International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 09 Issue 04 April 2021, Page no. – 2255-2258
Index Copernicus ICV: 57.55, Impact Factor: 7.184
DOI: 10.47191/ijmcr/v9i4.05

Survey on Smart Agriculture using Data Science

Madhuri Malode[1], Manisha Thombare[2], Dipika Tidke[3], Reena More[4]

[1,2,3,4] Assistant Professor, MVPS's KBTCOE, Nashik

ARTICLE INFO

Published Online:
17 April 2021

ABSTRACT

Data analytics can be used in the majority to obtain data that can be used by farmers to enhance their farming. Farmers can generate smart farming decisions using that data throughout the production cycle of any crop. Nowadays, the idea of data analytics has been widely used in various sectors for well-organized decision making and people are gradually responding to the importance of the value of data analytics. Recent technologies are now able to gain plenty of data to farmers regarding agricultural activities, which can then be studied to find significant facts. Machine learning concept is used in a variety of applications like banking, health care, agriculture sector, education, and many more. While converting data in a proper format, the main focus is on the conversion of data from unstructured format to structured format. The data is then provided as input to machine learning algorithms. Using the concept of data analytics, i.e using efficient decision-making methods, numerous options is made available to farmers for profitable farming. It will thus lead to the economic growth of various farmers. In this paper, we tried to propose a platform of data analytics, where we put various ideas to farmers depending on their needs in the agriculture sector. The outcomes of this study would certify the proposed platform and also - future research on innovative decision-making in the field of agriculture.

Corresponding Author:
Madhuri Malode

KEYWORDS: Data Analytics; Smart Agriculture; Machine Learning Algorithms

I. INTRODUCTION

As per the 2014 FAO world agriculture data, world's largest manufacturer of many fresh fruits like banana, mango, guava, papaya, lemon and vegetables like chickpea, okra, and milk, major spices like chili pepper, ginger, fibrous crops such as jute, staples such as millets and castor oil seed is India. Among, the world's major food staple, India is the second-largest manufacturer of wheat and rice. One of factor of India's economic growth is agriculture. Since few Decades, Agriculture in India has mislaid its impact over world Agriculture growth. There are many reasons for the shortcoming of agriculture in India like lack of education, lack of advanced technical knowledge [7].

From the past few decades, in the agriculture field, researchers have considered and suggested different algorithms and techniques which would be beneficial for the farmers to increase the crop as well as help them to increase the profit financially. Applications related to agriculture based on Machine Learning algorithms provide different ideas like irrigation based on soil moisture content using various sensors, improve crop yield productivity, efficient use of pesticide and fertilizers and avoid excessive cost related to harvesting. Nowadays for effective decision-making purposes, various organizations are using Spatial Data mining techniques to extract useful information from spatial data sets. Predictive analytics supports the farmers to make smart decisions in farming by collecting internal, external intelligent data and real-time data from field sensors [1].

II. RELATED WORK

Dr.K.Sumathi et.al. [1] proposed the support for the stakeholders involved in agriculture. They presented a platform of data analytics for intelligent agriculture. By using the systems farmer can get the details of crop maintaining, irrigation required as per the moisture and temperature of the field, pesticide details. Also the farmers can get the updates regarding weather forecast, agriculture goods availability in market.

Jintao Huang & Lichen Zhang (2017)[2] presented their paper on data processing platforms for intelligent agriculture and this research paper is kept as the basis for the proposed

platform. The paper also elaborates on the varied processing types namely Hadoop, storm stream, and Cassandra, and envisages the point on how the agricultural stakes would be beneficial from the stage of production till the end.

M. JoharanBeevi and C. Vijayakumaran (2016)[3] has thrown light on the same research and have conducted a that shows the results validating the point on terminologies about big data and several frameworks that are used in various fields. But more specifically it highlights the usage and appropriateness of Hadoop for the agricultural sector.

G. RajeshKumar et al.,(2015) [4] In economic growth and food security of the agro-based country, Agriculture plays an important role. Crop determination is the main issue from agriculture planning. Various parameters are used in the Agriculture industry that means it depends upon the productivity rate, market rate, market farm, government controls, and policies. Using machine learning techniques yield rate of the crop, determination of weather conditions, soil conditions, and classification crops are determined.

