AGRICULTURE CROP ANALYSIS USING HADOOP

Ramachandiran R,

ramachandiran08@gmail.com

Associate Professor,

Sri Manakula Vinayagar Engineering College,

Narmadha N,

narmadhanatraj24@gmail.com

Student,

Sri Manakula Vinayagar Engineering College,

ABSTRACT— Given their high vulnerability, developing countries have been particularly affected by the effects of weather change during in the recent years. In this way, providing knowledge and information could act as a strategic support service to help farmers become more resilient in their farming systems by enabling them to both adapt to and benefit from the new scenarios brought on by climate change as well as to take preventive measures. In order to prevent the effects of extreme weather occurrences, this study sought to understand the structure of interactions amongst farmers in the system. To do this, we want to comprehend the relations between farmers and the government as well as the different regions in weather patterns and farming methods.

Keywords:- **Agriculture; Machine Learning; Classifiers; K-Nearest Neighbor(KNN)**

I. INTRODUCTION

In order to prevent the effects of extreme weather situations, the study's focus was on knowing the structure of interactions between farmers and NAIS institutions regarding the transfer of climatic information. In order to achieve this, three variables are considered that have the potential to affect the network of relationships that has been observed: the institutions' roles, the ways in which farmers obtain information from them, and their regional behaviour, which is attached to the dominant farming systems, which are largely combinations of regions and the main agricultural products that are grown in such regions, with farms trying to serve as the system's central element (Meuwissen et al., 2019). Understanding these interconnections may have consequences for those who share the mission with policymakers, researchers, extension organisations, and the business sector.

The long-term, global climate change issue involves complex interactions between environmental, institutional, economic, social, and technological processes (Arnell et al., 2019; Fisher et al., 2002; Mendelsohn, 2008). Its negative impacts have a significant negative impact on the agricultural sector, which now has delayed two basic objectives of agriculture today: ending hunger and achieving sustainable development. Because developing countries have less financial resources to devote to combating it, they are more susceptible to its impacts, which are particularly severe. In fact, between 2008 and 2018, the damage brought on by extreme weather events in the poor world—one of the most obvious effects of climate change—cost roughly US$1700 billion. The farming sector, which is particularly prone to these situations.

Smallholders declared in Antwi-Agyei and Stringer's work (2021) that they used weather information (i.e., information related to long-term climate change impacts or to short-term impacts more related to changing weather patterns) to make critical farming decisions about the preparation of the land, the preference of crop varieties, changing cropping patterns, or planting time adjustments. Similarly, Gebrehiwot and Van Der Veen (2013) found that connected to climatic data increased the probability that farmers is using several crop varieties, implement irrigation and soil conservation techniques, and alter planting dates. It was also confirmed by Mulwa et al. (2017) and Ponce (2020) that farmers' decisions to adopt adaptation practises are highly altered by their capacity to receive climatic-related information.

A variety of enabling technologies provide the Internet of Things' sturdy backbone. Big Data, Embedded Systems, Cloud Computing, Wireless Sensor Networks, Security Protocols and Architectures, Communication Protocols, Web Services, Internet, and Search Engines are just a few examples of technologies that are discussed. A WSN is a wireless sensor network.is made up of a variety of sensors and nodes that work together to track different types of data. Cloud computing, commonly referred to as on-demand computing, is a category of Internet-based computing that makes pooled processing resources and data available instantly to computers and other devices. It can take many different forms, including IaaS, PaaS, SaaS, and DaaS.

Analyzing huge data sets with a variety of data kinds, or "Big Data," is the procedure for finding undiscovered patterns, unidentified relationships, market trends, customer preferences, and other pertinent business data. Communication protocols: These are the building blocks of IoT systems that enable connectivity and coupling to applications. By Providing formats for data communication, data encoding and data addressing, communication protocols also make it easier to share data over networks. Embedded systems are a type of computer system that combine hardware and software to carry.

