Internship Report

Artificial Intelligence & Machine Learning

DLithe Consultancy Services Pvt. Ltd.

Internship Report

Trainee/Intern Name: Karthik Betageri

Reg. no:

Period: 30 Days

Job Assignment: Crop Yield Prediction Model using AI and ML

Organization: DLithe Consultancy Services Pvt. Ltd.

Supervisor's Name: Medini B V

Observations:

- The report introduces the crop yield production models that typically involves analyzing various factors affecting yields
- The various factors that are taken into consideration in crop yield prediction models are weather conditions, soil quality, crop varieties, and agricultural practices.
- The various advancements to be made while using AIML in agricultural area

Submitted to

Signature of Training Supervisor

Date:

Signature of Co-ordinator

Date:

DLithe™
Be Agile..Be in Demand

Letter of Transmittal

To,

Program Co-ordinator
DLithe Consultancy services
Bengaluru

Dear Sir,

I am writing to submit my report on Artificial Intelligence (AI) and Machine Learning (ML). The training program was an invaluable learning experience, and I am grateful for the opportunity to participate.

The training program covered various aspects of AI and ML, including basic concepts, algorithms, programming languages, and practical applications. I gained a comprehensive understanding of the role of AI and ML in modern technology and industry, and also gained hands-on experience with AI and ML tools and platforms. The training highlighted the potential of AI and ML to revolutionize various fields, including healthcare, finance, and manufacturing.

The report includes a detailed overview of the training program, including the topics covered, the learning objectives, and the outcomes achieved. It also provides observations and insights into the potential benefits and challenges of implementing AI and ML solutions in different fields.

I believe that the knowledge and skills that I acquired during the training program will be valuable to our organization. AI and ML are rapidly becoming more ubiquitous in various industries, and the ability to work with AI and ML tools and platforms will be increasingly important for our organization's success.

I hope that the report provides useful insights into the benefits of on-job training and the potential of AI and ML.

Sincerely,

Name: Karthik Betageri

DLithe™
Be Agile..Be in Demand

Table of Contents

Introduction	5
Literature survey	5
Proposed work	6
Implementation	8
Conclusion	9
References	10

1. INTRODUCTION

Agriculture plays a critical role in the global economy. With the continuing expansion of the human population understanding worldwide crop yield is central to addressing food security challenges and reducing the impacts of climate change. Crop yield prediction is an important agricultural problem. The Agricultural yield primarily depends on ***weather conditions (rain, temperature, etc)***, pesticides and accurate information about history of crop yield is an important thing for making decisions related to agricultural risk management and future predictions. The basic ingredients that sustain humans are similar

Crop prediction is one of agriculture's complex challenges. Because crop production is affected by many factors such as atmospheric conditions, type of fertilizer, soil, and seed, this challenge necessitates using several datasets. This implies that predicting agricultural productivity is not a simple process; rather, it entails a series of complicated procedures. Crop yield prediction methods can now reasonably approximate the actual yield, although more excellent yield prediction performance is still desired.

Python libraries like Pandas, NumPy, and scikit-learn are often used for data manipulation, analysis, and modeling. Techniques such as regression analysis or machine learning algorithms like Random Forests can help predict and understand crop yield variations based on input variables.

2. LITERATURE SURVEY

All the published papers and research articles cover the following points. These survey aim to identify the current state of the field, key methodologies, challenges, and potential areas for improvement. They typically includes:

1. Introduction to ALML in Agriculture: Exploring literature introducing ALML concepts and its applications in agriculture, specifically for crop yield prediction. Considering the advantages of ALML over traditional methods.

2. Existing Crop Yield Prediction Models: Identifying and reviewing existing crop yield prediction models that leverage ALML. Examine their methodologies, algorithms, and performance metrics.

3. Data Sources and Features: Investigate the types of data sources used in previous studies for training ALML models in crop yield prediction. This includes remote sensing data, weather data, soil information, and other relevant features.

