isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2024 ,ISSN_NO: 2321-3337

MULTI FACTOR DATA ANALYSIS FOR EFFECTIVE CROP CULTIVATION SYSTEM USING BIG DATA ANALYSIS

Anjali R[1], Charumathi J[2], TejaSri V M[3], Dr Kamal N[4]*

[123] UG Scholar-Dept CSE, Grt Institute of Engineering and Technology Tiruttani, India.

[4*]Professor-Dept CSE, Grt Institute of Engineering and Technology Tiruttani, India.

anjali265580@gmail.com , jcharuprakash@gmail.com , vmtejasri@gmail.com

[4]*Corresponding Author: kamal.n@grt.edu.in

Abstract

In today's rapidly advancing technological landscape, the agriculture sector stands at the forefront of innovation, striving to address the complex challenges faced by farmers worldwide. This paper presents a multifactor analysis framework leveraging machine learning and big data methodologies to optimize crop selection based on key determinants including rainfall patterns, soil characteristics, and geographic location. By harnessing vast datasets and sophisticated algorithms, our approach aims to mitigate the impact of natural disasters and financial constraints on agricultural productivity. Through the integration of predictive models and real-time data analytics, we provide actionable insights to empower farmers in decision-making processes, ultimately enhancing crop yield and fostering economic sustainability in farming communities. This research underscores the pivotal role of technology in revolutionizing agricultural practices and ensuring food security for future generations.

Keywords: *Machine Learning, Big Data, Data Analytics, Multifactor Analysis.*

1. Introduction

The primary idea behind the project is to recommend the best crop to farmers based on big data concepts and multi-factor analysis. The different criteria include rainfall data from the previous year, the kind of soil, a list of crops that may be grown there, and the water level needed to nurture those crops. We are implementing the big data notion in our project since each produced concept need to be scalable. Our program has to be scalable if we wish to add more parameters and apply the same idea to more than one location. We implemented this application on big data to ensure scalability. We can state with certainty that Framers are better protected against financial and natural disasters thanks to the implementation of this project. In most cases, farmers won't account for all of these variables, and there aren't any automated processing tools available to support them. In order to get effective results, we develop this concept utilizing the Support Vector Algorithm.

2. Related Work

Humanity will require more food from less land and water resources during the coming decades. The effects of four potential development scenarios on food production are quantified in this study using data from the Special Report on Emission Scenarios and the Millennium Ecosystem Assessment. The effects of population growth and technological advancement on the availability of land and water, as well as the changes in demand for agricultural and forest products due to these factors, are all taken into account, albeit in part and in tandem. Dynamic elasticities are used in the computation of the effects of income on food demand. Global, partial equilibrium model simulations of the forestry and agricultural sectors demonstrate that all development scenarios under consideration result in increases in per capita food levels, with only small effects on food costs. There will be a 14% increase in agricultural land worldwide between 2010 and 2030. Restrictions on deforestation have a significant effect on the cost of land and water resources, but they have minimal effect on the amount of food produced globally and its price. Population expansion results in the largest increase in overall food production, whereas predicted income changes have the greatest partial influence on per capita food consumption levels. Adaptations in land management intensities either increase or decrease the impact of technical change. [1]

Knowledge produced by agricultural systems science enables academics to tackle challenging issues or make wise agricultural decisions. This science's long history serves as an example of the variety of systems and scales that it operates and has been investigated on. Scientists from a wide range of disciplines have contributed concepts and tools for modeling, a fundamental tool in agricultural systems science, throughout more than six decades. It is crucial to reflect on this history and its lessons as agricultural scientists now consider the "next generation" models, data, and knowledge products required to address the progressively complex systems problems that society faces. By doing so, we can make sure that we avoid re-inventing the wheel and work to take into account all aspects of related challenges. In order to achieve this, we provide a brief overview of the development of agricultural systems modeling throughout history and highlight key takeaways that will inform the creation of new instruments and techniques for agricultural systems. Agricultural system modeling has evolved as a result of several historical occurrences as well as general technological advancements in other domains. These include the creation of process-based bio-physical models of crops and livestock, statistical models derived from historical data, and economic optimization and simulation models at local, regional, and global scales. The characteristics of models of agricultural systems have been considerably diversified, contingent upon the sizes of the systems involved and the diverse range of goals driving their development and application across disciplinary boundaries by researchers. It appears that the groundwork has been laid for the significant advancements in agricultural systems science required for the creation of the upcoming models, databases, knowledge products, and decision support systems. These advances are expected to come from recent trends in increased collaboration between academic institutions, disciplines, and the public and private sectors. The community should take historical lessons into account when creating this next generation of agricultural systems models in order to help prevent obstacles and traps .[2]