II. RELATED WORK

K-NEAREST NEIGHBORS

The k-Nearest Neighbors method (k-NN) is a supervised classifier, like Naive Bayes. The accuracy of this algorithm is highly dependent to the value of k. If k is set too high, irrelevant data will be taken into account and thus, the accuracy will decrease. If k is set too low, such as k=1, the one neighbor that will be considered may be noise, and hence, the test data may be labeled incorrectly. In other words, if k is low, the information will be local, and if k is high, the information will be global. In k-NN, to predict the label of a new instance, its Euclidean distance to all of the instances should be calculated. Then, we take the k nearest neighbors and predict its label based on them. Algorithms like k-NN are called instance-based or lazy learners, because: 1) There are no models in these methods to apply to our data. The data should be always available to classify new instances. Hence, it is called instance based. 2) There is no training and testing in this algorithm and the classification is accomplished in one phase. Thus, it is called lazy.

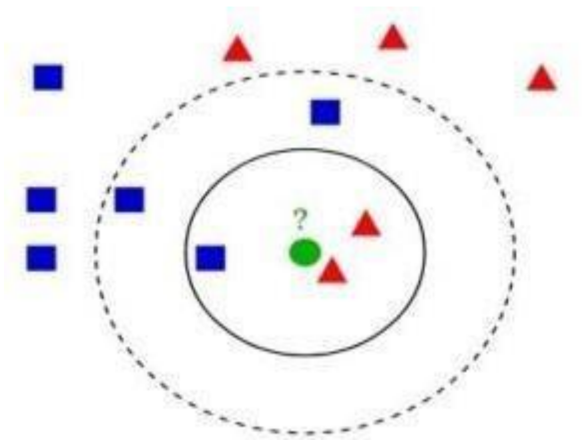

Figure 1 : KNN

III. DATA SET

A.MAPPER

The map module scans a data chunk and invokes the user-defined map function to process the input data. After generating the intermediate results (a set of key/value pairs), it groups the results based on the partition keys, sorts the tuples in each partition, and notifies the master node about the positions of the results.

B.REDUCE

The reduce module pulls data from the mappers after receiving the notification from the master. Once all intermediate results are obtained from the mappers, the reducer merges the data by keys and all values with the same key are grouped together. Finally, the user-defined function is applied to each key/value pair, and the results are output to DFS.

IV. METHODOLOGY

The current work's main objective is to successfully analyse massive data in attack detection systems. The data collection from an attack detection system, which contains both common and odd packets, is used to undertake a full experimental investigation. The data is preprocessed using the information entropy approach. The information entropy method is used to construct faster and more accurate devices. The hadoop(KNN) Classifier technique is frequently used to divide the input into dataset. The results of the classifier are evaluated using the efficiency measures. Finally, an examination of the comparison between the suggested and several existing classifiers is offered. Thefollowing subsection typically contains a detailed discussionof each step used in the particular suggested model.

V. PREPROCESSING

Preprocessing is a critical stage in the organization of real-world datasets into a comprehensible manner. Undoubtedly, the real-world datasets have been noisy and sparse in several behaviors. For the purpose of analyzing large-scale data trends, preprocessing is essential. In order to enhance the machine learning method for pattern categorization in massive data intrusion detection systems, preprocessing techniques are consequently necessary. In order to further increase the accuracy and effectiveness of the resultant machine learning work. The information gain method is employed in the current research study to extract important features from the data collection. The next subsections provide a full description of the information gain method. The values in the dataset should be accurate and not null in order to apply the machine learning method. For some machine learning models, data in a specific format is required because the KNN approach does not allow null values.

VI. INFORAMTION GAIN METHOD

The aim of this study was to understand the structure of the interactions between farmers and the transfer of information on preventing the effects of extreme weather events.To do so, we aim to understand how farmers and institutions are connected and how the climate and farming systems' particularities of the territories, the main roles of the institutions, and the media through which farmers get information from each of them could influence such information transfer processes. This study identifies how actors involved in preventive information transfer are connected, quantifies their relevance, and detects deficiencies in such transfer processes. Additionally, we detected some groups of farmers who rely on other institutions too, with regional differences in access to information. Results suggest that there is room for improvement regarding the transfer of information on preventing extreme weather events.