4. ALML Techniques: Examining the various ALML techniques applied in crop yield prediction, such as supervised learning, unsupervised learning, reinforcement learning, and ensemble methods. Analyze their effectiveness in handling agricultural data.

5. Evaluation Metrics: Looking into the evaluation metrics used to assess the performance of ALML-based crop yield prediction models. Common metrics include accuracy, precision, recall, and F1 score.

6. Challenges and Limitations: Summarize challenges and limitations encountered in existing studies, such as data scarcity, model interpretability, and scalability issues. Identify gaps in current research that require further exploration.

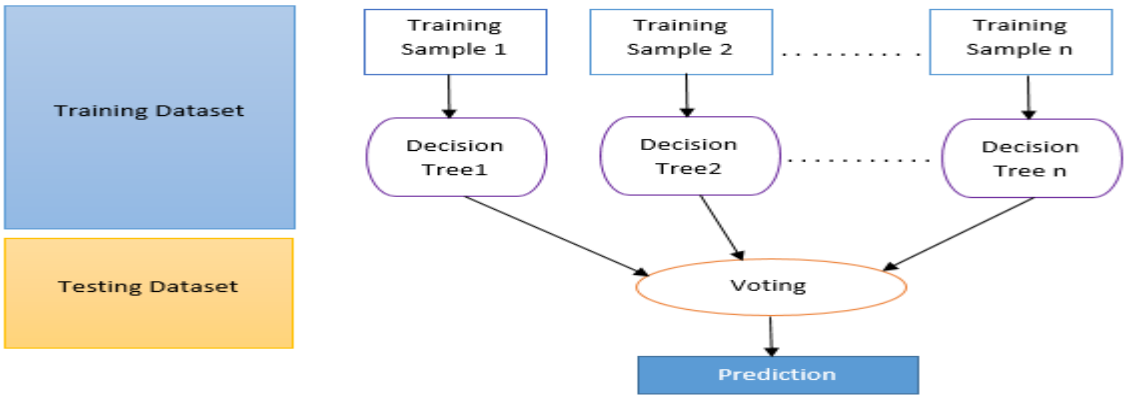
7. Integration with Precision Agriculture: Explore literature discussing the integration of ALML models with precision agriculture practices. Understand how these models contribute to optimized resource allocation and improved crop management.

8. Future Directions: Look for insights into the future directions of research in ALML for crop yield prediction. Identify potential innovations, emerging technologies, and areas where improvements can be made.

3. PROPOSED WORK

One of the major challenges for the Crop Yield project model will be the collection of input datasets for the algorithm. For conducting the test, we are using Random Forest

Classifier algorithm. The dataset for testing the algorithm is collected from FAO (Food and Agriculture Organization), World Data Bank. and agricultural websites. Various models are first tried and then the best one, that is Random Forest model is used as it is best suited for recognising the complex patterns present in agricultural datasets. Then, the collected information is fed to the Random Forest algorithm. Finally, the algorithm evaluates the data on the basis of the patterns present and displays the result. To give you an idea, here is a diagrammatical representation of the whole process we will follow for the conduction of the project.



Random Forest Classifier algorithm

Data Collection

Crop Prediction Dataset, a comprehensive collection of agricultural data designed to aid in the prediction of crop yields based on various environmental factors. This dataset serves as an invaluable resource for researchers, data scientists, and agronomists seeking to enhance their understanding of crop behavior and optimize agricultural practices.

