An efficient crop suggesting system using machine learning models

by

Md Saif Mokarrom
20301121
Md Anonto Shuvo
23141036
Nazmul Hasan Oyon
20101528
Rifha Hossain Munaja
20301466
Soumik Roy
20101573

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

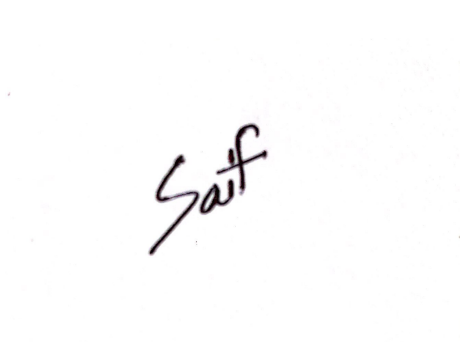
Department of Computer Science and Engineering
Brac University
January 2023

Declaration

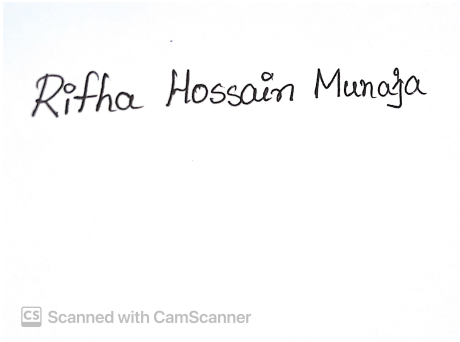
It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:



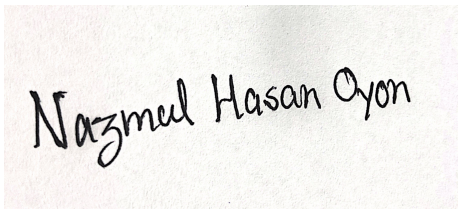
Md Saif Mokarrom
20301121



Rifha Hossain Munaja
20301466

Md Anonto Shuvo
23141036

Soumik Roy
20101573



Nazmul Hasan Oyon
20101528

Approval

The thesis titled "An efficient crop suggesting system using machine learning models" submitted by

1. Md Saif Mokarrom (20301121)
2. Md Anonto Shuvo (23141036)
3. Rifha Hossain Munaja (20301466)
4. Soumik Roy (20101573)
5. Nazmul Hasan Oyon (20101528)

Of Fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on August 23, 2015. **Examining**

Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam
Assistant Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Arif Shakil
Lecturer
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Table of Contents

Abstract

As an agricultural country, a huge economy depends on the yearly agricultural production of Bangladesh. However, the agro production depends on multiple factors such as loamyness of the soil, soil nutrients, weather conditions and amount of cultivable land. Negligence and lack of utilization of these factors often result in less agro production. Our research proposes an efficient system of suggesting a list of crops using Machine Learning based on these factors. In this research, we are using NPK values as a factor. Moreover, we will offer production rate per unit area unit and price according to farmers' financial status. Our study focuses on to help farmers to know and utilize these factors and cultivate crops in a systematic way that is beneficial economically with the help of Machine Learning. We are going to use Ensemble Model Technique with models like Support Vector Machine(SVM), Random Forest, Naive Bayes. K-Nearest Neighbour (KNN) will also be studied. Then we are going to construct a strong classifier using Adaboost. Furthermore, Explainable AI will be used to explain the ensemble model. Also implement a frontend to take parameters as inputs. Farmer's financial situation is also taken into account to suggest a series of cultivable crops.

