ISSN 2277-2685
IJESR/Jul-Sep. 2024/ Vol-14/Issue-3/413-423
M.A Ghani *et. al.,* / International Journal of Engineering & Science Research

AGRICULTURE SOIL ANALYSIS, CLASSIFICATION AND CROP SUITABILITY RECOMMENDATION USING MACHINE LEARNING

M.A Ghani[1], Syed Rizwan Ul Haq[2], M.D Mohib Khan[3], Bhargavi Bendalam[4]
[1,2,3]B.E. Student, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
[4] Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
bhargavi@lords.ac.in

Abstract: *Increasing crop yield is crucial to meet the growing demands of the population. Many Indian farmers own small, fragmented plots of land, and their crop productivity is influenced by factors like soil quality, rainfall, and environmental conditions. India experiences an annual soil loss of approximately 5.3 billion tonnes, leading to reduced productivity on degraded land. Soil fertility significantly impacts agricultural output, varying with nutrient levels and suitability for different crops. Evaluating soil's physical, chemical, and biological properties helps determine fertility, plan cultivation strategies, and forecast crop yields.*
Keywords: *Soil analysis, crop suitability, machine learning, supervised learning, classification.*

I. Introduction

Agriculture forms the foundation of India's economy. As of 2011, approximately 60.5% of India's land is dedicated to agriculture, including arable land (52.8%), land for permanent crops (4.2%), and pastures (3.5%). Agriculture and related activities contributed 17.1% to the GDP in 2017-18 and employ about 42% of the country's workforce. Major crops covered 15 million hectares during the Kharif season and 57 million hectares during the Rabi season in 2013-2014.

Many farmers now take soil samples to nearby Krishi Vigyan Kendra (KVK) centers for testing soil nutrients and characteristics. Soil testing involves analyzing a sample to assess its fertility and identify any deficiencies that may affect crop growth. While Soil Health Cards (SHCs) record this data, they do not directly advise farmers on optimal crop choices for maximum yield.

Despite advancements in soil analysis through SHCs, decisions on crops, fertilizers, and their proportions largely rely on traditional knowledge and discussions with local experts. This traditional approach, while beneficial in some respects, contributes to challenges such as soil degradation from excessive fertilizer use, declining yields, and negative impacts on ecosystems and human health over time.

Concurrently, there has been significant progress in computational and data science. Agriculture has embraced digitization, with digitized land maps, extensive satellite imagery, and datasets detailing soil nutrient compositions. Farmers now access these resources through mobile phones and network connectivity, facilitating tasks like SHC registration via mobile apps.

II. Literature Survey

1) Crop Recommendation System for Precision Agriculture. Authors: S. Pudumalar et al. [1]

This study employs data mining techniques to recommend crops based on site-specific soil characteristics and crop yield data. By integrating multiple learning algorithms (Random tree, CHAID, K-Nearest Neighbour, Naive Bayes) through an ensemble model with majority voting, the system aims to reduce incorrect crop choices and enhance productivity.

2) A Machine Learning Approach to Assess Crop Specific Suitability for Small/Marginal Scale Croplands. Authors: Bhimanpallewar R et al. [2]

Proposes a machine learning framework to assess crop suitability based on soil composition, environmental factors, and selected crops. The system aids decisions on improving soil suitability or optimal land usage, thereby enhancing agricultural productivity.

3) Using Parallel Random Forest Classifier in Predicting Land Suitability for Crop Production. Authors: Senagi K et al. [3]

Utilizes optimized machine learning algorithms such as Parallel Random Forest (PRF), Linear Regression (LR), Linear Discriminant Analysis (LDA), KNN, Gaussian Naïve Bayesian (GNB), and Support Vector Machine (SVM) to predict land suitability for sorghum production. The study demonstrates experimental setups and comparative analyses to determine the most effective predictive models.

4) An Analysis of Agricultural Soils by using Data Mining Techniques. Authors: Ramesh Babu, Rajesh Reddy [4]

Examines agricultural soil parameters using data mining techniques, focusing on classification and analysis to enhance understanding of soil behavior and predict crop yield. The study contributes insights into leveraging computational methods for soil management and crop production.

5) Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. Authors: Kussul N et al. [9]

Introduces a deep learning architecture for classifying land cover and crop types from multi-temporal, multi-source satellite imagery. The study employs unsupervised neural networks for optical imagery segmentation and supervised approaches like convolutional neural networks (CNNs) for accurate crop classification, showcasing advancements in remote sensing applications for agriculture.

literature survey highlights various research efforts employing machine learning and data mining techniques to enhance agricultural practices, specifically focusing on soil classification, crop recommendation systems, and remote sensing applications. These studies provide foundational insights that can inform the development of innovative solutions for optimizing crop yield and sustainable agriculture practices..

III.System Analysis

The current practices of soil analysis and crop recommendation show several deficiencies leading to suboptimal crop yields. Accessibility to soil testing facilities across the country is limited, particularly for small-scale farmers who may find the associated costs prohibitive. Moreover, reliance on traditional methods and soil health cards often overlooks advancements in technology and data sciences.

1) Farmers traditionally find it difficult to choose which crop is most suited and financially advantageous to their soil, their circumstances, and their location because of the variability in soil types across the country, andas a result they frequently suffer losses.

2) Due to unpredictable weather patterns, it is currently astronomically difficult for farmers to predict the yield for a specific planting season and the profit that they may make.

3) Due to the "farm to market" system, which involves hundreds of intermediaries who victual up the majority of the revenues by transporting and selling crops, farmers get dismally little returns for their labor. Artificial intelligence and machine learning are widely used in modern agriculture. Precision farming methods, crop recommender systems, and yield prediction methods may all be used to improve farm output, plant pest detection, and overall harvest quality. AI system deployment might provide the struggling agriculture industry a boost. One of the emerging technologies in agriculture is machine learning. The agriculture industry may use machine learning to improve crop quality and output. Finding trends in the agricultural data and relegating them to more important data might become habitual. You can use this information for further procedures.

Gathering data, processing it, and training it before testing it on samples of data are the steps that machine learning approaches often take. For the relegation of soil and crop prognostication based on prior patterns followed and the kind of soil, an algorithm like SVM may be used. The following datasets are needed for the project: a soil dataset with multiple chemical qualities and a crop dataset with geographical information.

There is minimal integration of modern techniques like machine learning and remote sensing for data-driven soil classification and crop suitability analysis. Additionally, existing solutions lack comprehensive computerized systems, advanced geospatial capabilities, and coordination across initiatives.

Deploying and scaling up research efforts remains challenging, highlighting the need for an integrated framework leveraging computational power, digital data access, and expertise in soil and crop sciences to enhance agricultural productivity effectively.

Disadvantages of the present model

1.Soil Health Cards may lack actionable insights for farmers regarding optimal crop selection.

2.Lack of an integrated end-to-end solution covering soil analysis and crop recommendation.

3.Limited real-world implementation of research using machine learning for soil classification.

4.Potential bureaucratic obstacles and resistance to integrating modern systems with traditional practices.

5.Inconsistent quality of soil testing and recommendations from different facilities.

6.Inadequate computerized systems and geospatial capabilities for robust data analysis.

Proposed System

The paper proposes an integrated system that utilizes machine learning and data science to improve crop productivity. The system operates in two phases: first, classifying soil based on compositional data and imagery analysis, and second, matching crops to these soil classifications based on crop requirements. Soil classification leverages biochemical composition, soil images, and satellite data, while crop mapping considers factors such as nutrient needs and pH levels. Modern ML techniques including neural networks, SVM, and decision trees, alongside statistical tools, are employed. The primary goals are accurate soil classification, optimal crop recommendations, mitigating soil degradation, and enhancing crop yields. Overall, the system aims to provide a comprehensive solution from soil analysis to crop suitability recommendations, using technological advancements and data-driven approaches to boost agricultural productivity.