In recent decades, data mining has been a prominent research area aimed at extracting usable and implicit knowledge. Humans are easily able to absorb this knowledge. Initially, statistical methods were used to manually compute and assess this knowledge extraction. The development of technology led to the subsequent emergence of semi-automated data mining techniques. Another example of this progress was in storage, which raises the need for analysis. In these situations, semi-automated methods have lost their effectiveness. As a result, automated data mining techniques were developed for effective knowledge synthesis. This study presents an overview of the existing literature on pattern recognition and data mining for soil data mining. Agricultural soil dataset data mining is a relatively new area of study. Data mining can be used to discover and customize effective methods for handling complicated soil datasets .[3]

The factors influencing the acceptance of novel agricultural practices are well understood, but there hasn't been much work done to create quantitative models of adoption that may be used to predict adoption in advance for those organizing policy, research, development, and extension activities in the field of agriculture. The outcome of such an attempt is ADOPT (Adoption and Diffusion Outcome Prediction Tool), which estimates the significance of several factors influencing adoption and provides projections of a practice's expected rate and peak level of adoption. It makes use of a conceptual framework that takes into account a number of factors, such as those pertaining to risk, economics, environmental effects, farmer networks, farm and farmer characteristics, and the new practice's ease of use and convenience. A key component is the capacity to ascertain the relative benefit of the practice based on attributes of the practice and possible users. ADOPT users answer 22 questions about the following topics: a) the practice's characteristics that influence its relative advantage; b) the population's characteristics that influence their perceptions of the practice's relative advantage; c) the practice's characteristics that influence how quickly and easily people can learn about it; and d) the characteristics of potential adopters that influence their capacity to learn about the practice. ADOPT predicts the practice's diffusion curve and offers sensitivity studies of the variables affecting adoption's rate and peak level. This study describes the model and uses examples of new crop kinds, new cropping technologies, and grazing alternatives to show how well it can predict the diffusion of agricultural practices. ADOPT is intended to improve conceptual knowledge and consideration of the adoption process by people working in agricultural research, development, extension, and policy, in addition to offering forecasts .[4]

Supervised machine learning (ML) approaches have been utilized more and more recently to analyze remote sensing (RS) observation data for crop production prediction, thanks to the development of satellite missions and artificial intelligence tools. However, supervised machine learning models typically have poor spatial transferability because of the domain shift between diverse locations Because of this, models that are trained using labelled data from one spatial location (the source domain) frequently become invalid when they are applied directly to another (the target domain). We suggested a multisource maximum predictor discrepancy (MMPD) neural network, an unsupervised domain adaptation (UDA) method for county-level corn production prediction, to solve this problem. The study's novelties include the following: 1) we suggested using crop yield response in the target domain to align source and target domains and maximize the discrepancy between two source-specific yield predictors; and 2) we used the multisource UDA strategy to prevent negative interference between labelled samples from various sources. Case studies in Argentina and the U.S. corn belt showed that the suggested MMPD model beat a number of other cutting-edge deep learning (DL) and UDA techniques and successfully decreased domain changes .[5]

3. Objective

This project's primary goal is to advise farmers on the best crops to plant. A number of interfering factors, such as the kind of soil, the amount of water needed to cultivate that specific crop, and the amount of rain that fell the year before, can affect the growth of any crop. A farmer will eventually cease farming if they suffer any losses, which would lead to a shortage of food and grains. We have begun work on this project with the goal of protecting the lives of farmers.