Keywords: NPK value; Soil nutrition; Machine learning; Prediction; Production rate; Ensemble Framework; Explainable AI

Chapter 1

Introduction

1.1 Background

The backbone of Bangladesh's economy is agriculture. Over the past years it's agriculture has seen major changes and development. According to the FAO, most Bangladesh's land about seventy percent is used for cultivating crops and 80-90 percent population works mainly in agriculture. The massive domination of agriculture in Bangladesh, illustrated by the fact that in our country almost 3 out of 4 people in the nation are employed in agriculture, is the most remarkable aspect of the country's economic existence . Practically most of our food grains are produced by agriculture, which also produces larger quantities of raw materials like cotton, jute, oil and seeds are the main production sectors. The development of the national economy of Bangladesh is influenced by the agricultural sector. Information from the past years' analysis provides a thorough understanding of the phenomenon. Agriculture has been making less contribution to Bangladesh's economy each day.
Although Bangladesh's economy mostly depends on agriculture, there are so many things to look at. But it is carried out in poor conditions, and as a result, the average yield per acre of the various crops is substantially lower than in countries where agriculture is properly organized. Besides, agricultural agriculture in Bangladesh has consistently increasingly faced the hazards of climate change, including flood, drought, and salt. In addition, ineffective management techniques, particularly those relating to insects and diseases, fertilizer, water, and irrigation, have significantly decreased agricultural production. Small and small-scale farmers, who make up the bulk of farmers, are restricted by limited financial resources and are unable to pay the high management expenses related to high-input technologies.
Farmers use their previous knowledge to plan the agricultural process. They end up growing undesired crops as a result of their lack of precision farming knowledge. Farmers will have a long-term advantage since they may anticipate healthier crop development, increased yield, and higher revenue if they cultivate their crops while taking into account their financial situation, soil characteristics, and facilities that are accessible. Therefore, farmers' well-being is secured if they are led in the proper path utilizing new technology, and national development is also possible. The proper composition of each parameter in the soil, which includes a number of components or parameters, will help the growth of plants. In agriculture Machine Learning is an important model which helps to achieve specific results by using data and algorithms.Nitrogen (N) is most useful for various types of crops as it is a primary

and essential source of protein. Phosphorus is the main element of enzymes for the energy conversion process in various types of plants by restricting plant growth, phosphorus deficiency lowers growth and yield. Moreover, Potassium (K) limits plant growth and reduces production.

1.2 Problem Statement

The population of the world is increasing day by day whereas the cultivable lands are the same throughout the years. The farmers have to cultivate more crops every year to meet the needs of the large population. And so the cultivable lands are being used for cultivation crops throughout the year. As a result the soil moisture of the land is decreasing. Moreover, less time is now given for the land to recover from cultivating one crop to another. Thus the land is unable to recover its moisture, fertility after each cultivation. Again to earn more profit, the farmers are cultivating the same type of crops on the same land over and over decreasing the land's soil fertility and properties. Thus, next crops can not grow properly or in good condition and amount due to lack of soil nutrients. Furthermore, nowadays, the weather is unpredictable due to global climate change. Untimely drought and rain both scenarios are seen frequently nowadays. The farmers can not early predict the weather like before and thus the crops end up damaged due this sudden change in the weather. Due to such various factors and continuous cultivation, the agricultural soil is getting deteriorated and agro production is decreasing which is a bigger threat to the farmers as well as the future of agriculture. Moreover, budget is also an issue for them while choosing crops.
In this paper, we are going to address this issue of inefficient crop selection for land. We want to automate the process of crop selection. For that, we not only consider the soil nutrients and previous production rate data which are the most important factors for crops to grow but also we want to take care of the economic factor of a farmer while suggesting the crops. So, to deal with this issue, Machine learning algorithms can help by predicting crops based on the appropriate dataset so that farmers can cultivate cost effective crops and get benefit from our research.

1.3 Research Objective

The research aims to use Machine learning models to suggest crops to the farmers based upon the land's NPK(Nitrogen, phosphorus and Potassium) value, land area, weather conditions and farmers' financial economy. The objectives of the research are -

1. Soil nutrients availability and need for cultivating different crops on the lands.
2. Predicting list of crops for the land that are profitable based upon the farmers financial condition.
3. To explore various algorithms for attaining better accuracy at predicting the suitable crops in different conditions.
4. Enhancement of Machine learning algorithms by applying hypertuning methods on models.

5. To build a front end side so that non tech people can easily benefit from our study. This application will take region, economic type as an input and get a result of a list of crops for that region.