Improvements necessities

1.Encourages crop rotation over mono-cropping through tailored recommendations.
2.Potential for increased crop yields and farmer incomes by suggesting suitable crops.
3.Accessibility to recommendations via mobile apps and computerized systems, surpassing traditional soil testing methods.
4.Utilizes advanced computation capabilities such as cloud-based GPUs for swift and precise data analysis.
5.Incorporates remote sensing and geospatial data for accurate farm mapping and insights.
6.Simulation of various conditions for precise crop recommendations and system improvements.
7.Automated classification and crop suggestions to minimize human errors.

IV.System Viabilities

The viability study for the project "Agriculture Soil Analysis, Classification, and Crop Suitability Recommendation Using Machine Learning" evaluates the practicality and viability of implementing the proposed system.

This phase includes a business proposal outlining the project's general plan and initial cost estimates, ensuring that the system development does not impose an undue burden on the organization.

Economical practicability assesses the financial impact of the system on the organization. It examines whether the project can be completed within a reasonable budget and whether the anticipated benefits outweigh the costs. In this case, the feasibility study indicates that the majority of technologies required for the system, such as machine learning algorithms and open-source software, are freely available. Customized components may need to be procured but are expected to be within budgetary constraints.

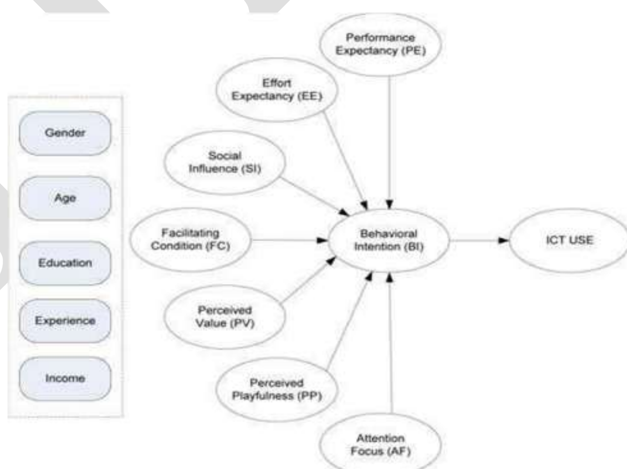
Technical Feasibility: Technical feasibility evaluates the technical requirements of the system and ensures that it does not place excessive demands on existing technical resources. The proposed system must be compatible with current hardware and software infrastructure with minimal modifications or additional investments required. By leveraging widely supported machine learning frameworks and existing geospatial and remote sensing tools, the system aims to maintain modest technical requirements and compatibility.

Social Feasibility: Social feasibility assesses the acceptance of the system by its users, particularly farmers and agricultural stakeholders. User acceptance is crucial for the successful implementation and adoption of the system. Efforts will be made to train users effectively, ensuring they feel confident and comfortable using the system. User training programs will emphasize the benefits of data-driven crop recommendations and soil analysis, encouraging constructive feedback and user engagement. By addressing user concerns and demonstrating the system's value in enhancing agricultural productivity, social acceptance can be effectively fostered.

Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms " - Agriculture is a vital a part of human lives. it's one in every of the most important supply of employment in India. More than 1/2 the population depend on agriculture. it's the backbone of our economy. Crop yield depends on several factors. one in every of the most important factors that have an effect on the yield of the crop is soil. Improvising the techniques to predict crop yield in several atmospheric condition will facilitate farmers and other stakeholders in higher higher cognitive process in terms of scientific agriculture and crop choice. Crop yield prediction includes foretelling the yield of the crop from previous historical knowledge that consists of things like temperature, humidity, pH, rain and crop name. It provides United States of America an inspiration for the best foreseen crop which can be cultivate within the field weather. In the planned work, a comparative analysis on soil properties to predict fertility and crop yield has been performed mistreatment machine learning algorithms. The analysis has been done on self –obtained dataset, for 3 crops - tomato, potato and chili. The crop yield prediction has been done mistreatment K Nearest Neighbour algorithmic program, Naïve Thomas Bayes algorithmic program and call Trees classifier.

The viability study concludes that the proposed system for Agriculture Soil Analysis, Classification, and Crop Suitability Recommendation Using Machine Learning is economically viable, technically feasible with minimal impact on existing resources, and socially acceptable with appropriate user training and engagement strategies. These considerations provide a solid foundation for proceeding with the development and implementation phases of the project.