4. Proposed System

We incorporate an application in our suggested method to determine the types of soil and if the property receives its water from boreholes or from rainfall and recommend a crop that would be appropriate for that soil. Thus, using this program, we can eventually recommend a beat crop after thoroughly analysing a number of factors, including soil type, list of crops that may be grown there, rainfall data from previous years, and water level requirements. Farmers are the key people we can save, along with framing techniques. We forecast the kind of crop that would thrive in that specific soil, as well as the temperature and other factors. In order to determine the crop for the appropriate soil, we are employing machine learning with a set of datasets. In order to determine the crop for the appropriate soil, we are employing machine learning with a set of datasets.

isrjournals
International Journal of Advanced Research in Computer Science Engineering and Information Technology
Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2024 ,ISSN_NO: 2321-3337

5. Architecture Diagram

The suggested system, which is dependent on where we gather the datasets. Following the dataset's extraction, transformation, loading, and processing. The processed dataset is divided into train and test sets, which are then subjected to processing and analysis via the Supported Vector Machine-machine learning algorithm. Big Data analysis is being used to analyze a variety of metrics in the HDFS (Hadoop Distributed File System) system, including temperature, water requirements, soil type, and rainfall data from previous years, in order to forecast the current problem condition. indicating, at last, the ideal crop to grow in that particular area.

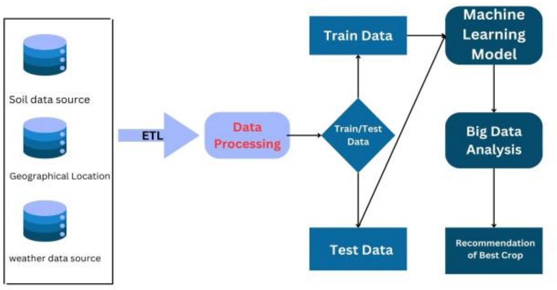

Fig.5.1: Architecture Diagram

6. Algorithm

Backing An n-dimensional space is classified by Vector Machines, a type of discriminative algorithms, by identifying the virtual line, or hyperplane, that best divides the classes into distinct groups. For instance, you might use a scatterplot to show all data if you had two features, and the algorithm would draw a line dividing the classes. The maximum margin is the largest minimum distance, or "best," that the SVM algorithm finds in the feature space between observations belonging to the two classes. The distance measured perpendicularly from the hyper plane to only the nearest points is used to compute the margin. These are the only points that matter for defining the hyper plane and building the classifier. We refer to them as the support vectors.

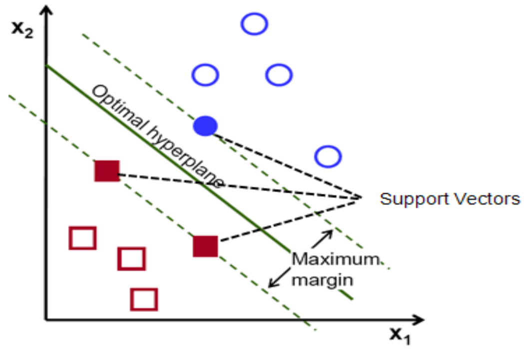

Fig-6.1: Support Vector Machine

7. Implementation

7.1 User Interface Design Module

The Farmer can enter their location and other details into our application with the help of the User Interface Design. Following the farmer's input, our application begins processing the data and compares each factor—such as location and rainfall statistics—with a pre-programmed trained data set. Ultimately, this GUI suggests the Best Crop as the Output after evaluating Multi Variant variables using Big Data. We utilize MSQL as a back end and Java to construct our application. The only thing that will go through this IDE is all input and output.

7.2 Multifactor Analysis Module

In this module, we will process the farmer's inputs appropriately by comparing all the data saved in the server-side dataset. Our application begins processing by comparing the trained data set from the backend as soon as the Farmer provides location data. In order to process the data and retrieve the result, we employ the Support Vector technique. The farmer's location information is first compared with rainfall data from the previous year for that particular location. Following the achievement of the rainfall prediction, the soil type for that location is compared, followed by the types of crops that can be grown there specifically, and lastly the amount of water needed to nurture those crops is computed. Our application will give the desired farmer

isrjournals
International Journal of Advanced Research in Computer Science Engineering and Information Technology
Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2024 ,ISSN_NO: 2321-3337

the best crop to plant based on all of these factors.

7.3 Best Crop Suggestion Module

The system in this module will contrast the fresh input with the data from the training set. following thorough analysis of the trained data and comparison with the farmer's inputs. Our application determines the best crop to grow in a given location by analyzing all of the trained data sets, including rainfall data from prior years, soil types, crops that may be grown there, and the amount of water those crops require.