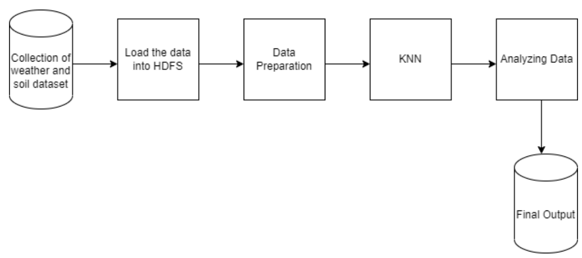

Figure 2 : Proposed System For KNN Method

VII. KNN ALGORITHM

K-nearest Neighbour (KNN) to predict the risk of cerebral infarction disease. For T-data, we propose KNN- based unimodal disease risk prediction (KNN-UDRP) algorithm to predict the risk of cerebral infarction disease. In the remaining of the paper, let KNN(T-data) denote the KNN algorithm used for T-data. For S&T data, we predict the risk of cerebral infarction disease by the use of KNN algorithm, which is denoted by KNN (S&T-data) for the sake of simplicity.

VIII. EXPERIMENTAL SETUP

In the existing system, the effects of climate change have impacted heavily on the agricultural sector, particularly in developing countries provided their high vulnerability. In this sense, knowledge and information transference could act as a strategic support service to improve the resilience of their farming systems, as this would help farmers to adapt and take advantage of the new scenarios brought by climate change, as well as to take preventive actions.

DISADVANTAGES OF EXISTING SYSTEM

• Radio and TV are the predominant media to reach smallholders.

• Problems with the contents of information, its perception and a lack of resources may prevent a wider information transfer.

PROPOSED SYSTEM

The purpose of this study was to comprehend the framework of farmer relationships and the dissemination of knowledge about mitigating the effects of extreme weather occurrences. In order to do this, it is important for us to understand the relationships between farmers and institutions as well as how local climate and farming practises, institutional functions, and the media used by farmers to receive information from each of these entities may all have an impact on how information is transferred. This study identifies the relationships between the people involved in the transfer of preventive knowledge, measures their importance, and identifies flaws in these transfer procedures. Additionally, we identified certain farming communities that depend on other institutions as well, with regional variations in information availability.

ADVANTAGES OF PROPOSED SYSTEM

These mainly concern issues with trust among institutions and farmers, which are related, among other things, towards how farmers view formal information. They also concern issues with getting and utilising the information. The agricultural industry in Peru might become more resilient to climate change by addressing these challenges by assisting more farmers in receiving official information on mitigating its effects and knowing how to apply it. More in-depth analysis and more data would be required to support these first findings.

IX. COMPARISON AND PERFORMANCE OFPROPOSED MODEL

The purpose of this study was to comprehend the framework of farmer relationships and the dissemination of knowledge about mitigating the effects of extreme weather occurrences. In order to do this, it is important for us to understand the relationships between farmers and institutions as well as how local climate and farming practises, institutional functions, and the media used by farmers to receive information from each of these entities may all have an impact on how information is transferred. This study identifies the relationships between the people involved in the transfer of preventive knowledge, measures their importance, and identifies flaws in these transfer procedures. Additionally, we identified certain farming communities that depend on other institutions as well, with regional variations in information availability.