Chapter 2

Related Work

In this paper, Schwalbert et al. conducted a study [1] to predict soybean yield based on satellite and weather data (Vegetation indices like NDVI, EVI and temperature) in Brazil. It showed a comparison between different algorithms and concluded that LSTM neural networks works well than Random forest and OLS Linear Regression. Though this research was conducted in southern Brazil, it can easily be implemented in other regions if sufficient data is available. Since satellite data are required to conduct this study, adverse weather conditions can affect our research poorly.
In this study, Jain et al. [2] , proposed a crop sequencing model to contribute a sequence of crops for a region. This model takes into account various environmental factors such as soil type, climate type, temperature, ph value to select appropriate crops using WEKA classifiers and regression models. To provide a sequence of crops for a particular season authors used a factor value(multiplication of predicted yield rate and price). Crop diseases, effect of one crop on another crop are important to determine soil fertility which was not taken into consideration in this study.
In This paper, Kulkarni et al. [3] represents a crop suggestion model to increase crop production. Soil type such as NPK values, PH, temperature, rainfall used as independent variables to build this model. Authors have used ensemble techniques. Various base learning concepts such as SVM, RF, Naive bayes were used and by combining, it achieved more than 99 percent of accuracy. Although in this study, authors take into account only four crops, it can be applicable for more crops and for wider regions.
In this research paper, Ramesh and Vardhan conducted a study [4] on predicting crop yield rate using data mining technology. This study used data from 1965 to 2009 of a district of Eastern India. To estimate the data four variables were used such as year, sowing area, rainfall, production. Clustering was done based on the 'rainfall' attribute. Multiple Linear Regression (MLR) showed an accuracy of 98 percent whereas K- means algorithm gave 96 percent accuracy. Soil quality such as NPK, other micro and macro nutrients, play a critical role in estimating crop output rates. So, there is a scope to take into consideration all of these factors to get a viable result.
In this research paper, Shakoor et al. [5] have used Decision Tree Learning - ID3 (Iterative Dichotomiser 3) and KNN Regression algorithms to predict the crops. They have worked on 6 crops over 10 regions and have used temperature, rainfall, crops per hectare as parameters. For training the models, they have used 12 years of data from every region. For some crops, Decision Tree shows better prediction

and for others KNN shows better prediction. Their results show that Decision Tree Learning- ID3 algorithm performs better in comparison with KNN in terms of the error percentage of the crop yield rate predictions.
In this research paper, Jha et al. have studied [6] in the agro based sector, the authors have studied about various experts and wireless systems employed in these regions. A technique was suggested by the authors for the identification of flowers and leaves as well as watering using IOT. Deep learning and other advanced methods are utilized to identify the type of crop assisting farmers in creating an environment that will help the plant grow. Use of Thermal Images, can give farmers real time data about the water quantity in the lands.
In this paper, Kalimuthu et al. [7] have used the supervised learning algorithm Naive Bayes to predict the crops. They had designed an Arduino Uno Board with DHT11 sensor moisture sensor for collecting data of temperature, humidity and soil moisture over a month along with previous collected data. With Naive Bayes the accuracy was 97% and to improve the accuracy, they had used a boosting method along with the Naive Baise method. They also developed an application where after giving parameters values like temperature, humidity, soil moisture and pH, a crop will be suggested. There is scope to improve using source of sunlight and crop health into account to suggest fertilization, crop and the land.
In this pape,r Nischitha et al. [8] has used the SVM algorithm to predict the rainfall based on the previous collected data and Decision tree algorithm for the crop prediction. As perimeter, pH level, temperature and humidity were taken. The preprocessing rainfall data is trained by an SVM classifier with Radial basis function(RBF) kernel. Their model shows the predicted crops along with the required NPK value, seed for cultivation (kg/ha), approximated yield (quantity/acre) and market price(Rs/quantity).
In this paper, Kumar et al. [9] has proposed a model called Crop Selection Method (CSM) that will solve the problem of crop selection by maximizing net yield rates across seasons and generate the greatest possible economic growth for the nation. This model determines a list of crops whose output per day is at its highest during the growing season using the crop's sowing date, plantation days, and estimated yield rate as input. The CSM method's effectiveness and accuracy depend on the anticipated value of impacted variables, hence it's necessary to take a prediction technique with higher accuracy performance.
The capability of farmers to predict which crop can be grown in a certain soil type is greatly helped by the classification of soil according to soil nutrients, [10] presents a system for categorizing soils according to their macro- and micronutrient contents and for forecasting the kinds of crops that can be grown in each type of soil. Machine learning KNN, Bagged tree, SVM, and logistic regression are implemented in the paper. The proposed system involves two phases: the training and testing phases. It uses two databases: the soil database based on soil parameters and the crop database. Chemical and geographical features are included in the soil database. On the other hand, the crop database is based on time and production rate. The current model is the most accurate of the ones currently in use. In the future, proper fertilizers are advised for the cultivation of crops that grow well. The future models incorporate real-time data that is directly received from agricultural land that is connected to sensors, whereas the current models deal with old data that is currently available. The sensors detect other soil minerals and the fertility of the