Maria. B &Vassilios.D (2013)[5] in their research work have included the study based on appropriateness of ICT procedure in the agricultural sector. The pilot research conducted is presented in this paper.KMW recently launched their innovative farm machine, MEGA T, which has the charge of a tiller and the functionality, comfort, and ease of use of a tractor. Mega T has been considered to address four main regions of concern, viz. ease, security, simple of use and pride of ownership. These were found deficient in regular power tillers that are walk-behind machines that are dull, hard to run, and insecure. The Mega T is lighter in heaviness and smaller in size, yet just as influential. Moreover, it can increase efficiency in areas where larger machines cannot be used.

III. SYSTEM ARCHITECTURE

A. Introduction

What is big data?

Big data means large and complex data sets that are processed and analyzed to uncover valuable information which is useful for business and organizations.

Characteristics of Big Data

In 2001, Gartner analyst Doug Laney planned the 3 'V's of Big Data – Variety, Velocity, and Volume. Discussion on characteristics of big data:

1. Variety:

Variety of big data is the data gathered from various sources which can be structured, semi-structured, or unstructured. In the past data was collected from spreadsheets and databases. but nowadays data is collected from various resources such as emails, photos, PDFs, videos, audios, social media posts and so much.

2. Velocity:

Velocity is the speed at which data is being created in real-time.

3. Volume:

A large Volume of data is considered as Big Data. The data is generated daily from various sources like social media, business processes, human interactions, machines, and all. This generated data is stored in data warehouses.

Types of big data

There is 3 types of big data

Structured

Structured data is processed, stored, and retrieved in a structured manner. just like the student information stored in a table in the student database as student roll number, name, and other details. The data can be searched easily by search engine algorithms.

Unstructured

There is no specific form or structure to this type of data. So it is very difficult to analyze and process this type of data. An example of this is email.

Semi-structured data

The data having both structured and unstructured formats are semi-structured data. The data does not obey the tabular structure of data models but it contains tags, markers to separate semantic elements, and hierarchies of records.

Data Analytics

Data analytics is the science of evaluating raw data to create results about the data. Data analytics deals with the process of generating valid output based on various results already generated by researchers. In application areas like industries, organizations use data analytics concepts to make efficient decisions as well as to update present models and theories.

Data science

Data science consists of operations like data cleaning, data preparation, and analysis of structured and unstructured data. Data science deals with different concepts of statistics, mathematics, programming, problem-solving, and capturing [6].

B. Framework

In this paper, the concept of a data analytics platform for intelligent agriculture is discussed. This platform consists of four phases as Data discovery phase, Data Preparation Phase, Data Analytics Phase, Visualization Phase.

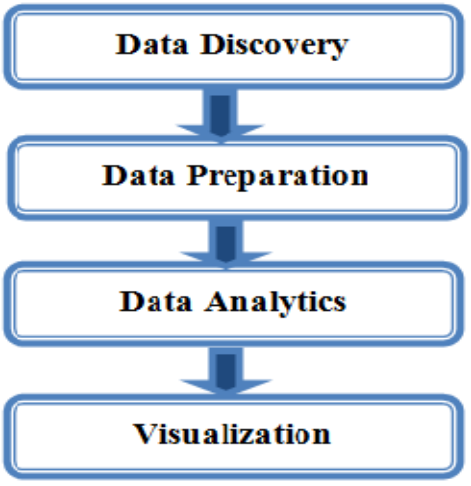

Fig. I. Framework

- **Data Discovery Phase**

In this phase, we define the purpose of the data and process to achieve the data throughout the process of Data Analytics. Firstly, we identify the important objectives by going through the available data. During this, we check whether we have worked on similar projects before or not. If so, we can refer to such projects to understand the ongoing project. In this phase, the team calculates technology, people, data, and time. Frome various heterogeneous databases, data is collected together. Source from where data is gathered can be internal or external or any other source. As data is gathered from various applications, data can be of different formats. The following are the different kinds of data sets that can be used for testing. Internal database, database of field sensor, database of soil moisture content, database of local weather, database of wind influence, database of weather forecasting, database of marketing, database of stakeholder, and other databases.

