ABSTRACT

Agroecology is one of the oldest and noblest professions in India. Farmers face numerous hardships while using traditional methods of farming in today's technologically centric world. Precision farming is a modern approach in comparison to traditional cultivation techniques.

We are predicting the right crop using parameters like district, rainfall, temperature, area, a crop which would help the farmer to predict the crop yield prior to making the decision to cultivate their final crop. This method can provide the farmer with valuable insights and assist them. In this paper we are using various techniques like XGB Regressor, Ridge Regression and LGBM Classifier.

We have used Hyperparameter Tuning on these models to get a better accuracy. We have also planned to combine both the models and also notify the farmers using SMS or E-mail.

INTRODUCTION

Agriculture is one of the oldest and noblest professions in India. It plays a critical role in the global economy and is certainly the largest livelihood provider in India. The production of Agroecology is minimal. As the demand for food is growing exponentially, brilliant minds throughout the world are working on finding new methods to help farmers to fulfil the ever-rising need for agricultural production.

Farmers being a crucial part of our society are not given their due credit. They face a lot of hardships while using the traditional methods of farming in today's technologically centric world. Farmers find it difficult to understand the technicality of the advanced science used to ease their hardships. The parameters generally used in other research papers focus more on factors like soil type and the nutrients they possess which are complicated for the farmers to comprehend. So, in this paper, we have used parameters like district, rainfall, temperature, and area which are easier for the farmers to understand.

Precision farming is a modern approach in comparison to traditional cultivation techniques. These inputs are utilized in precise amounts to get increased average yields. Initially, we have started by predicting crop yield by taking the parameters such as district, rainfall, temperature, area, a crop which would help the farmer to predict the crop yield prior to making the decision to cultivate their final crop. This

method can provide the farmer with valuable insights and eventually assist them.

In addition to crop yield, we are also helping them in predicting the right crop for cultivation. Selecting a crop for farming is one of the essential decisions farmers have to make as their whole revenue and yield depends on this decision. We are predicting the right crop using parameters like district, rainfall, temperature, area.

We have used the above-mentioned methods to minimize the errors and give better predictions which in turn would help the farmers.

1.1 Motivation

Our application provides a digital application to the farmers to help them evaluate the crop yield and decide which crop to grow.

Farmers, who are an important component of our society, are undervalued. In today's technologically focused society, they confront several challenges while employing traditional farming practice. Farmers struggle to comprehend the complexity of the modern technology employed to alleviate their sufferings. The criteria commonly employed in other study articles place a greater emphasis on aspects like soil type and nutrient content, which are difficult for farmers to grasp. So, in this study, we've chosen factors like district, rainfall, temperature, and area that are more easily understood by farmers.

1.2 Problem Statement & Objectives

Digital Farming and Precision Agriculture allow precise utilization of inputs like seed, water, pesticides, and fertilizers at the right time to the crop for maximizing productivity, quality and yields. Most of the farmers practice traditional farming patterns to decide crops to be cultivated in a field. Thus, the solution to this Challenge can be a crop recommendation system that helps farmers to decide the right crop to sow in their field and forecast the yield & revenue.

Objectives:

- Building an application that recommends the farmers about the best crops to be cultivated based on climatic parameters and also predicts the yield and revenue that would be generated for cultivated land.
- Participants must make use of IBM services (Watson studio / Auto ai/ build machine learning model) and can use any SDK to create a web Interface.
- Build Machine Learning model that predicts yield, revenue and also recommends crops using IBM capabilities.
- Integrate the model with UI(User interface)
- Build Machine Learning model that predicts yield, revenue and also recommends crops using IBM capabilities.

LITERATURE SURVEY

In this Literature Survey, we are going to give a brief idea of the existing farming system that uses conventional methods of farming, the drawbacks that come along with it due to its manual setting and labor and how our proposed system can overcome them. We also talk about the limitations that our system possesses.

2.1 Survey of Existing System

It discusses the needs and strategy required for designing a precision farming software model. It delves into the fundamentals of precision farming. The authors begin with the fundamentals of precision farming before moving on to constructing a model to support it. This research offers a model that uses Precision Agriculture concepts to control variability on small, open farms at the individual farmer and crop level with the help of easy to use technologies like SMS and emails. The model's overall goal is to provide direct advising services to even the smallest farmer at the level of his or her smallest crop plot. This model was created for the situation in Kerala, where the average holding size is significantly lower than the rest of India. As a result, with

minimal alterations, this model can be used everywhere in India.

