International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.147

Agro Prognostics Based on Characteristics of the Agricultural Environment Using Machine Learning

UdayKiran Reddy Aduri
Department of Data Science and Business Systems,
School of Computing,
SRM Institute of Science and Technology,
Kattankulathur, Tamil Nadu, India

PavanKiranKuchipudi
Department of Data Science and Business Systems,
School of Computing,
SRM Institute of Science and Technology,
Kattankulathur, Tamil Nadu, India

M. Prakash
Department of Data Science and Business Systems,
School of Computing,
SRM Institute of Science and Technology,
Kattankulathur, Tamil Nadu, India
prakashm2@srmist.edu.in

***Abstract*—Agriculture is an important field of study and agricultural forecasting is an important aspect that depends on a lot of factors. In the past, farmers were free to choose their crops based on factors such as soil composition, rainfall, humidity and temperature conditions. They then observe the growth of these crops and determine the ideal time to harvest them. However,due to rapidly changing environmental conditions, this approach is no longer sustainable.**

Crop forecasting has become predominantly reliant on machine learning techniques in recent years, which have been extensively employed in this area of research. To ensure that machine learning models perform with high accuracy, it is important to use efficient feature selection methods to preprocess uncleaned and basic data and create datasets suitable for machine learning. Only data features that are highly relevant to defining the final model output should be used to eliminate redundancy and improve model accuracy. it is of utmost importance to select the most appropriate features for a model, ensuring that only the most relevant ones are incorporated. Failure to consider the significance of the data characteristics during the modeling process will result in an unnecessarily intricate model, incorporating all the features of the raw data.

Also, adding features that don't contribute much to a machine learning model increases temporal and spatial complexity, which hurts its accuracy. The results of this study demonstrate that the ensemble approach outperforms current classification techniques in terms of predictive accuracy.

Keywords—Agriculture, classification, Agro Prognostics, feature selection.

I. INTRODUCTION

Agricultural forecasting is a complex process in agriculture that requires the use of multiple data sets due to the reliance on the process of sowing seeds is dependent on both living and non-living factors. Biological factors include environmental components resulting from the direct or indirect action of living species (microbes, plants, animals, parasites, predators, pests) on other living organisms, as well as human variables such as fertilization, protection of plants, irrigation, air, and water pollution, soil quality. These factors lead to various variations in crop production, such as internal errors, yields. Abiotic and biotic factors have an impact on the environmental formation, development and quality of plants. Abiotic factors include physical, chemical, and other factors such as mechanical vibration, radiation, climatic conditions, soil type, topography, atmosphere, and water chemistry.

Chemical components such as environmental toxins, nitrogen fertilizers, pesticides and heavy metals also play a role. Abiotic factors such as bedrock, topography, climate and water conditions all impact crop quality. Soil formation variables also have broad impacts on soil formation and agricultural values.

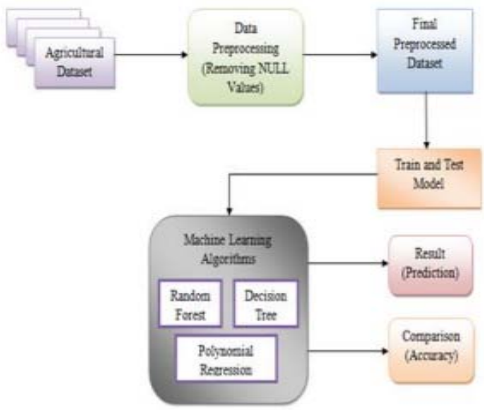

Fig.1: Example figure

Analyzing the agricultural production process is a difficult and complex task. strategies are essential for predicting crop area, which is a continuous improvement optimization process. These strategies can also be used in the design, development and formulation of new and improved products. However, to perform statistical analysis, it is necessary to have access to numerical data, which allows to make inferences about different events and thus to take binding economic decisions.

According to Muriithi [6], the representation of a specific event numerically increases the amount of information that can be obtained, and the improved data quality allows for more accurate information and more accurate decision making.

