Big data analytics in Agriculture

Debdeep Bose
Letterkenny Institute of Technology

Big data analytics in Agriculture

Debdeep Bose. Author,MSc. Big Data Analytics Artificial Intelligence (Group A), Letterkenny Institute of Technology, Donegal, Ireland

***Abstract*—In the sector of agriculture which is very vast and requires a lot of planning, decision making, security and various other intricate factors influencing it. Though, the field of agriculture is less impacted by the recent technological advancements. However, agriculturalists are rapidly moving towards working with modern tools and technologies. One such up to the minute technology is Big Data analytics. Big data has been introduced to almost every other sector even agriculture is not outdistanced from it [1]. Agriculturists, Agribusinesses, institutions and researchers have been dependent on various techniques to collect related data. Henceforth, the collected data is further modified or turned into quality from quantity type. The sole focus is to extract acumens from it which can be utilized by the farmers or the end users and can be implemented to gain and achieve assured outcomes. Such as Apt crop forecasting, precision farming, smart agriculture, achieving high quality seeds, climate predictions and much more [2]. However, In-order to attain these niches a lot of big data analytic techniques have to be understood such as Predictive analytics, Machine learning, Classification and clustering, Recommendation system, Time series analytics, Regression analytics and Data mining. These are just a few that have been addressed here. Furthermore, a reviewed assimilation of various big data analytic techniques and its implementation in the field of agriculture has been obtained. Nevertheless, every technology has its drawbacks. Hence, the challenges of big data analytics in agriculture has been discussed with further augmenting to its future scope of work in the area of agriculture.**

***Index Terms*—Agriculture, Analytics, Big Data,**

I. Introduction

Agriculture is considered to be one of the major sectors of any nation and can be termed as the backbone of its economy. It is also one of the oldest vocations known to mankind. Predominantly, the judgment-making of a farmer is based on experience-based human intelligence and knowledge. Due to which it suffers from a huge number of unpredicted natural disasters such as climate change, lack of or untimely monsoon, droughts, floods. Also, factors such as institutions, governing bodies failing to support agriculture in terms of better farming schemes, loan facilities etc. Nevertheless, the time has come for technology to take over the transition and to resolve this situation. As we know, with recent trends and rise in "Big data analytics which is often the complex process of examining large and varied data sets, or big data, to uncover information – such as hidden patterns, unknown correlations, market trends and customer preferences which helps organizations make informed business decisions" [3] and has influenced innovation and advancement in virtually every field of the globe, neither is agriculture left behind. Based on scientific knowledge agriculturalists are moving away rapidly from traditional methods to more conventional tools and techniques for decision making and modern analytical modeling which further would help in increasing yield and revenue. Farmers do recognize that they all have massive amounts of data at their disposal. Hence, farmers and agricultural companies that suit their needs deal with huge amount of structured and unstructured data. In this paper we would see how Big data analytic techniques combined with farm data and data from several other sources such as satellite-drone imaging, corporate agronomical data, public data (data from government/governing bodies), Emerging data (Consumer based data), field-level sensors, weather stations and various historic data obtained from a variety of growers and growing conditions can provide information and insight to farmers and allow them to make appropriate decisions to increase their yields and reduce economic losses due to unpredictable changes. [2]

Under the supervision of Dr. Shagufta Henna, Letterkenny Institute of Technology, Letterkenny, CO. Donegal

A. Source for Agricultural data

Modern day farming produces huge amount of data by itself from sensors for example soil related, crop related, intercultural management, crop-pattern and harvesting related data. However, there are many official databases that are maintained and governed by institutions where they have long-term census, weather patterns and data of various kinds that can be analyzed and co-related. A few examples are Agricultural meta data element set (AgMES), [3] farmOS.org and agribusiness data etc. The data infrastructure should be defined in such a way that it can bear many independent modules with several API's and services. The infrastructure should be compatible enough to adapt to new modules without any specific changes to the core commands. The mentioned tools and databases are used in agricultural sector for data collection which further can benefit from its data infrastructure by identifying trends, patterns or any kind of disorder by visualizing them.