soil.
In this paper, N. Choudhary stated [11] that nutrients like Nitrogen, Phosphorus, Potassium, Zinc, Iron are important for the growth of the crops. Moreover, weather conditions, temperatures, soil pH level, also have effects in the growth of the crops. Here artificial intelligence uses soil parameters that are mentioned above to predict the fertility rate. After that, it will recommend to the farmer the crops that will take less time and less cost. With the help of deep learning neural networks, many machine learning algorithms like classification and regression are performed then will predict the fertility rate and later on suggest the right crop. After implementing all the models, then it has been found that deep learning gives 87 percent accuracy because of its self-learning ability and robustness.
In this paper Rajak et al. stated [12] in his proposed model that farmers will get the recommendation of crops based on soil in their lands. For this, SVM and ANN algorithm has been used in this model. The dataset has been collected from soil laboratories, and crops like groundnut, banana, vegetables, and sugarcane are used for the dataset. After implementation, it has been found that to increase productivity in agriculture, gives better accuracy with the right crop suggestions as a result it prevents soil degradation. Furthermore, it also utilizes the water resources more efficiently.
In this study, Akshatha Shreedhara stated [13] in their paper about precision agriculture. It is a technique that suggests crop to the farmers by their land. It is based on both soil characteristics of the land and also farmers' financial condition. Random tree, KNN and naive bayes algorithm used here. After the implementation, it's been able to predict the crop with high accuracy and efficiency.
In this findings Authors have used Machine Learning Algorithms Random Forest,Naive Bayes for precision farming.ML algorithm model is generated for prediction after preprocessing is done Anguraj et al. [14] .Applying the techniques in this suggested system ,the shortcomings of the current system are resolved. These ways include improving agricultural yield, real-time IOT crop analysis, and selecting necessary. The main purpose for a crop suggestion model is to approve the required crop for a specific field by applying this system crop loss will be prevent.Suggesting crops accuracy levels may vary depending on which algorithm with specific attributes is picked and applied. Selecting parameters, making smarter decisions and getting better yield.The model then accurately predicts and recommends the crops to be planted with a 96.89 percent accuracy rate.
This system helps the farmers to select suitable crop based on season and region of sowing Vanathi et al. [15] . Different datasets of weather parameters like rainfall, temperature, humidity, soil parameters and geography of a place are collected and then analyzed. This dataset included 5 environmental parameters, three biotic parameters and two related variables to find out different regions crop production. By integrating this system with other sectors that contribute to the development of a nation's agriculture, such as horticulture, sericulture, and many others, the system can be improved. In order to form a harvest for the site-distinct parameters, it proposed a recommendation system through an ensemble model employing KNN, Polynomial regression, Random forest tree, Decision Tree, and Naive Bayes as learners. The effectiveness of crop variety in reducing their unfavorable effect is also evaluated in this study, which used a lively panel data method to quantify heat stress effect and rainfall on agricultural fertility. Climate change has a significant