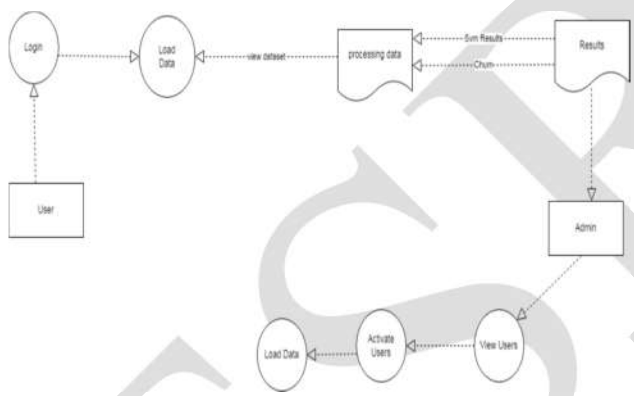
V. The Process Architecture



The bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system. The data flow diagram is one of the most important modelling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

It shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

It may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.



UML stands for Unified Modelling Language. UML is a standardized general-purpose modelling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modelling Language is a standard language for specifying, Visualization, Constructing and documenting the artefacts of software system, as well as for business modelling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modelling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

Goals:

The Primary goals in the design of the UML are as follows:

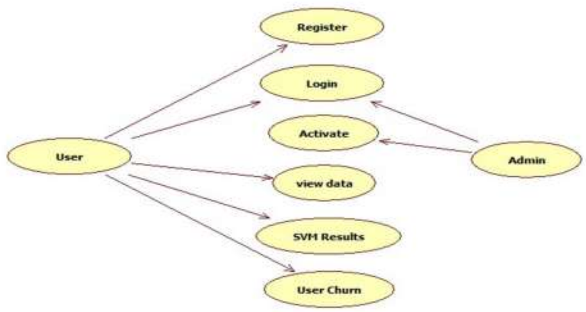
Provide users a ready-to-use, expressive visual modelling Language so that they can develop and exchange meaningful models.

Provide extendibility and specialization mechanisms to extend the core concepts.

Be independent of particular programming languages and development process.

Provide a formal basis for understanding the modelling language.

Encourage the growth of OO tools market.



Support higher level development concepts such as collaborations, frameworks, patterns and components.

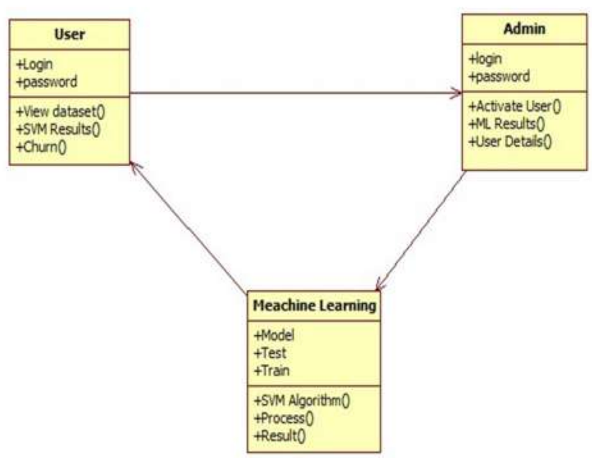
Integrate best practices.

A use case diagram in the Unified Modelling Language (UML) is a type of behavioural diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases.

The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

Class Diagram

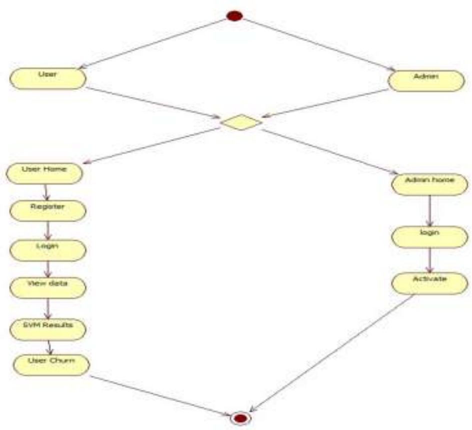
In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