B. Concept of Big data in Agriculture

Big data has an enormous potential in the field of agriculture. The detailed understanding of Big data can help us overcome the significant challenges for any kind of traditional farming techniques that can lead to an increase in crop quality and production. Therefore, this technology in agriculture can be used to collect and analyze huge data that can be obtained from numerous stages in agriculture The major influence of implementing big data in agriculture are benchmarking, analytics, model prediction, visualization, marketing and management. It is also dependent on two major factors. [4]

1) Push factor: The primary push factor in agriculture are adoption of smart devices in all the stages of agriculture. The infrastructure is such that they all are interconnected to each other which keeps the overall system data ready to provide any kind of insights. This basically enhances decision making more positively.

2) Pull factor: This factor is basically addressing problems from a business perspective. It can be gained by investing on state-of-the-art technology related to big data. It can definitely reduce and optimize cost and efficiency of farming by achieving smart agriculture. It focuses on safety, security, productivity and quality by using big data.

C. Interpreting and Visualizing data

Enabling visualization of agricultural data obtained from various sources is an important phase because it makes sense of all the enormous complex, structured, and unstructured data. An accurate understanding from the huge amount of data can be interpreted in the form of overviews, summary, verifiable models etc. However, most of the interpretation and visualization is organized in an Ad-hoc manner. The models are basically visualized in the form of tables, graphs, spreadsheets etc. [5]

II. Big Data Analytic techniques in Agriculture

This study uses systematic literature review to answer the research questions. The literature review has been divided primarily into two phases. The first phase basically focuses on systematic survey of well-known analytic techniques, technologies, and its implementation in agriculture. It has usage of important keywords like Data-driven agriculture, smart agriculture, spatial analytics, recommendation systems and precision agriculture. Whereas, the second phase emphasizes more on snowball approaches of its challenges and future scope of research for a better and detailed study.

A. Predictive analytics

It is a technique used in big data analytics which can predict the future outcome based on gathered or historic data. It is one of the very few technologies that have the potential to remold any business positively. It majorly consists of three important techniques also called as the “Big 3” techniques: which are regression analysis, decision trees, and neural networks. [6] However, there are various other techniques that are used here such as esemble models and random forests. Predictive analytics emphasizes on building models that result in fit statistics. It is a type of statistics that is further used to solve complex business problems In terms of agriculture it is for used for crop prediction, yield prediction and consumers buying habits. A few predictive analytics techniques that have been discussed here are

1) Decision trees: This technique uses classification algorithms to influence the forecast risks or rewards with respect to various actions committed that further determines the possible outcomes that can be visualized by the humans in the form of flowcharts, graphs and various other pictorial formats. A decision tree has a root node which is basically the starting point, a leaf node and branches which generally ask questions with its possible answers.

2) Simple Statistical Modelling: In predictive analytics statistical modelling can be explained and used from a simple mathematical model to an intensive deep learning model. However, multiple linear regression model is the most used method in predictive analytics. Multiple linear regression is accustomed to build models for predicting future trends, forecast changes dependencies of one variable on another etc.

B. Recommendation System

The information system that offers output based on behavioral data and functional patterns. Recommender system generally gives advice as an output based on its approach and categories. It can be considered as an intelligent application system to help users assist with their decision making. In agricultural sector recommendation engines use big data algorithms such as collaborative filtering, Content based, Hybrid and Apriori with Association rule to analyze data and recommend crops based on weather and soil conditions, purchase of Agri-items. Recommendation system can be majorly classified into three based on its approaches. [7]

1) Content based: The content-based system totally depends on collations and similarities in order to make recommendation. Here the amplitude of similarity is based on the previous interests or heuristics. Especially, content-based recommendation system is determined on recommending items with related textual context. A few techniques such as Bayesian classifiers such as clustering, neural network which can be used to calculate the “utility prediction” with the data.

2) Collaborative Filtering (CF): It is a type of recommendation system which collects data from user feedbacks and converts them as ratings in any given area. It is also termed as a social filtering system. The general notion of it is if a set of users that agreed for a set of products will likely agree again to it in the future. Collaborative filtering exploits commonalities in rating behavior among various users. It can be either model based or memory-based approach. A few algorithms that are used in CF are K-nearest neighbor (Clustering) using matrix factorization and neural nets.

