and a pickle file containing the trained machine learning model. These various components are integrated using the Flask micro web framework, which allows web servers to pass requests to web applications or frameworks.

To ensure that the website is accessible to all farmers, we have scripted it in Hindi, as it is the most commonly used language in India. By navigating through the website in Hindi, farmers can easily fill in the required fields such as market name, crop type, and district name. Flask then feeds these entries to the machine learning model, and when the user requests a forecast, the output is rendered and displayed.

One of the key advantages of our website is that it is designed to be farmer-friendly, which is particularly important given that not all farmers in India are highly literate. By providing a user-friendly interface in a language that is familiar to them, farmers can avoid the frauds and misleads they may face during the sale of seasonal crops. The predictions provided by the website can be used as a reliable guide for pricing crops, and farmers can make informed decisions based on accurate market forecasts.

Fig 4. Website preview

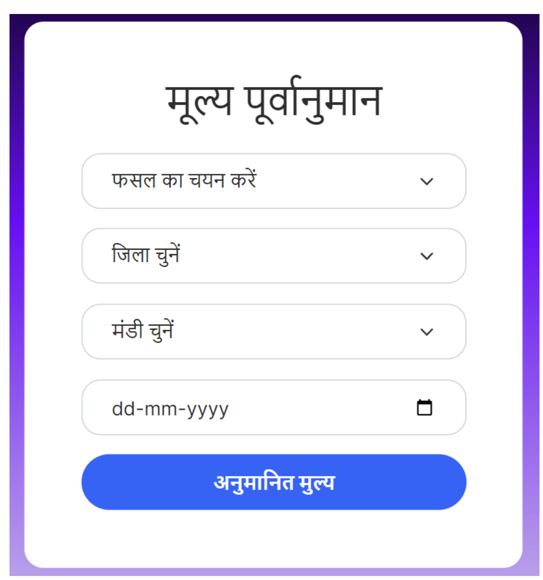

Fig 5. Close-look of the options available

IV. CONCLUSION

The present study addresses the pressing issue of crop price prediction with the primary objective of estimating the profit that a farmer can expect for a given crop even before sowing it. To this end, we have conducted an experiment comparing the Random Forest model with a novel LightGBM model. Our findings demonstrate that the LightGBM model is not only more accurate but also significantly faster than the Random Forest model, making it a more efficient tool for crop price prediction.

The potential applications of this model are extensive, particularly in the context of supporting the farming community. Collaborating with non-governmental organizations (NGOs) and government agencies would enable us to disseminate the tool more widely, making it accessible in rural areas and at the village panchayat level. Such collaborations would expand the tool's reach and impact, enabling farmers to access valuable information and resources that can help them make informed decisions and improve their agricultural practices.

By leveraging the capabilities of this tool, we can support farmers in addressing various challenges related to farming and assist them in making the right decisions regarding crop selection. Therefore, we propose that this model could be a valuable resource for the farming community and contribute to the overall development of agriculture in India.

ACKNOWLEDGEMENT

We are immensely grateful for the opportunities provided to us by VIT Bhopal University to showcase our abilities and put our skills into practice. The university's facilities and resources have been instrumental in enabling us to develop and execute our project successfully.

We would also like to thank our project guide for their constant guidance, support, and encouragement throughout the project. Their insights and feedback have been invaluable in shaping our project and helping us overcome challenges along the way.

We hope that our project will not only meet the expectations of our university but also contribute to the larger goal of promoting digitalization in the agriculture sector. We are proud of our work and grateful for the opportunity to have undertaken such a meaningful project.

REFERENCES

[1] Rachana P S, Rashmi G, Shravani D, Shruthi N, and Seema Kousar R, "CROP PRICE FORECASTING SYSTEM USING SUPERVISED MACHINE LEARNING ALGORITHMS," International Research Journal of Engineering and Technology (IRJET), vol. 06, no. 04, pp. 4805–4807, Apr. 2019, [Online].

[2] Pandit Samuel, etal."Crop Price Prediction System using Machine Learning Algorithms." Quest Journals Journal of Software Engineering And Simulation, Vol. 06, No. 01, 2020, Pp. 14-20

[3] Brunda, N. L, and A. Khan, "Crop Price prediction using Random Forest and Decision Tree Regression," *International Research Journal of Engineering and Technology (IRJET)*, vol. 07, no. 09, pp. 235–237, Sep. 2020, [Online].

[4] K. Moraye, A. Pavate, S. Nikam, and S. B. Thakkar, "Crop Yield Prediction Using Random Forest Algorithm for Major Cities in Maharashtra State," *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)*, vol. 9, no. 2, pp. 40–44, Mar. 2021

[5] G. S. Kakaraparthi and P. Rao, "CROP PRICE PREDICTION USING MACHINE LEARNING," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 03, no. 06, pp. 3477–3481, Jun. 2021, [Online].

[6] D. A. McCarty, H. J. Kim, and H. S. Lee, "Evaluation of Light Gradient Boosted Machine Learning Technique in Large Scale Land Use and Land Cover Classification," *Environments*, vol. 7, no. 10, p. 84, Oct. 2020

[7] K, Sriram. (2019). A Survey on Crop Prediction using Machine Learning Approach. International Journal for Research in Applied Science and Engineering Technology. 7. 3231-3234. 10.22214/ijraset.2019.4542.

[8] Priya P, Muthaiah U, Balamurugan M. (2018). Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Sciences and Research Technology. 7(4), 1-7, ISSN: 2277-9655

AUTHORS

First Author – Akash Verma, Student, VIT Bhopal, akash.verma2020@vitbhopal.ac.in

Second Author – Tanvi Bansal, Student, VIT Bhopal, tanvi.bansal2020@vitbhopal.ac.in

Third Author – Roshan Patidar, Student, VIT Bhopal, roshan.patidar2020@vitbhopal.ac.in

Fourth Author – Dr. M.R. Thiyagupriyadharsan, thiyagu.priyadharsan@vitbhopal.ac.in

Correspondence Author – Lamamuri Puneet Saravanan, Student, VIT Bhopal, lamamuri.puneet2020@vitbhopal.ac.in, +916383440851