II. LITERATURE REVIEW

Using naive Bayes classification techniques to classify improved agricultural land soils:

DOI: rp-9788770040723.147

The proliferation of computing and data storage technologies has led to the accumulation of large amounts of data. Extracting useful information from this raw data has always been a challenge, which has led to the development of new methods and techniques such as data mining, which help bridge the knowledge gap. This study aimed to evaluate these new data mining methods and apply them to soil science databases to determine if relevant associations could be found. The Department of Soil Science and Agrochemistry of SV College of Agriculture, Tirupati provides an extensive soil database containing soil profile measurements from many locations in Chandragiri Mandal, Chittoor district. This study investigates whether soils can be classified using different data mining methods.

In addition, a comparison of naive Bayesian classifications is performed and the best performing methods are analyzed. The results of the study could have wide applications in agriculture, land management and environmental protection.

Biotic factors

Potato production in Canterbury has remained steady at around 60 tonnes per hectare for the past decade. However, potato growth models predict potential yields as high as 90 tonnes/ha, which some commercial growers have already achieved. A two-year study conducted by industrial and academic partners examines constraints to agricultural production. During the first growing season, 11 transformed crops were fully evaluated, including tests for final yield, plant health and soil quality. Persistent reasons for declining yields have been found to include soil-borne diseases such as Rhizoctonia stem canker and sponge root infection, as well as subsoil compaction and poor management. irrigation.

A crop history of potatoes grown over the past decade has resulted in more rapid development of Rhizoctonia stem canker symptoms (by emergence) compared to periods of grass growth and non-crop areas previous potato crop (8 weeks after emergence). In year two, a controlled field trial was conducted on a cash crop known to have high levels of soil-borne pathogens to identify and quantify the impact of soil-borne diseases. on performance. The treatments consisted of soil fumigants (90, 112 and 146 kg/ha of chloropicrin) without pesticide control before the application of azoxystrobin (1.5 L/ha) or fluosulfa (400 mL/ha). DNA testing of soil-borne pathogens before and after treatment showed a slight decrease in DNA levels of R. solani and subterranean sponges in soil (plots treated with the fumigant), but results varied. considerably.

The final total fresh yield averaged 58 tonnes/ha and was unaffected by the treatment. The severity of R. blight on rhizomes treated with azoxystrobin was consistently lower than all other treatments throughout the season.

Response surface methodology: A retrospective and literature survey

Response Surface Methodology (RSM) is a combination of statistical design and numerical optimization methods used to improve process and product design. The roots of exploration in this field trace back to the 1950s and have garnered extensive utilization, notably within chemical and process industries. In recent times, RSM has undergone a gamut of inventive upgrades and has found diverse applications over the last 15 years. Our analysis focuses on RSM developments since 1989, highlighting contemporary fields of investigation and recommending potential avenues for future inquiry.

Utilizing response surface methodology to enhance potato tuber production:

The study investigated the optimal operating parameters for obtaining maximum yields of potato tubers in Kenya, with the aim of helping potato growers reduce input costs. To achieve this goal, the authors used a 2 3 factorial design and response surface methods to improve the potato manufacturing process. Using response surface methodology, the authors studied and optimized the combined effects of the mineral nutrients of water, nitrogen, and phosphorus. They determined that the optimal production parameter for potato tuber yield was 70.04% irrigation water, 124.75 kg/ha of nitrogen in the form of urea and 191.04 kg/ha of phosphorus in the form of triple superphosphate. Under ideal conditions, a 1.8m x 2 plot can produce up to 19.36kg of potato tubers.25 meters. Increasing potato production could improve livelihoods and reduce input costs for smallholder potato farmers in Kenya. Additionally, the methods used in this study can be applied to other crop studies to better understand overall crop productivity.

Enhancing potato crop yield forecasting through the integration of cultivar details and unmanned aerial vehicle (UAV) remote sensing data via machine learning:

Precision agriculture requires accurate yield maps to identify patterns of geographic variation in yields, identify key variables driving yield variation, and provide site-specific management information. Yield forecasting of potato (Solanumtuberosum L.) tubers using remote sensing is critical, as cultivar variation can have a significant impact on yield predictions. with a machine learning approach using unmanned aerial vehicle (UAV) remote sensing.