PROPOSED SYSTEM

Dataset Collection

The dataset constitutes of region-specific attributes which are collected from districts of Karnataka, India like, Bagalkot, Chamarajanagar, Gadag, Belagavi (Belgaum), Tumakuru (Tumkur), Chikballapur, Koppal etc. A total of 29 districts are taken into consideration. The crops considered in our dataset are Arecanut, cotton, Jowar, Maize (Corn), Bajra, Rice. Figure 1 depicts the distribution of crops in that particular district. We can get a brief overview about the type of crops that grow best in that specific region.

A crop requires specific climatic as well as regional conditions to give the optimal yield. So, it becomes crucial to consider these factors for predicting the crop and its yield. While predicting a crop or its yield numerous features are taken into consideration. Thus, the features we have taken into account for predicting “Crop Yield” are District, Crop, Average Temperature, Rainfall and Area. Similarly, while predicting the “Crop” the features we have taken into consideration are District, Average Temperature, Rainfall, Area and Production (Yield). Temperature is one of the most important factors as humidity and moisture affects the production of

the crop as high air temperature reduces the growth of shoots and in turn reduces root growth. In the same way Rainfall plays a vital role in the growth and nourishment of the crop as excess or minimal amount of rainfall would destroy the crop. According to the Area under cultivation the yield of the crop is estimated.

Introduction

Our Model will help in predicting the farm yield and it will also be giving the recommendation of the crop for farming.It will basically help the farmers and give a better guidance to add to the traditional farming techniques that are currently in use.It will help the farmers to get the desired results and help them decide which crop is the best for cultivation according to the available resources. Other than giving the right crop it will also be focusing on providing the farmers the right amount of seeds, water, pesticides and fertilizers that should be used for maximum production.

3.1 Algorithm and Process Design

Ridge Regression

Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the

standard errors. It is hoped that the net effect will be to give estimates that are more reliable. Another biased regression technique, principal components regression. Ridge regression is the more popular of the two methods.

It is a model tuning method that is used to analyze any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values to be far away from the actual values.

XGB Regressor

XGBoost is a powerful approach for building supervised regression models. The validity of this statement can be inferred by knowing about its (XGBoost) objective function and base learners.

The objective function contains loss function and a regularization term. It tells about the difference between actual values and predicted values, i.e how far the model results are from the real values. The most common loss functions in XGBoost for regression problems is reg: linear, and that for binary classification is reg: logistics.

Ensemble learning involves training and combining individual models (known as base learners) to get a single prediction, and XGBoost is one of the ensemble learning methods. XGBoost expects to have the base learners which are uniformly

bad at the remainder so that when all the predictions are combined, bad predictions cancel out and better one sums up to form final good predictions.

LightGBM

LightGBM is a gradient boosting framework based on decision trees to increase the efficiency of the model and reduce memory usage.

It uses two novel techniques: Gradient-based One Side Sampling and Exclusive Feature Bundling (EFB) which fulfills the limitations of histogram-based algorithm that is primarily used in all GBDT (Gradient Boosting Decision Tree) frameworks. The two techniques of GOSS and EFB described below form the characteristics of LightGBM Algorithm. They comprise together to make the model work efficiently and provide it a cutting edge over other GBDT frameworks.

3.2 Framework

This project was implemented using the following technologies -

1. Frontend - Node Red, Json flow
2. Backend - IBM AutoAi, Model Deployment, API IBM, Pipeline

Node-RED is a flow-based development tool for visual programming developed

originally by IBM for wiring together hardware devices, APIs and online services as part of the Internet of Things.

Node-RED provides a web browser-based flow editor, which can be used to create JavaScript functions. Elements of applications can be saved or shared for re-use. The runtime is built on Node.js. The flows created in Node-RED are stored using JSON. Since version 0.14, MQTT nodes can make properly configured TLS connections.

JSON - It is an open standard file format and data interchange format that uses human-readable text to store and transmit data objects consisting of attribute–value pairs and arrays (or other serializable values). It is a common data format with a diverse range of functionality in data interchange including communication of web applications with servers.