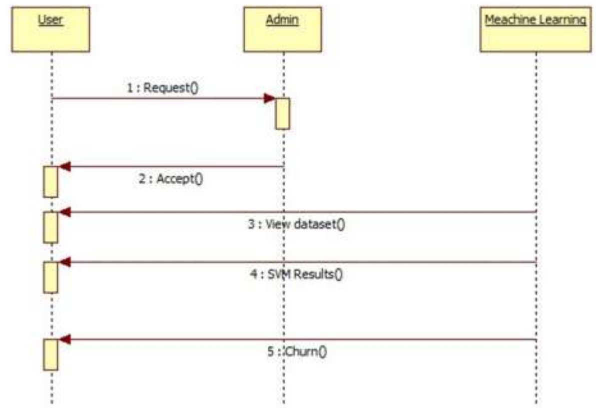


Sequence Diagram

A sequence diagram in Unified Modelling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order.

It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams





Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

VI.Modules Vaibilities

In an large data set more number of data consisting of soil images and their features will undergo for Training the model. After the Training of the model only by using our we can predict the soil type and its features. And the remaining data will be used for testing the model which trained. This Data set of soil which will be divided for both the training and testing. In an large data set more number of data consisting of soil images and their features will undergo for Training the model. After the Training of the model only by using our we can predict the soil type and its features. And the remaining data will be used for testing the model which trained. The crop data set which will be also divided to both training and testing. In an large data set more number of data consisting of crop images and their details will undergo for Training the model. After the Training of the model only by using our we can predict the crop which can be cultivated in that soil. And the remaining data will be used for testing the model which trained.

User: The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the user. Once admin activated the user then user can login into our system. User can upload the dataset based on our dataset column matched. For algorithm execution data must be in float format. Here we took Three Customer Behaviour dataset for testing purpose. User can also add the new data for existing dataset based on our Django application. User can click the Classification in the web page so that the data calculated Accuracy and F1-Score, Recall, Precision based on the algorithms. User can click Prediction in the web page so that user can write the review after predict the review that will display results depends upon review like positive, negative or neutral.

Admin: Admin can login with his login details. Admin can activate the registered users. Once he activate then only the user can login into our system. Admin can view the overall data in the browser. Admin can click the Results in the web page so calculated Accuracy and F1-Score, Precision, Recall based on the algorithms is displayed. All algorithms execution complete then admin can see the overall accuracy in web page.

Data Preprocessing: A dataset can be viewed as a collection of data objects, which are often also called as a records, points, vectors, patterns, events, cases, samples, observations, or entities. Data objects are described by a number of features that capture the basic characteristics of an object, such as the mass of a physical object or the time at which an event occurred, etc. Features are often called as variables, characteristics, fields, attributes, or dimensions. The data preprocessing in this forecast uses techniques like removal of noise in the data, the expulsion of missing information, modifying default values if relevant and grouping of attributes for prediction at various levels.

Machine learning:

Based on the split criterion, the cleansed data is split into 60% training and 40% test, then the dataset is subjected to four machine learning classifiers such as Support Vector Machine (SVM). The accuracy, Precision, Recall, F1-Score of the classifiers was calculated and displayed in my results. The classifier which bags up the highest accuracy could be determined as the best classifier.

VII. CONCLUSION

In conclusion, the implementation of Agriculture Soil Analysis, Classification, and Crop Suitability Recommendation Using Machine Learning offers significant advantages for agricultural practices. By leveraging advanced technologies such as machine learning algorithms, remote sensing, and geospatial data analysis, the system provides a comprehensive approach to optimizing crop yields and soil management.

The integration of machine learning allows for precise soil classification based on biochemical composition, soil images, and satellite data, enhancing the accuracy of crop suitability recommendations. This data-driven approach not only mitigates soil degradation but also promotes sustainable agricultural practices by recommending appropriate crops based on soil conditions and environmental factors.

Furthermore, the system's ability to discourage mono-cropping through crop rotation suggestions contributes to long-term soil health and resilience against pests and diseases. Farmers and agricultural stakeholders can benefit from easy access to recommendations through user-friendly interfaces and mobile applications, facilitating informed decision-making and enhancing productivity.