3) Hybrid: It is basically a combination of Content based and collaborative filtering. These evens the disadvantages of both the approaches and provides a better output for recommendation system. Some of the techniques are feature combination, augmentation, mixed, cascade, switching.

C. Data Mining

The data mining technique plays a very important role in agriculture. There are many data mining techniques that can be used in the area of farming. Especially, pattern mining which can be used to discover patterns in large datasets. An example, soil characterization is analyzed and studied by clustering various kinds of soils types by K-means and GPS based technology that can be used in precision agriculture and farm management. A few of the techniques have been discussed below. [8]

1) Association: This one of the most used data mining technique also known as the relation technique. In this mining technique a pattern is discovered based on the connection between items of the same activity or transactions. This technique is used in market basket analysis to research trends of purchase habits of a user.

2) Classification: Classification is a typical mining technique pertaining to machine learning. It used mathematical context in order to classify a set of data into predefined classes. A few techniques are statistics, linear regression, decision trees.

3) Clustering: Clustering is also defined as a data mining technique derived from machine learning. Here a cluster of objects is formed which have similar characteristics.Unlike classification where classes are predefined in clustering objects are defined and put under classes.

4) Prediction: As the name implies prediction data mining technique detects the relationship between dependent and independent variables in a given domain. It is generally used for predictions with the help of data such as text-mining, sentiment analysis etc.

5) Sequential Patterns: Sequential pattern or pattern mining is a data mining technique that explores similar kind of pattern, trends or transactions in any given data set. "The task of sequential pattern mining is a data mining task specialized for analyzing sequential data, to discover sequential patterns." [9]

D. Spike and Slab regression analytic technique

It is a kind of Bayesian technique used in big data for probability distribution of co-efficient in linear regression models.Here spike and slab are referred to a type of regression coefficients. The analytic technique offers unique advantages in dimensional issues.In the field of agriculture, it is very popular for tool and model selection. [8]

E. Time series analytic technique using Big data

In a time series analytic model, formerly observed values are used to forecast the output.It uses the concept of time series which is a series of data points arranged in time. However, here time is considered as an independent variable with an ambition to forecast the future. It is used in forecasting crops, vegetation price movement and price fluctuation in market. [10]

III. Implementing analytic techniques in Agriculture

The above-mentioned techniques have several applications in the field of agriculture and can be implied to enhance productivity and reduce manual inputs. A few of them have been discussed as follows.

A. Intelligent crop recommendation system

This is an intelligent system which is a combination of machine learning and big data analytics and is used for decision making, recommending crops and making predictions. This system takes into consideration all the parameters such as rainfall, soil conditions, temperature and location. The recommendation system is generally split into two sub-systems which are interrelated to each other. Firstly, the sub system consists of crop predictor and the second system is the rainfall predictor [11].

1) Crop predictor: This sub system is the primary function of the recommendation system which helps agriculturists recommend crops. Hence the initial part is acquiring of training data set because the accuracy of a machine learning algorithm is based on the selection of parameters and the correctness in its training data. The dataset generated has as schema such as soil type, aquifer thickness, soil PH, thickness of top-soil, precipitation, temperature and location.

The next part of the crop predicting system is the prepossessing of data. This is further divided into two steps. Basically, there could be missing values in the original training dataset which has to be removed and replaced. The reason for removing and replacing the values is because the presence of these can cause deterioration to the values and can affect the machine learning algorithms performance. Following that, the second step is to generate class labels for the dataset before applying it to the algorithm. The major reason behind labelling the data is because the recommender system is being trained using supervised learning. Since the system uses Multi-labeled classification. Therefore, the model has more than one class which is being assigned to a single instance for which the best machine learning algorithms would be K nearest neighbors (K-NN), Random forest, Decision trees and Artificial neural network [11].

Hence, after applying the appropriate machine learning algorithms. The trained models are used to recommend crops to the farmers; this can be done by saving the weights of the model. Henceforth, recommendation can be availed by just giving in raw inputs given initially which were soil type, aquifer thickness, soil PH, top-soil thickness, precipitation, temperature and and location.