In the years 2018 and 2019, various cultivars and levels of nitrogen (N) were tested on small plots to assess their effectiveness.

Gather multispectral imagery via drones at different stages of plant growth and employ machine learning algorithms like random forest regression (RFR) and support vector regression (SVR) to integrate multiple plant metrics with varietal data. The research discovered that spectral data obtained during the onset of tuber initiation (in late June) exhibits a stronger correlation with marketable potato yields compared to data obtained during the maturity of the tuber in the later stages of the growing season. Nonetheless, the optimal nutrient markers and time for anticipating potato yields differ depending on the potato variety. Model that do not use variety information. Further research is needed to improve potato yield predictions by combining more specific data on varieties, soil and landscape characteristics,

DOI: rp-9788770040723.147

management data and advanced machine learningAlgorithms.

III. Methodology

Assessing agroclimatic parameters that influence the production of winter plant species in the temperate temperature zone, primarily grains, presents a significant challenge. The primary factor that impacts wintering yield is the number of days with temperatures above certain thresholds, such as 5 degrees Celsius, as well as the frequency of these days and the number of days with temperatures above 0 degrees Celsius and 5 degrees Celsius during the wintering period. While some of these parameters can be approximated using public data, fluctuations in these components pose a challenge. Accurate predictions of agrometeorological parameters are crucial for precise production forecasting, especially in light of rapid changes in environmental circumstances that make farming increasingly challenging.

Disadvantages

Despite the progress made in the field of agro analytics, there are still various challenges that need to be overcome, including the need for improved crop forecast models that account for soil and environmental factors such as rainfall, humidity, and temperature. In this regard, a research study proposes an enhanced crop forecast model that uses two essential techniques: feature selection (FS) and classification. The research study utilizes sampling approaches to balance an unbalanced dataset before using FS methods to identify data characteristics that are highly important while checking the end output of the model.

Advantages

One of the advantages of this proposed model is that it avoids redundancies and improves the accuracy of the machine learning model by only including data characteristics with a high degree of importance. Additionally, an ensemble method outperforms the previous classification technique, resulting in increased prediction accuracy. This improved crop forecast model has the potential to overcome some of the difficulties in predicting agrometeorological parameters for winter plant species, particularly grains, and provide farmers with more accurate information for decision-making in the face of rapid environmental changes.

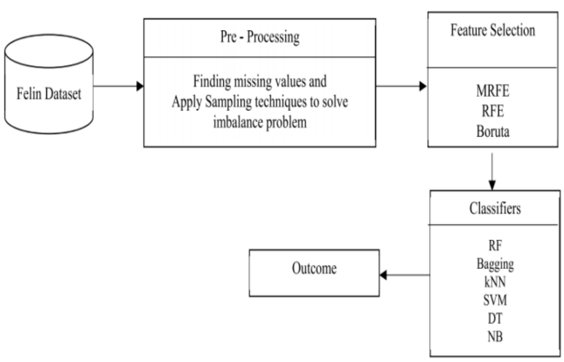

Fig.2: System architecture

Modules

- The following modules were developed for the project at hand:
- Data Exploration: This module will facilitate the input of data into the system.
- Data Processing: This module will read and process the input data.
- Train-Test Split: This module will be used to separate the data into training and testing sets.
- Model Generation: This module will be responsible for building the machine learning models, including models with and without feature selection. The feature selection techniques to be used include SMOTE, ROSE, RFE, MRFE, BORUTA, and MEMOTE. The machine learning algorithms used for model building include Naive Bayes, KNN, Bagging Classifier, Random Forest Decision Tree, SVM, Gradient Boosting, and Voting Classifier. The accuracy of the algorithms will also be calculated.
- User Signup and Login: This module will enable users to register and log in to the system.
- User Input: This module will allow users to input data to be predicted.
- Prediction: The final prediction will be displayed through this module.

IV. Implementation

Algorithms

KNN (K-Nearest Neighbor): KNN is a supervised machine learning algorithm that can handle both classification and regression problems. In this algorithm, the number of nearest neighbors (denoted by "K") of a new unknown variable is used to predict or rank the variable.