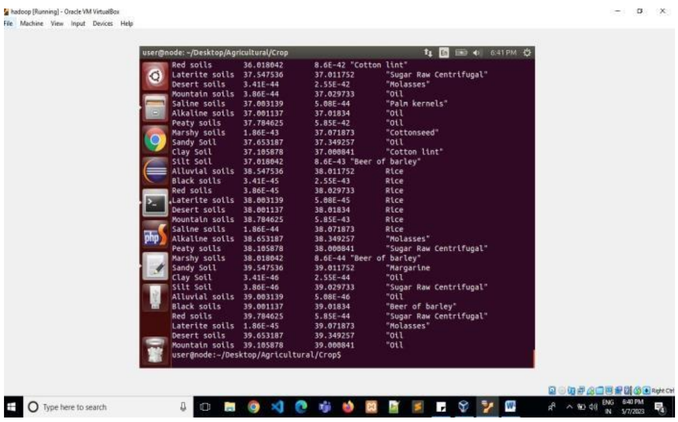

Figure 3 : Proposed model's performancecompared to current models

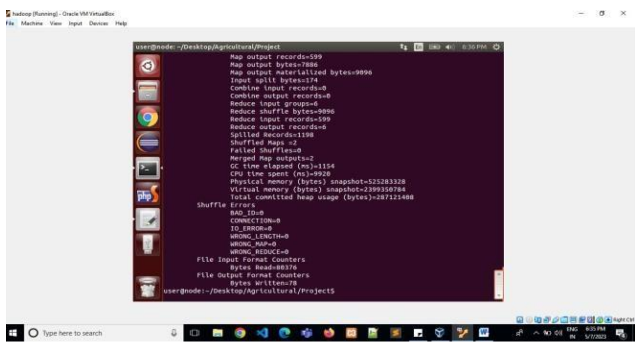

Figure 4 : Proposed model's performancecompared to current models

Figure 5 : Proposed model's performancecompared to 3D Bar.

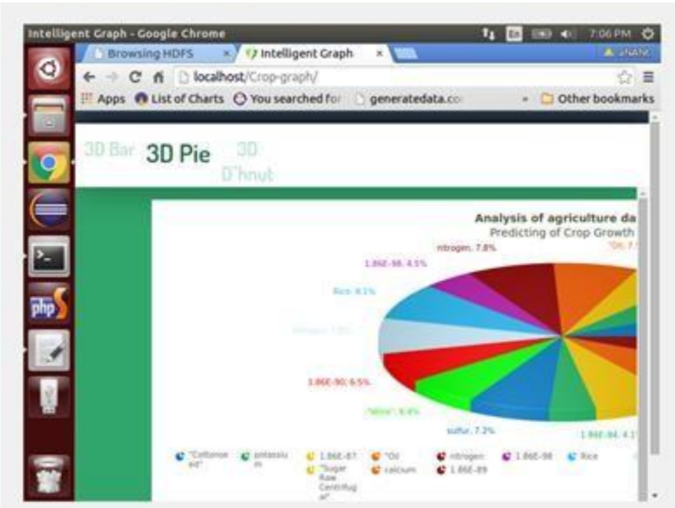

Figure 6 : Proposed model's performancecompared to 3D Pie.

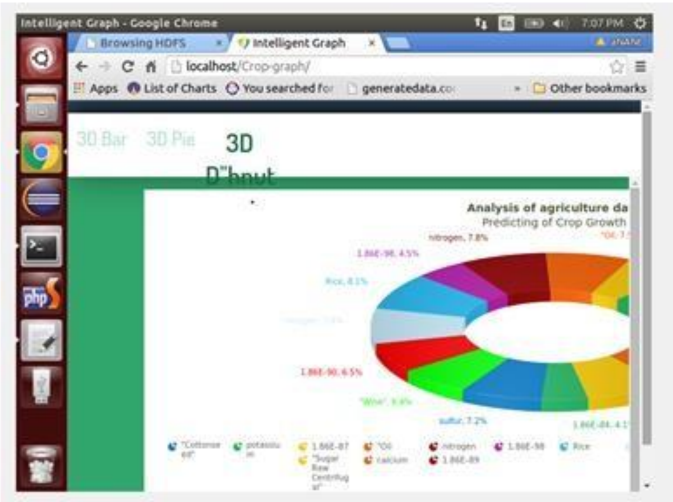

Figure 7 : Proposed model's performancecompared to 3D –Hurt.