JSON is a language-independent data format. It was derived from JavaScript, but many modern programming languages include code to generate and parse JSON-format data. JSON file names use the extension '.json'.

Json flow - In computer programming, flow-based programming (FBP) is a programming paradigm that defines applications as networks of "black box" processes, which exchange data across predefined connections by message passing, where the connections are specified externally to the processes. These black box processes can be reconnected endlessly to form different applications without

having to be changed internally. FBP is thus naturally component-oriented.

FBP is a particular form of dataflow programming based on bounded buffers, information packets with defined lifetimes, named ports, and separate definition of connections.

IBM Auto Ai - AutoAI is a variation of AutoML. It extends the automation of model building to the entire AI lifecycle. Like AutoML, AutoAI applies intelligent automation to the steps of building predictive machine learning models. These steps include preparing data sets for training; identifying the best type of model for the given data, such as a classification or regression model; and choosing the columns of data that best support the problem the model is solving, known as feature selection. Automation then tests a variety of hyperparameter tuning options to reach the best result as it generates, and then ranks, model-candidate pipelines based on metrics such as accuracy and precision. The best performing pipelines can be put into production to process new data and deliver predictions based on the model training.

API - An application programming interface (API) is a connection between computers or between computer programs. It is a type of software interface, offering a service to other pieces of software. A document or standard that describes how to build such a connection or interface is called an API specification. A computer system that meets this standard is said to implement or

expose an API. The term API may refer either to the specification or to the implementation.

In contrast to a user interface, which connects a computer to a person, an application programming interface connects computers or pieces of software to each other. It is not intended to be used directly by a person (the end user) other than a computer programmer who is incorporating it into software. An API is often made up of different parts which act as tools or services that are available to the programmer. A program or a programmer that uses one of these parts is said to call that portion of the API. The calls that make up the API are also known as subroutines, methods, requests, or endpoints. An API specification defines these calls, meaning that it explains how to use or implement them.

One purpose of APIs is to hide the internal details of how a system works, exposing only those parts a programmer will find useful and keeping them consistent even if the internal details later change. An API may be custom-built for a particular pair of systems, or it may be a shared standard allowing interoperability among many systems.

3.3 Results

The model predicting Crop-Yield we found that XGB Regressor gave us the most optimized solution compared to Ridge Regressor. Further when XGBRegressor used with Hyperparameter Tuning gave us a better result as RMSE value before Hyperparameter Tuning was 1802.774 and after using Hyperparameter Tuning the RMSE value was 1704.057. Hence, XGB Regressor accompanied with Hyperparameter Tuning gave us the most optimal solution.

In the model predicting Crop we found that LGBM Classifier gave us the most optimized solution compared to XGB Regressor. Further when LGBM Classifier used with Hyperparameter Tuning gave us a better accuracy. Hence, LGBM Classifier accompanied with Hyperparameter Tuning gave us the most optimal solution.

3.4 Conclusion & Future Scope

We applied the regression and classification technique. However, We got improvised results using Hyperparameter Tuning.. The future implementation for our research would be creating a website and an application for the convenience of the farmers which would provide them a platform which has

the facility of receiving an email or SMS. We can integrate both the models i.eCrop-Yield Prediction and Crop Prediction all together in one model.

References

- Babu, S. (2013). A software model for precision agriculture for small and marginal farmers. *2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS)*. https://doi.org/10.1109/ghtc-sas.2013.6629944
- NCSS Statistical Software. (n.d.). *Ridge Regression*. https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Ridge_Regression.pdf
- Brownlee, J. (2021, March 6). *XGBoost for regression.* Machine Learning Mastery. https://machinelearningmastery.com/xgboost-for-regression/.
- Nikki. (2020, September 2). *Xgboost for regression*. GeeksforGeeks. https://www.geeksforgeeks.org/xgboost-for-regression/.
- *LightGBM (light gradient Boosting Machine).* GeeksforGeeks. (2020, July 16). https://www.geeksforgeeks.org/lightgbm-light-gradient-boosting-machine/.
- Google. (n.d.). *Overview of hyperparameter tuning | ai platform training. Google.* https://cloud.google.com/ai-platform/training/docs/hyperparameter-tuning-overview.
- https://www.ibm.com/in-en/cloud/watson-studio/autoai