impact on farmers' willingness to adopt productivity-boosting inputs, which in turn affects agriculture fertility. The proposed system's prediction accuracy is 98 percent. If there is enough data, this project could be expanded to a wider area.
the main purpose of Dharani et al. [16] was to use deep learning algorithms to predict crop yields. In this research RNN-LSTM, ANN,CNN hybrid networks following observations were made.In this research Deep learning and its hybrid techniques,including deep and artificial neural networks and recurrent neural networks have been used.Artificial Intelligence is used for monitoring the crops and provides technique for anticipate production because crop production is entirely dependent on timely observation and advice . It was seen that ANN's performance in the findings obtained is between 60 and 70 percent. With its visual matrix assessment, the CNN outperforms ANN and achieves an estimation accuracy of about 87 percent where actual rate of yield was 85.1 percent as opposed to the CNN's estimated yield prediction of 86.5 percent. With an estimation accuracy of 89 percent, RNN with LSTM and the hybrid network works better than other networks. The hybrid network combines 90 percent accuracy at the highest level.
According to John et al. [17] Indicators of soil quality such as soil organic carbon (SOC) are important because they directly affect soil fertility and plant productivity. SOC also contributes significantly to the improvement of soil structure by giving nutrients to the soil.ANN,SVM, cubist regression,RF and MLR were used as machine learning techniques in this research to improve the prediction of Soil Organic Carbon. Within the research region total 60 soil samples were taken at a soil depth of 30 cm, a method named Walkley-Black was used to determine the SOC content of each sample.The mean of SOC is 1.62% and coefficient of variation is 47%. With the help of geospatial technology, soil characteristics, remote sensing data, digital elevation models (DEM), micro-climatic data, and geology, precision agriculture has now been developed to spatially vary nutrients and soil parameters within a field. Farm managers can control within-field variability with precision agriculture to increase the crop enterprise's cost-benefit ratio.Fathololoumi et al. used (RF) model and cubist models to improve digital soil mapping utilizing multitemporal remotely sensed satellite data fusion in Iran.
In this paper, Erel et al. studied an experiment [18] in the irrigation solution of fruit setting of olive trees and flowering in response to the N , P, K of the soil. In the study, it has been shown that the amount of N,P and S impacted the olive trees' flowering intensely and N and P levels had an impact on the fruit set but not K levels. Six replications of a randomized block design were employed in the experiment. The findings of this study clearly show that in the root sector of the olive plant's, with the increase in the concentration of N, P, K, the amount of each of these nutrients also increases. Furthermore, it's also found out that, for the growth and production of olive trees, the macronutrients have an important role.
In this study, Kasinathan et al. conducted a study [19] to detect insects and their classification. Wang and Xie dataset's shape attributes were used in experiments to organize the dataset's nine and 24 insect classes.Image augmentation is used since there aren't as many insect photos in the Wang and Xie databases. Different types of models are utilized here such as Artificial Neural Networks, Support Vector Machine, KNN, Naive Bayes, CNN model. The CNN model generated the highest rate of classification for both the nine and 24 classes of the insects. Its rates were 91.5 percent and 90 percent for the classes, respectively. The outcomes of classifying and

detecting various insect datasets using machine learning and an insect pest detection method were compared in this research.The results of this research is to identify the insect in an early stage and decrease the period of time taken for crop production and also improve its quality.

Chapter 3

Methodology and WorkPlan

3.1 Research Methodology

The main motive of our study is to suggest a sequence of crops for a region. In our case, data collection is a long and manual process. We have to manually create a dataset from the government's audit book. Then the dataset needs to be pre-processed by filling missing data, handling noisy data, transforming data and encoding the categorical data. After that, data is ready to be split. It will be split into an 80 : 20 ratio, which means the training dataset will contain 80 percent of the total data whereas the testing dataset will get the remaining 20 percent. Models we will be using in our research are SVM, Naive Bayes, RF, Adaboost and K-NN. Furthermore, we'll employ the ensemble framework on our models. To add on this, Explainable Artificial Intelligence will be used to explain ensemble models. Finally, we will create an application where location and economic type are given as inputs and the user will see a list of crops.