Overall, Agriculture Soil Analysis, Classification, and Crop Suitability Recommendation Using Machine Learning represents a pivotal step towards modernizing agriculture, improving yield predictability, and ensuring food security in a technologically advanced and sustainable manner.

XII. REFERENCES

[1] S. Pudumalar, E. Ramanujam, R.Harine Rajashree, C. Kavya, T. Kiruthika, J. Nisha., "Crop recommendation system for precision agriculture," IEEE Electron Device Lett, June, 2017. https://ieeexplore.ieee.org/document/7951740.

[2] Bhimanpallewar R et al., "A Machine Learning Approach to Assess Crop Specific Suitability for Small/Marginal Scale Croplands" International Journal of Applied Engineering Research, ISSN 0973 4562 Volume 12, Number 23 (2017) pp. 13966-13973. http://www.ripublication.com

[3] Senagi K et al., "Using parallel random forest classifier in predicting land suitability for crop production", Journal of Agricultural Informatics (ISSN 2061-862X) 2017 Vol. 8, No. 3:23-32.

[4] Ramesh Babu, Rajesh Reddy, "An Analysis of Agricultural Soils by using Data Mining Techniques" IJESC, 2017, Volume 7 Issue No.10.

[5] Supriya D., "Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining Approach", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2017, DOI: 10.15680/IJIRCCE.2017. 0505067.

[6] Sirsat M et al., "Classification of agricultural soil parameters in India" Article in Computers and Electronics in Agriculture, April 2017.

[7] B. Tanmay et al., "Crop Recommendation System Using Neural Networks", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 05, May-2018.

[8] M. Mokarram, et al., "Using Machine Learning for Land Suitability Classification".

[9] Kussul N et al., "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data" https://ieeexplore.ieee.org/document/7891032

[10] Driessen B et al., "Improving crop classification with landscape stratification based on MODIS-time series", Laboratory of Geo Information Science and Remote Sensing, Wageningen University and Research Centre, The Netherlands.

[11] Ji S et al., “3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images”, Remote Sensing. 2018, 10, 75; doi:10.3390/rs10010075 www.mdpi.com/journal/remotesensing

[12] Rose M. Rustowicz, “Crop Classification with Multi-Temporal Satellite Imagery”, Stanford University, CA, USA.

[13]https://medium.com/sciforce/machine-learning-in-agriculture applications-and-techniques-6ab501f4d1b5

[14]https://en.wikipedia.org/wiki/Convolutional_neural_network

[15]https://medium.com/machine-learning-101/chapter-2-svm-support vector-machine-theory-f0812effc72

[16] https://www.kdnuggets.com/2017/11/real-world-deep-learning-neural networks-smart-crops.html

[17]Pranay Malik, Sushmita Sengupta, Jitendra Singh Jadon, ” Comparative Analysis of Soil Properties.

[18.] Predict Fertility and Crop Yield using Machine Learning Algorithms “, 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence).

[19] Jie Ren , Jing Liang , and Yuanyuan Zhao “ Soil PH Measurement Based on Compressive Sensing and Deep Image Prior “ , IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 4, no. 1, February2020.

[20] Vitalik Buterin, Evgeniou, Theodoros Pontil, Massimiliano. (2001). Support Vector Machines:Theory and Applications. 2049. 249-257. 10.1007/3- 540-44673-712.

[21.] M. Gopal P S and B. R, “Selection of important features for optimizing cropyield prediction,” Int. J. Agricult. Environ. Inf. Syst., vol. 10, no. 3, pp. 54– 71,Jul. 2019.

Online references:

[1] https://www.tensorflow.org/guide

[2] https://towardsdatascience.com/the-actual-difference-between statistics-and-machine-learning-64b49f07ea3

[3]https://www.nature.com/articles/nmeth.4642.pdf

[4]http://www.fao.org/3/I9066EN/i9066en.pdf

[5]https://soilhealth.dac.gov.in/

[6]https://pib.gov.in/newsite/PrintRelease.aspx?relid=186413