Naive Bayes: Naive Bayes is a probabilistic classifier based on a probabilistic model with high independence assumptions. The algorithm assumes that the input variables are independent, which may not be the case in reality.

Bagging Classifier: A Bagging Classifier can be thought of as a team of base classifiers that work together to make predictions. Each member of the team is trained on a random sample of the original data, and their predictions are combined to form a final output. This technique is particularly useful for reducing the uncertainty of complex machine learning models.

Random Forest: Random Forest is a supervised machine learning algorithm that can handle both classification and regression problems. It is based on the idea of ensemble learning, which consists of integrating several classifiers to improve the performance of the model. A random forest consists of many decision trees over different subsets of the provided dataset, and the end result is a majority vote based on predictions.

Decision trees: Decision trees use several methods to determine whether a node should be split into two or more child nodes.

DOI: rp-9788770040723.147

SVM (Support Vector Machine): SVM is a popular supervised learning method used for classification and regression tasks. The SVM algorithm is designed to find the best straight line or decision boundary for classifying an n-dimensional space so that new data points can be easily assigned to the correct class in the future. The optimal choice boundary is represented by a hyperplane.

Gradient Boosting is a powerful technique in the realm of machine learning, widely applied to tackle regression and classification challenges. Essentially, it constructs a predictive model by aggregating a series of feeble models, which are usually decision trees. By leveraging decision trees as weak learners, the method is commonly referred to as boosted gradient trees.

Speech Classifier: A speech classifier is a machine learning estimator that trains multiple base models, or estimators, and makes predictions based on the results of each base estimator. The voting decisions from each estimator output can be aggregated to make a final prediction.

REFERENCES

[1] R. Jahan, “Applying naive Bayes classification technique for classification of improved agricultural land soils,” International Journal of Research and Applied Sciences, Engineering and Technology, vol. 6, no. 5, pp. 189-193, 2018.

[2] B.B. Sawicka, and B.KrochmalMarczak, “Biotic components influencing the yield and quality of potato tubers,” Herbalism, vol. 1, no. 3, pp. 125-136, 2017.

[3] B. Sawicka, A.H. Noaema, and A. Gaowacka, “Predicting the size of the potato acreage as a raw material for bioethanol production,” In B. Zdunek& M. Olszáwka (Eds.), Alternative Energy Sources Lublin, Poland: WydawnictwoNaukowe TYGIEL, pp. 158-172, 2016.

[4] B. Sawicka, A. H.Noaema, T.S. Hameed, and B. KrochmalMarczak, “Biotic and abiotic factors influencing on the environment and growth of plants,” In BioróżnorodnośćŚrodowiskaZnaczenie, Problemy, Wyzwania. MateriałyKonferencyjne, Puławy, May 2017. Retrieved from https://bookcrossing.pl/ksiazka/321192

[5] R.H.Myers, D.C. Montgomery, G.G. Vining, C.M. Borror, andS.M. Kowalski, “Response surface methodology: A retrospective and literature survey,” Journal of Quality Technology, vol. 36, no. 1, pp. 53-77, 2004.

[6] D.K. Muriithi, “Application of response surface methodology for optimization of potato tuber yield,” American Journal of Theoretical and Applied Statistics, vol. 4, no. 4, pp. 300-304, 2015. doi: 10.11648/j.ajtas.20150404.20

[7] Moshika, A., Thirumaran, M., Natarajan, B., Andal, K., Sambasivam, G., &Manoharan, R. (2021).Vulnerability assessment in heterogeneous web environment using probabilistic arithmetic automata. IEEE Access, 9, 74659-74673.

[8] J.R. Olędzki, “The report on the state of remote sensing in Poland in 2011-2014,” Remote Sensing and Environment, vol. 53, no. 2, pp. 113-174, 2015.

[9] K. Grabowska, A.Dymerska, K. Poaarska, andJ. Grabowski, “Predicting blue lupine yields based on selected climate change scenarios,”ActaAgroph., vol. 23, no. 3, pp. 363-380, 2016.

[10] Rajesh, M., &Sitharthan, R. (2022). Image fusion and enhancement based on energy of the pixel using Deep Convolutional Neural Network. Multimedia Tools and Applications, 81(1), 873-885.