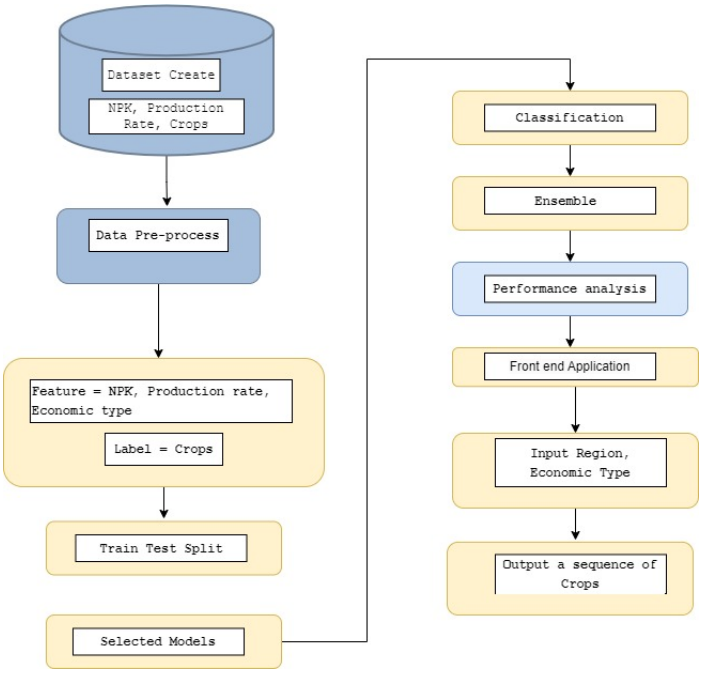


Figure 3.1: Methodology of our Research

3.2 Support Vector Machine (SVM)

SVM is a set of methods for supervised learning . Basically it is for data classification, the process of regression inspection, and finding outliers. SVM is appropriate for more than one class classification for determine of crops [10]. The SVM method generates a hyperplane by picking the vectors that will act as the best result. This is employable to make decisions and separate the data points into the categories they belong in [12] . .

3.3 Random Forest RF

Random forest can be applied to both classification and regressive problems in ML. It uses more than one classifier to handle challenging problems and enhance the performance of models.This is known as ensemble learning which means the best result will be taken. There are several decision trees in a random forest.To increase the accuracy of the dataset, the predicted average is used.RF uses each prediction made by a decision tree rather than relying just on one.It anticipates the ultimate output based on the majority of predictions.Random forest's primary aspect is that it can handle enormous datasets with high dimensionality.It prevents overfitting by increasing the model's accuracy [14] .

3.4 Naive Bayes

Naive Bayes is a widely applied classification algorithm in machine learning. It makes probabilistic prediction. Naive Bayes algorithm is easy to understand as well as implement when it comes to categorical data. In our proposed work, categorical. The 'Bayes theorem' serves as the foundation for this algorithm. It is recognized as Naive Bayes because a naive classifier assumes that every feature is independent of other features. This study [3] uses naive bayes with ensemble models for crop recommendations. Also, [14] uses naive bayes to analyze soil.

3.5 K-Nearest Neighbor (KNN)

The Algorithm KNN is non parametric , it follows sample supervised learning technique which detain all previous data and choose value for new choosen predictor.This method presume the previous cases that are comparable with the new data and it places the new cases in that category which is mostly comparable to the current categories and saves information which is accessible. It can category data points based on similarity. KNN algorithm mostly used for classification problems and this technique learns slowly because it accomplish an action on the dataset. Distance function such as Manhattan, Minkowski Euclidean distance function figure distance from the new input sample predictor to all training sample predictors and then selects the k closest distances with matching values is done by KNN and looks their class levels. KNN is mostly efficient for all ML approaches and for prediction of new data point from all data samples space complexity and time are comparable. In study from [5]. KNN has used to predict yearly production rate of crops. This can be useful for our study as well.

3.6 Adaboost

A sort of ensemble model method called ”boosting” converts weak classifiers strong. Weak models are created by using it to build a model. A model is first created with the help of training data.The second model is created so that it may correct the first model's flaws.This process is repeated until the maximum model is added or the complete training data set has been fully forecasted. AdaBoost was the first boosting algorithm which was designed for binary categorization. It is short for Adaptive Boosting and is a very common boosting method that improves weak classifiers.