8. Experimental Results

This result discusses about the implementation of the multi factor data analysis for effective crop cultivation system using big data analysis for various cases are identified and the below Fig. 8.5., Fig. 8.6., Fig. 8.7., Fig. 8.8., shows the implementation of the user input for the recommendation of the best crop to be cultivated. Fig. 8.1., Fig. 8.2., Fig. 8.3., Fig. 8.4., shows the user input and registration detail for the processing of previous year datasets. The final result will be shown in the Fig. 8.9.

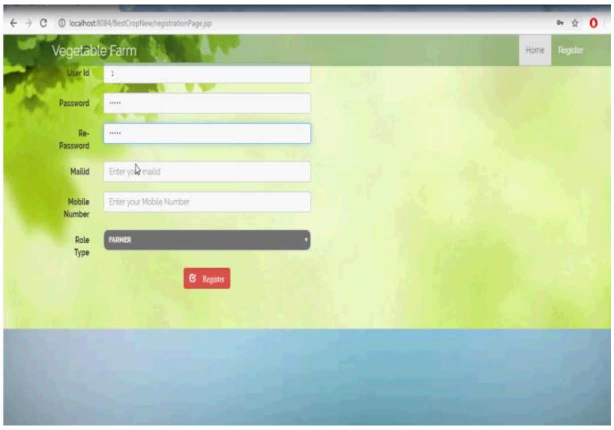

Fig.8.2. User Registration

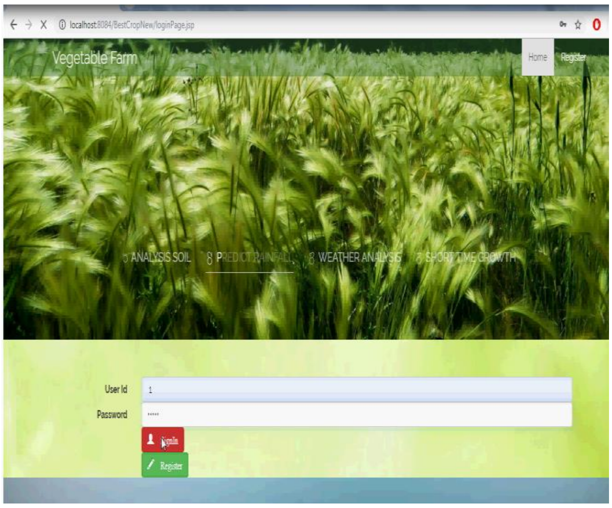

Fig.8.3. User Login

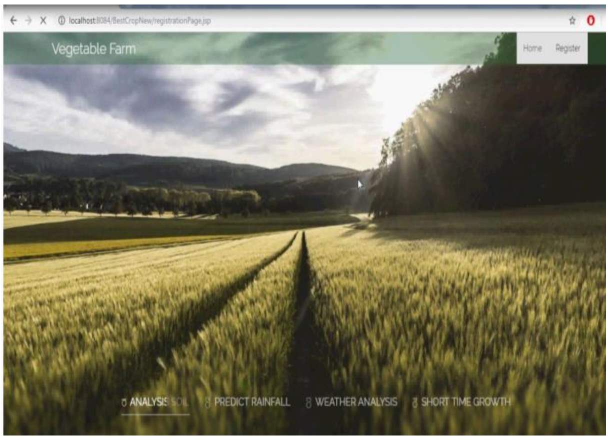

Fig.8.1. Application Frontend

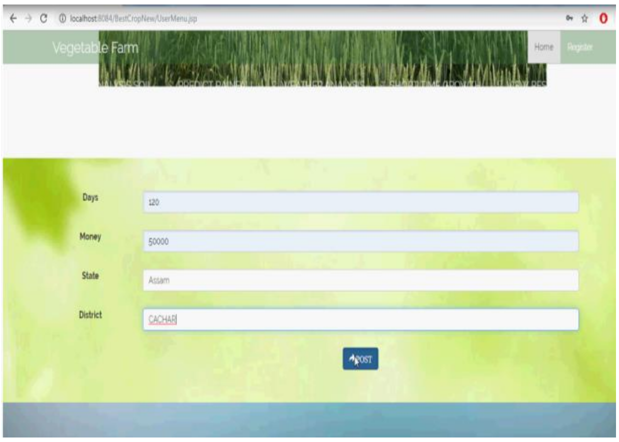

Fig.8.4. User Input

isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2024 ,ISSN_NO: 2321-3337