2) Rainfall Prediction system: This sub system as the name suggests predicts the occurrence of rainfall. As we know each crop has its individual rainfall requirement and if the requirement is not met then the yield may suffer. However, if the rainfall is surplus it may also cause adverse consequences. Hence, rainfall can be considered one of the critical parameters in the field of agricultural analytics. However, prediction of rainfall accurately during the sowing and harvesting season can be one of the major challenges. Therefore, this sub system predicts rainfall for each month across the year. The dataset is obtained from the governments meteorological official site which consists of month wise rainfall data. In millimeters (mm) [11].

Similarly, as the previous sub system data must be preprocessed here the missing values are removed and are replaced with negative values in order to avoid deterioration. The machine learning algorithm that is implemented here is linear regression which is again a supervised learning approach where the parameters are considered as X (Location) and Y (precipitation values) which further can be used to predict the quantitative value of rainfall occurring in the region.

In conclusion, this is a successful implementation of an intelligent recommendation system that can be used by the agriculturists and farmers for making an informed decision.

B. Precision Agriculture using Map-Reduce

Map-reduce is defined as the processing technique and a program model for distributed computing [Cite]. It splits data into value and key. Firstly, map-reduce maps the data with the relative key value and the distributes or reduces the data with distributing it to several other nodes. Every node that are mapped with the key nodes can store data. This is one the best techniques to use in terms of getting fast and reliable 3D data view which makes decision making and reporting quite effortless. This technique however is used in the field of agriculture allowing the rational handling of inputs and variable rates which in return grants an enhanced understanding of temporal and spatial variability in soil and plant parameters. Such a technique is used for deep analysis of data which in return allows the agriculturists, farmers and few computer scientists in field of agriculture make precise decisions or forecasting. Hence termed as Precision agriculture [12].

The initial stage of the processing technique is to gather datasets. The datasets are obtained from meteorological departments and government official websites. The datasets that are collected is huge in volume and can exceed petabytes for a single research. The parameters in the dataset include water level, soil analysis, FCI storage and capacity and market consumption on daily basis.

Once the data has been obtained and preprocessed. Big data for distributed computing with map-reduce algorithm is to be implemented which compiles data from multiple sources together and generates a report. The attained report generally consists of details about current markets status or weather patterns etc. Further, generated report is visualized by means of an integrated business analytical application such as power BI which gives a 3D visual representation as output with the various parameters combined with charts and graphs which helps decision making an easy task for the farmers.

The major reason behind using visualization is because graphs and charts are in general easy to analyze because the human mind can perceive images and visual representation much faster and simple.

C. Crop prediction using various Machine Learning approaches

Agricultural data which is obtained in heterogeneous and homogenous form is being generated continuously. A few typical datasets are obtained from Historical data, streamed data, social and web-based data, agricultural sensor data and official institution websites. Hence, with abundance in agricultural big data and combining it with machine learning algorithms facilitates acquiring trends and prediction insights from data. Therefore, below discussed are a few Machine learning algorithms that are implemented.

1) Grey wolf optimization (GWO) technique: This is a Machine learning optimization technique that uses feature selection for creating a subset of classification. However, like other techniques the achieved data is preprocessed in order to increase the quality of data that is used for training the algorithm. The GWO technique selects the optimal feature subset and optimization is used as a relative weight to the features based on its training error and relative information. This approach is the best when it comes to find the semantic relations instead of geometrical cohesion [13].

The dataset which is obtained from official government institutions and applied for processing via map-reduce which reduces the execution time during feature selection and classification. The output is visualized with the help of a visualization tool which allows making decision easier.

Therefore, grey wolf optimization technique is used to carry out efficient relationship between features which in return reduces the training error in the model. Hence, increasing the precision and accuracy of prediction in the crop prediction model that is used by agriculturalists for their business.

2) K-means clustering: K-means is an unsupervised machine learning algorithm. In this algorithm each cluster is defined by the mean value of objects which are iterative, nondeterministic and numerical in nature. The data objects are found to be similar in each individual cluster compared to the other clusters. As we know, K means is a segregation-based cluster analysis technique the K value is selected as cluster center. Following that the data value is calculated and between each distance vectors and cluster center in order to assign it to the nearest cluster. The final goal is the prediction of K-centroids for individual datapoints. In the field of agriculture K-means clustering is used to predict crops by the help of graphical analysis of the obtained output. It has more enhanced graphs as compared to other approaches which helps decision making easier [14].