3.7 Ensemble Model Technique

Ensemble model is an immensely important machine learning technique to get optimal results. It ensembles multiple base learning models to produce a best model. We implement several machine learning models to get the best predictive models. Ensemble models take into account multiple models and average of those models create an optimal model. From datasets, several subset of data is generated and each of the sub dataset is used on the models. In bagging technique, a decision tree is formed. Machine learning algorithms are run in order to find a final result. On the other hand, Random forest models are better than decision tree based models to predict with high accuracy. Ensemble technique has dependent and independent technique. In a dependent framework, one classifier is dependent on the other other framework. We are using an Independent framework in our proposed research where every classifier works parallelly. Naive bayes, RF, SVM has been used in research [3] to create an ensemble model. In our study, we are going to use ensemble learning techniques on our dataset to get a best model. Advantages of using ensemble learning in our work-

1. Compared to single models, it will have a high accuracy value.

2. As several models are used in ensemble technique, it can handle if the dataset contains both linear and non-linear data.

3.8 Workplan

Our thesis will be completed in three phases. Each phase is four months long. Firstly, In pre-thesis 1, we studied multiple research papers and chose our topic. After that, start with the abstract. We read other related publications to obtain a more in-depth understanding of this issue. We have studied related machine learning and deep learning models and decided to use SVF,RF,K-NN and AdaBoost algorithms to make our research more efficient. Following that, we focused on methodology. In the pre-thesis 2, we have to collect data from the government's survey book and we have to create a dataset of our own. Pre-processing was next carried out to help us spot any issues.The models are then applied to our dataset.In the final thesis , we carry out the implementation as planned. We analyze model efficiency and compare output results. We will be working on a interactable application for non technical

users so that they can benefit from our research. Finally, we will work on the final report according to all the information we got.

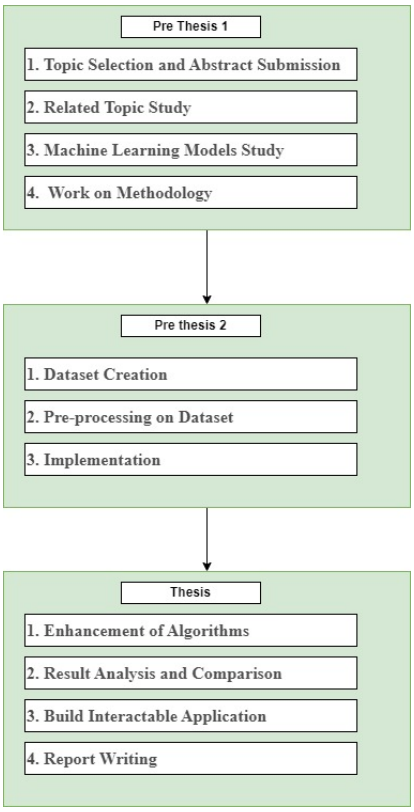


Figure 3.2: Working Plan of Our Research

Chapter 4

Conclusion

4.1 Conclusion

Agriculture continues to be the main source of income for the majority Bangladeshi people and it is the main economic sector of the country. To increase the productivity in agriculture we have to monitor the agriculture system to precision agriculture where we will collect NPK value, fertility rate of growing crops, previous year crop production of a region. After that we will implement an efficient machine learning model for recommending the suitable crop for that particular region. To get accurate results, we will be using ensemble technique. The suggested study also proposes explainable AI on the ensemble models for model explainability. Finally, we would like to present a front end application for this research. To sum up, it can be concluded that ML models with ensemble technique can be used efficiently to select a list of suitable crops for a region.

Bibliography

[1] R. A. Schwalbert, T. Amado, G. Corassa, L. P. Pott, P. V. Prasad, and I. A. Ciampitti, "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern brazil," *Agricultural and Forest Meteorology*, vol. 284, p. 107 886, 2020.