X. CONCLUSION

This thesis focuses on analyzing the agricultural soil data using data mining for crop yield prediction The Association rule mining algorithms must perform efficiently. For a given soil data the soil physical factors are grouped using KNN. In future the work can be expanded and enhanced using the climatic factors, crop prediction, fertilization techniques based on seasons edict specific crop on various soil.

XI. REFERENCES

1. Aguilar-Gallegos, N., Munoz-Rodríguez, ˜ M., Santoyo-Cort´es, H., Aguilar-Avila, ´ J., Klerkx, L., 2015. Information networks that generate economic value: A study on clusters of adopters of new or improved technologies and practices among oil palm growers in Mexico. Agric. Syst. 135, 122–132. https://doi.org/10.1016/j. agsy.2015.01.003.

2. Aguilera, E., Díaz-Gaona, C., García-Laureano, R., Reyes-Palomo, C., Guzm´ an, G.I., Ortolani, L., S´ anchez-Rodríguez, M., Rodríguez-Est´evez, V., 2020. Agroecology for adaptation to climate change and resource depletion in the Mediterranean region. A review. Agric. Syst. 181 (February), 102809 https://doi.org/10.1016/j. agsy.2020.102809.

3. Aker, J.C., 2011. Dial –A‖ for agriculture: A review of information and communication technologies for agricultural extension in developing countries. Agric. Econ. 42 (6), 631–647. https://doi.org/10.1111/j.1574-0862.2011.00545.x. Altea, L., 2019.

4. Perceptions of climate change and its impacts: a comparison between farmers and institutions in the Amazonas Region of Peru. Clim. Dev. 0 (0), 1–13. https://doi.org/10.1080/17565529.2019.1605285. Antwi-agyei, P., Stringer, L.C., 2021.

5. Improving the effectiveness of agricultural extension services in supporting farmers to adapt to climate change: insights from northeastern Ghana. Clim. Risk Manag. 1–30 https://doi.org/10.1016/j. crm.2021.100304.

6. Antwi-Agyei, P., Dougill, A.J., Abaidoo, R.C., 2021. Opportunities and barriers for using climate information for building resilient agricultural systems in Sudan savannah agro-ecological zone of North-Eastern Ghana. Clim. Services 22, 100226. https:// doi.org/10.1016/j.cliser.2021.100226.

7. Arnell, N.W., Lowe, J.A., Bernie, D., Nicholls, R.J., Brown, S., Challinor, A.J., Osborn, T. J., 2019. The global and regional impacts of climate change under representative concentration pathway forcings and shared socioeconomic pathway socioeconomic scenarios. Environ. Res. Lett. 14 (8) https://doi.org/10.1088/1748-9326/ab35a6.

8. Bergmann, J., Vinke, K., Fernandez, ´ C.A., Gornott, C., Gleixner, S., Laudien, R., Lobanova, A., Ludescher, J., Schellnhuber, H.J., 2021. Assessing the evidence: climate change and migration in Peru (issue august). https://publications.iom.int/s ystem/files/pdf/assessing-the-evidence-peru.pdf. Bhatia, P.S., Iovleff, S., Govaert, G., 2017.

9. Blockcluster: an R package for model-based coclustering. J. Stat. Softw. 76 (1) https://doi.org/10.18637/jss.v076.i09. BID, & CEPAL, 2014. La economía del cambio climatico ´ en el Perú. In: Ludena, ˜ C.E., S´ anchez-Aragon, ´ L., de Miguel, C., Martínez, K., Pereira, M. (Eds.), Monografía BID No. 222 y CEPAL LC/W.640. https://publications.iadb.org/es/la-economia-del-c ambio-climatico-en-el-peru. Borgatti, S.P., Halgin, D.S., 2011.

10. Borgatti, S.P., Halgin, D.S., 2011. Analyzing affiliation networks. In: Scott, J., Carrington, P. (Eds.), The SAGE Handbook of Social Network Analysis, pp. 417–433. https://doi.org/10.4135/9781446294413.n28. Issue 1941.