3) Apriori algorithm: The apriori algorithm is applied for mining frequent occurring items in the datasets. It generates a K-itemset followed by pruning and then it is clustered under the same data object. Frequent items sets are defined as the items that have the item with least support [14].

Once the data is obtained and preprocessed the attributes that are chosen as parameters are place, area, yield, soil type and rainfall. Following which the model is trained and the obtained output is visualized with the help of any visual analytics tool such as tableau or power BI for decision making and crop yield prediction.

Here in the field of agriculture it can be used in the prediction of crop yields by mining frequent item set.

4) Naïve Bayes: Naïve Bayes algorithm can be used for evaluating and understanding for a classification technique. Previous knowledge and functional learning algorithms can be conveyed with Bayes classification technique. It can calculate explicit probabilities combining with the observed data. Below is the defined formula:

$$P(X/D) = (P(D/X)P(X))/P(D)P(X) \tag{1}$$

In the above equation, D is the training data, X is the hypothesis and P is defined as the probability.Once the data is obtained and preprocessed the attributes that are chosen as parameters are place, area, yield, soil type and rainfall. Following which the model is trained and the obtained output is visualized with the help of any visual analytics tool such as tableau or power BI for decision making and crop selection [14].

In the field of agriculture, it is used to calculate the probability of growing a particular crop in an acre of land.

D. Smart Farming

Smart farming can be defined as the multidimensional combination of several different technological implementations [cite]. As per the domain of cognitive science "a process is considered smart if it has the following six characteristics which are adapting, inferring, sensing, anticipating, learning and self-organizing" [cite]. In order to achieve smart farming and to develop solutions; integration of technologies is required [15].

1) Internet of Things(IoT): Technologies such as Internet of things (IoT) in which it uses sensors placed in fields and farms that can gather data. Wireless sensors can also be used to analyze various agricultural parameters such as soil properties, temperature data etc.

2) Cloud Computing: Similarly, the collected data needs to be accumulated at a common platform that should be easily accessible. This can be achieved by introducing cloud computing which provides sharing of resources at a reasonable cost. It can be majorly used to store agricultural data and can be implemented alongside IoT.

3) Mobile Computing: Mobile computing which is used essentially in every domain including agriculture. It can be used to send daily or seasonal messages to agriculturists/farmers regarding weather, market or product details. Finally, these above-mentioned technologies are integrated with analytic techniques in order to find out new design, evaluation, concept and development.

The smart farming model which consists of the integrated technologies such as IoT, Cloud and Mobile technology uses the concept of Map-reduce in terms of data analytics. This is easy to handle data which uses multiple nodes. The process is divided in two parts which is map and reduce. Map is used to filter and sort. Whereas, reduce function can perform summary operations. Hence, this specific analytic technique is used for predictive analysis. It can also be implemented to perform Data mining because of the huge amount of generated data by the IoT sensors. The flow is data obtained from the sensors is stored in the cloud platform. Following which preprocessing of data is done then it focuses towards categorization and then attribute selection, algorithm implementation and later evaluating with pattern prediction. Hence the output that is obtained can be visualized using Business analytics tool and can be used for decision making such as fertilizer requirement, crop sequence, weather patterns etc [15].

E. Crop analysis using Data mining techniques

This implementation is aimed at analyzing greenhouse crops with the help of data mining technique to extract patterns. There are influencing factors such as crop growth, prediction of soil moisture which contains greenhouse variables such as soil, temperature, illumination level and humidity. The analysis process in terms of data mining technique requires establishing an objective variable and designing an analysis system [16].

In greenhouse farming data is generated through devices and sensors of Internet of Things, Internet of everything with wireless networks and embedded systems. The factors and variable that are selected as attributes in the datasets are soil moisture, level of carbon di oxide, relative humidity and luminosity. Once the data is obtained and preprocessed with the help of knowledge discovery in databases (KDD) following which pattern mining is implemented with the help of decision tress and classification rules. This approach provides a predictive model with the collected data. However, data mining is implemented in order to study new patterns with large amount of available data. After the final interpretation of the output it is evaluated and visualized with the help of tools in Graphical user interface. This user interface contains the visual classes of the different icons and algorithms.