[2] N. Jain, A. Kumar, S. Garud, V. Pradhan, and P. Kulkarni, "Crop selection method based on various environmental factors using machine learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 2, pp. 1530–1533, 2017.

[3] N. H. Kulkarni, G. Srinivasan, B. Sagar, and N. Cauvery, "Improving crop productivity through a crop recommendation system using ensembling technique," in *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, IEEE, 2018, pp. 114–119.

[4] D. Ramesh, B. V. Vardhan, *et al.*, "Data mining techniques and applications to agricultural yield data," *International journal of advanced research in computer and communication engineering*, vol. 2, no. 9, pp. 3477–3480, 2013.

[5] M. T. Shakoor, K. Rahman, S. N. Rayta, and A. Chakrabarty, "Agricultural production output prediction using supervised machine learning techniques," in *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, 2017, pp. 182–187. DOI: 10.1109/NEXTCOMP.2017.8016196.

[6] K. Jha, A. Doshi, P. Patel, and M. Shah, "A comprehensive review on automation in agriculture using artificial intelligence," *Artificial Intelligence in Agriculture*, vol. 2, pp. 1–12, 2019, ISSN: 2589-7217. DOI: https://doi.org/10.1016/j.aiia.2019.05.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2589721719300182.

[7] M. Kalimuthu, P. Vaishnavi, and M. Kishore, "Crop prediction using machine learning," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, IEEE, 2020, pp. 926–932.

[8] K. Nischitha, M. N. Dhanush Vishwakarma, and M. M. Ashwini, "Crop prediction using machine learning approaches," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 08, 2020.

[9] R. Kumar, M. Singh, P. Kumar, and J. Singh, "Crop selection method to maximize crop yield rate using machine learning technique," in *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, 2015, pp. 138–145. DOI: 10.1109/ICSTM.2015.7225403.

[10] N. Saranya and A. Mythili, "Classification of soil and crop suggestion using machine learning techniques," *Int J Eng Res Technol*, vol. 9, no. 02, pp. 671–673, 2020.

[11] D. Varshitha and S. Choudhary, "An artificial intelligence solution for crop recommendation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 3, pp. 1688–1695, 2022.

[12] R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A. Devare, "Crop recommendation system to maximize crop yield using machine learning technique," *International Research Journal of Engineering and Technology*, vol. 4, no. 12, pp. 950–953, 2017.

[13] K. Akshatha and K. Shreedhara, "Implementation of machine learning algorithms for crop recommendation using precision agriculture," *International Journal of Research in Engineering, Science and Management (IJRESM)*, vol. 1, no. 6, pp. 58–60, 2018.

[14] K. Anguraj, B. Thiyaneswaran, G. Megashree, J. P. Shri, S. Navya, and J. Jayanthi, "Crop recommendation on analyzing soil using machine learning," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 6, pp. 1784–1791, 2021.

[15] A. Vanathi, S. R. Sree, and K. Swaroopa, "Ensemble techniques for effective prediction of crop selection in the coastal andhra deltaic region," in *Intelligent Systems and Machine Learning for Industry*, CRC Press, 2022, pp. 199–222.

[16] M. Dharani, R. Thamilselvan, P. Natesan, P. Kalaivaani, and S. Santhoshkumar, "Review on crop prediction using deep learning techniques," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1767, 2021, p. 012 026.

[17] K. John, I. Abraham Isong, N. Michael Kebonye, E. Okon Ayito, P. Chapman Agyeman, and S. Marcus Afu, "Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil," *Land*, vol. 9, no. 12, p. 487, 2020.

[18] R. Erel, A. Dag, A. Ben-Gal, A. Schwartz, and U. Yermiyahu, "Flowering and fruit set of olive trees in response to nitrogen, phosphorus, and potassium," *Journal of the American Society for Horticultural Science*, vol. 133, no. 5, pp. 639–647, 2008.

[19] T. Kasinathan, D. Singaraju, and S. R. Uyyala, "Insect classification and detection in field crops using modern machine learning techniques," *Information Processing in Agriculture*, vol. 8, no. 3, pp. 446–457, 2021.