DESIGN ALGORITHM

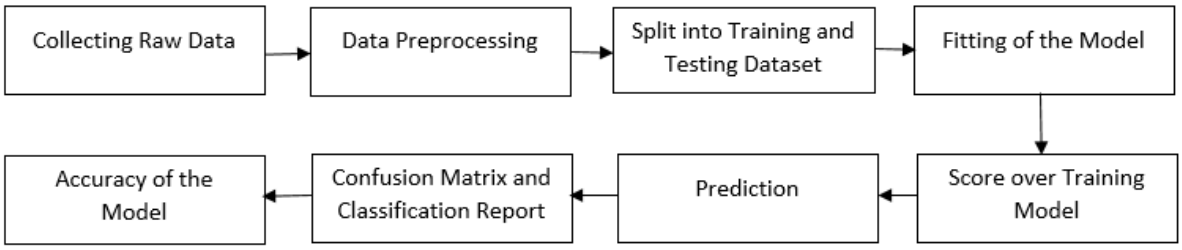
- Random Forest Classifier algorithm

The Random Forest method consists of multiple decision tree classifiers to enhance the model'sperformance. Here ensemble learning is the principle used to resolve complicated problems. It is asupervised learning algorithm. Decision trees are created at random using the instances from the training set. Each of the decision trees gives out predictions as their outcome. The final prediction for the model is decided by majority voting. One of the reasons for its popularity as a machine learning approach is that it can handle the issue of overfitting trees.

Here's how it works:

- Step 1: K instances are chosen at random from the given training dataset.
- Step 2: Decision trees are created for the chosen instances.
- Step 3: The N is selected for the number of estimators to be created.
- Step 4: Step 1 & Step 2 is repeated.
- Step 5: For the new instance, the predictions of each estimator is determined, and the category with the highest vote is assigned..

4. IMPLEMENTATION



1. Collecting the Raw Data: The practice of cumulating and scrutinizing data from different sources is known as data collection. Data collection is a way to keep track of past occurrences so that one can utilize data analysis to detect repetitive patterns.
2. Data Preprocessing: The process of modifying raw data into a form that analysts and data scientists can use in machine learning algorithms to find insights or forecast outcomes is called Data preprocessing. In this project the data processing method is to find missing values.

3. Train and Test Split: It is a process of splitting the dataset into a training dataset and testing dataset using train_test_split() method of scikit learn module. 2200 data in the dataset has been divided as 80% of a dataset into training dataset-1760 and 20% of a dataset into testing dataset-440 data.
4. Fitting the model: Modifying the model's parameters to increase accuracy is referred to as fitting. To construct a machinelearning model, an algorithm is performed on data for which the target variable is known. The model'saccuracy is determined by comparing the model's outputs to the target variable's actual, observed values
5. Checking the score over a training dataset: Scoring, often known as prediction, is the act of creating values from new input data using a trained machine learning model. Using model.score() method calculating the score of each model over a training dataset shows how well the model has learned.
6. Predicting the model: When forecasting the likelihood of a specific result, "prediction" refers to the outcome of an algorithm after it has been trained on a previous dataset and applied to new data. Predicting the model using predict() method using test feature dataset. It has given the output as an array of predicted values.
7. Confusion Matrix and Classification Report: Confusion Matrix and Classification Report are the methods imported from the metrics module in the scikit learn library are calculated using the actual labels of test datasets and predicted values. **Confusion Matrix** gives the matrix of frequency of true negatives, false negatives, true positives and false positives. **Classification Report** is a metric used for evaluating the performance of a classification algorithm's predictions.
8. Accuracy: The number of correct predictions divided by the total number of predictions is known as model accuracy. The accuracy of the model is calculated using accuracy_score() method of scikit learn metrics module.

5. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. Finally we predict the crop using machine

learning algorithm with different results. This brings some of the following insights about crop prediction. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated and lists out all possible crops, it helps the farmer in decision making of which crop to cultivate. Also, this system takes into consideration the past production of data which will help the farmer get insight into the demand and the cost of various crops in market.

6. REFERENCES

- Mayank Champaneri, Chaitanya Chandvidkar , Darpan Chachpara, Mansing Rathod, "Crop yield prediction using machine learning" International Journal of Science and Research ,April 2020.
- Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate, "Crop Prediction System using Machine Learning Algorithms", International Research Journal of Engineering and Technology, Feb 2020.
- Website used to procure information and datasets : https://www.kaggle.com/code/kushagranull/crop-yield-prediction