- **Data preparation phase**

Once the data discovery phase is completed, data is gathered from multiple resources, data preprocessing is performed. Process of preprocessing includes methods like handling the noisy data, data cleaning (removing unnecessary data). The main focus of this phase is that the data which is to be provided for processing must be efficient and effective. Actual data which is only needed is obtained from this phase. Processed Data can then be converted into different file formats like CSV, ARFF files, Table format, or any other format. The data preparation phase covers all actions to build the final dataset from the data collected in the discovery phase. Tasks include table, record, and attribute selection as well as conversion and cleaning of data for forming tools.

- **Data analytics phase**

This phase consists of operations that are needed to get a valid solution. There are different techniques for it:

Support Vector Machine (SVM):

SVM is a machine learning algorithm that is used to forecast crop yield. It is based on the concept of separating the concept from each other. One plane contains the same features and another contains different planes. The classifier uses the concept aircraft, lines, and hyperplanes to separate these features. Line is used for one-dimensional data, a plane for two-dimensional data, and a hyperplane for three-dimensional data. It is good to use SVM if the elements are tightly bound; if they are not then this concept does not work properly. Input variable space is split by the hyperplane. Variable space is split by the hyperplane which is the line. It separates the points by their classes as class 0 and class 1 in the input variable space. Margin in SVM is the distance between the hyperplane and its closest data points. The largest margin indicates the best hyperplane which separates the two classes. These points are relevant for constructing the classifier which is called support vectors. The support vector supports the hyperplane and the optimization algorithms are used to find the coefficient value which maximizes the margin.

- **Visualization phase**

In this phase, the output of the analytical model can be displayed in the form of Queries, Reports, Charts, and Diagrams. The process involves 7 steps like Acquire, Parse, Filter, Mine, Represent, Refine, Interact.

1. Acquire: In this step, from heterogeneous sources, the entire datasets are obtained.
2. Parse: The raw datasets are unstructured which is difficult for analysis and visualization. We need to parse the data to make it structured and ordered. Structuring and organization of data may be done based on tags, indices, and names.
3. Filter: In this step, non required data is removed.
4. Mine: By using some statistical and mathematical formula, analysis of data is performed on filtered data.in this process, filtered data is converted into variables that can be used to display through data visualization.
5. Represent: After mining, the format for data display is decided. Data must be displayed clearly and concisely.
6. Refine: in this phase, different graphics tools and technologies are used to represent the data in a more engaged and enriched way. Technologies used can be HTML5, CSS,SVG, etc.
7. Interact: In this phase, a user must be able to interact with the data values by selecting varying data ranges, data within specific time intervals. To achieve this, different graphics and programming tools and technologies are widely used for making the reports and dashboards interactive.

REFERENCES

1. Dr. K. Sumathi, Prof. KundhvaiSantharam, Prof. N. Selvalakshmi, "Data Analytics platform for intelligent agriculture", Proceedings of the Second International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC 2018) IEEE Xplore Part Number:CFP18OZV-ART; ISBN:978-1-5386-1442-6
2. Jintao Huang, Lichen Zhang, "The big data processing platform for intelligent agriculture", Cite as AIP Conference Proceedings 1864, 020033 (2017); Published Online: 03 August 2017
3. M. JoharanBeevi and C. Vijayakumaran, "A Survey on Bigdata Frameworks for Agro Environment", Middle-East Journal of Scientific Research 24 (2): 314-319, 2016, ISSN 1990-9233.
4. Rajesh Kumar G., et al., "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM).

5. Maria Botsiou, VassiliosDagdilelisa, "Aspects of incorporation of ICT in the Greek agricultural enterprises: The case of a prefecture", 6th International Conference on Information and Communication Technology in Agriculture, Food and Environment, Procedia Technology 8 (2013)
6. https://www.upgrad.com/blog/what-is-big-data-types-characteristics-benefits-and-examples/
7. http://www.fao.org/statistics/en/