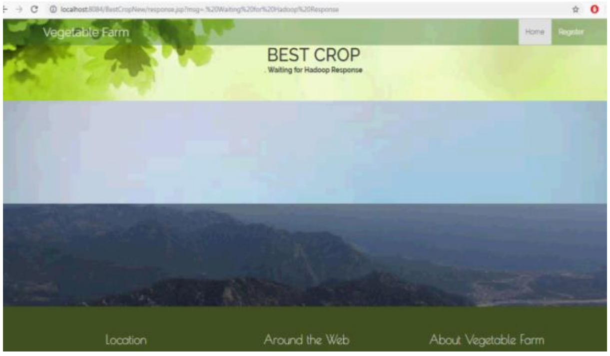

Fig.8.5. Big Data Hadoop Processing

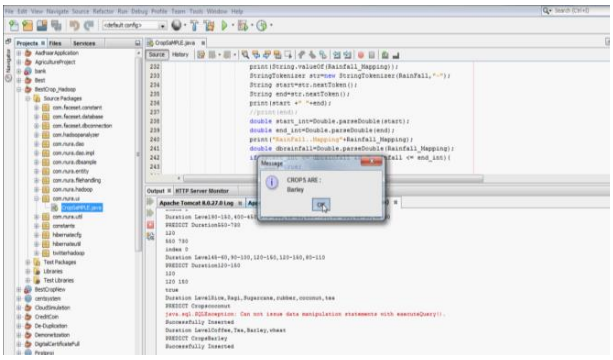

Fig.8.7. Best Crop Recommendation

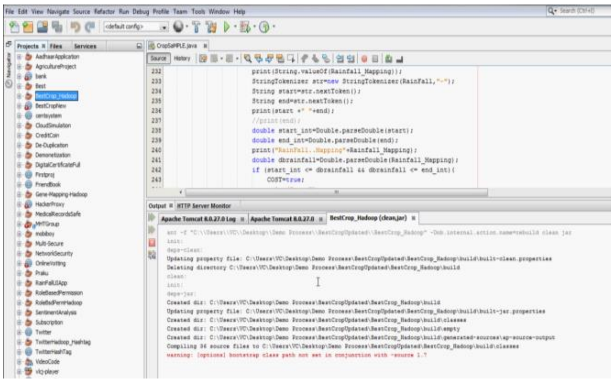

Fig.8.5.1. Hadoop Server Startup

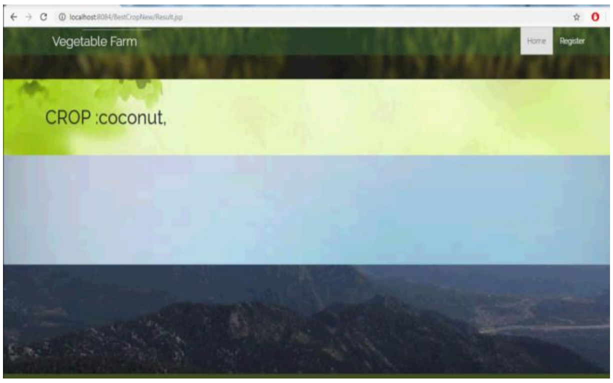

Fig.8.8. Output in Web Interface

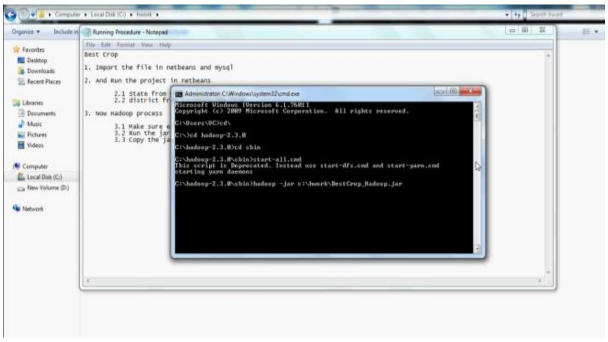

Fig.8.5.2. Hadoop Server Startup

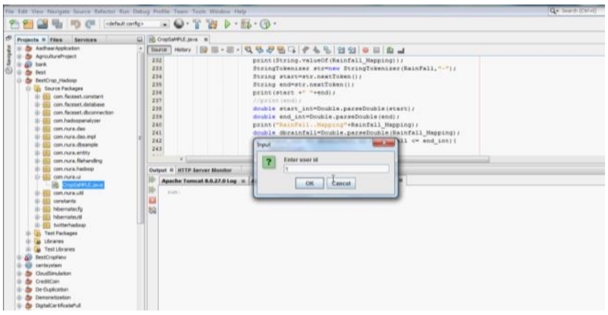

Fig.8.6. User Logs into Our Application

9. Conclusion & Future work

By comparing numerous parameters of a redefined dataset with the farmer's input, we have implemented a Crop Recommendation and Prediction system using our application and the Big Data Approach. We can reassure farmers that Best Crops can be suggested by using this interface/application. This will prevent the farmer from losing their source of income, and the application itself will encourage them to put in more hours.

We are able to link several fields at various locations so that all farmers, regardless of where they live, can use our project. For future use, all data can be kept on a cloud server. Block chain technology can be used to enhance security procedures.

Reference

[1] Uwe A. Schneider a,⇑, Petr Havlik b, Erwin Schmid c, Hugo Valin b, Aline Mosnier

b,c, Michael Obersteiner b,Hannes Bottcher b, Rastislav Skalsky´ d, Juraj Balkovicˇ d, Timm Sauer a, Steffen Fritz b" Impacts of population growth, economic development, and technical change on global food production and consumption" Agricultural Systems 104 (2011) 204–215 inelsvier

[2] James W. Jones a, ⁎, John M. Antle b, Bruno O. Basso c, Kenneth J. Boote a, Richard T. Conant d, Ian Foster e, H. Charles J. Godfray f, Mario Herrero g, Richard E. Howitt h, Sander Jansseni, Brian A. Keating g, Rafael Munoz-Carpena a, Cheryl H. Porter a, Cynthia Rosenzweig j, Tim R. Wheeler k "Brief history of agricultural systems modeling" in science direct.

[3] Dr. D. Ashok Kumar*1 and N. Kannathasan#2. A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining

[4] Geoff Kuehnea, Rick Llewellyna, ⁎, David J. Pannellb, Roger Wilkinsonc, Perry Dollingd, Jackie Ouzmana, Mike Ewinge," Predicting farmer uptake of new agricultural practices: A tool for research, extension and policy" in Elsevier science direct.

[5] Wahbeh, A. H., Al-Radaideh, Q. A., Al-Kabi, M. N., & Al- Shawakfa, E. M. (2011). A comparison study between data mining tools over some classification methods. International Journal of Advanced Computer Science and Applications, 8(2),18-26.

[6] Zhou, S., Ling, T. W., Guan, J., Hu, J., & Zhou, A. (2003, March). Fast text classification: a training-corpus pruning based approach. In Database Systems for Advanced Applications, year 2003. (DASFAA 2003). Proceedings. Eighth International Conference on (pp. 127-136). IEEE.

[7] Li, Y., & Bontcheva, K. year (2008). Adapting support vector machines for f-term-based classification of patents. ACM Transactions on Asian Language Information Processing (TALIP), 7(2), 7.

[8] Eiben, A. E., Raue, P. E., & Ruttkay, Z. (1994, October). Genetic algorithms with multi-parent recombination. In International Conference on Parallel Problem Solving from Nature (pp. 78-87). Springer, Berlin, Heidelberg.

[9] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37.

[10] Crone, S. F., Lessmann, S., & Stahlbock, R. (2006). The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. European Journal of Operational Research, 173(3), 781-800.

[11] Tubiello, F. N., Salvatore, M., Cóndor Golec, R. D., Ferrara, A., Rossi, S., Biancalani, R., ... & Flammini, A. (2014). Agriculture, forestry and other land use emissions by sources and removals by sinks. Rome, Italy.

[12] Agriculture Statistics of Pakistan, Pakistan Bureau of Statistical, Retrieved 10 September, at the year of 2016 by http://www.pbs.gov.pk/content/agriculture-statistics

[13] Doran, J. W., & Parkin, T. B. (1994). Defining and assessing soil quality. Defining soil quality for a sustainable environment, (definingsoilqua), 1-21.

[14] Larose, D. T. year (2014). Discovering knowledge in data: an introduction to data mining. John Wiley & Sons.

[15] Eiben, A. E., Raue, P. E., & Ruttkay, Z. (1994, October). Genetic algorithms with multi-parent recombination. In International Conference on Parallel Problem Solving from Nature (pp. 78-87). Springer, Berlin, Heidelberg.