Therefore, with the help of user interface and selection of specific greenhouse attributes farmers will be able to predict yield patters, crop patterns and further make important decisions based on them [16].

F. Spark based Agricultural information system

Geo-spatial data plays a very important role in many areas and applications. Similarly, the use of spatial data can be beneficial in the field of agriculture. For instance, technologies such as remote sensing, GPS and devices that use high resolution geo-spatial data for creating insights with advanced analytics algorithms. Hence, here we discuss about a spark-based system that can collect, learn, train, validate and visualize distributed Geo-spatial data.

The agricultural big data is collected from various sources few being institutions, open web data, government official websites and agricultural universities. The data that is obtained is raw and noisy in nature which must be cleaned and labelled in order to achieve consistency and train the model. The data which is preprocessed is now stored in Cassandra, which is a highly available and a high performance database. Once the data is stored successfully it is integrated with Apache spark framework which is a highly scalable cluster computing framework and has integrated API's in various languages such as Java, Scala and python also libraries such as GraphX, MLib, SparkSQL etc. After the data is imported from Cassandra it is further combined with a spark extension named Geospark. This is an open source third party extension that is used for analyzing spatial data. The algorithm that is used here in Multiple linear regression and is done by performing numerous transactions on the data. Finally, when the required model is obtained then its output is visualized with help of web based json interactive maps.

This method of data analytics can be used for current weather trend, crop yield prediction and perform insights on Agricultural market data [17].

IV. Challenges of Big data in agriculture

There are two major challenges in Big data analytics for agriculture (i) Technical Challenges (ii) Organizational challenges[2]. Technical challenges are related to the technological issues such as device installation, information technology (IT) platforms, technical expertise and troubleshooting, security and basically smooth functioning of the Infrastructure. On the other hand, Organizational challenges are more on to the business side like Financial gains, investment, safety, maintaining a team of expertise in the field, personnel recruitment and an overall management. The primary attention are towards revenue and investment with proper maintenance of quality, yield, safety and its security. As we can observe that implementing state-of-the-art technologies requires a huge investment and planning to deal which can be accepted by agriculturists of developed nations. However, it is very difficult for developing nations to cope up with in terms of infrastructure and expenditure [18].

A few other barriers are, from a sociopolitical viewpoint there is huge dependence of farmers on the monopoly of agri-food business and corporations due to the advancement in their practices. Concentration of big data technology in the hands of these groups limits the potential of growth and freedom of it because it is used in ways that can benefit only the individual. Nevertheless, there are a lot of questions raised on the issues such as practice of data collection, controlling of data and policies for monetizing it. Agriculturists are skeptical about sharing their farming data because they have a fear of loosing it to their competitors [21].

Therefore, the related points discussed above indicated are a few open challenges of big data analytics in the field of agriculture.

V. Future scope of work in Agriculture

This section lists a few potential scope of applying big data analytics in the area of agriculture [19].

- As farmers sometimes are unable to sell their yield due to excessive supply. Better demand and yield prediction models should be created.
- Supply chain have the access to high quality products and processes. Hence, integrating international supply chain with the field of agriculture would give the sector an immense boost overall.

Stages of the data chain	**State of the art**	**Key issues**
Data capture	Sensors, Open data, data captured by UAVs Biometric sensing, Genotype information Reciprocal data	Availability, quality, formats
Data storage	Cloud-based platform, Hadoop Distributed File System (HDFS), hybrid storage systems, cloud-based data warehouse	Quick and safe access to data, costs
Data transfer	Wireless, cloud-based platform, Linked Open Data Machine learning algorithms	Safety, agreements on responsibilities and liabilities
Data transformation	normalize, visualize, anonymize	Heterogeneity of data sources, automation of data cleansing and preparation
Data analytics	Yield models, Planting instructions, Benchmarking, Decision ontologies, Cognitive computing	Semantic heterogeneity, real-time analytics, scalability
Data marketing	Data visualization	Ownership, privacy, new business models

Fig. 1. Challenges of Big data analytics in agriculture [5]

- Product trace ability and shipment tracking with retail monitoring of consumers can enhance agri-product business.
- Application of better optimization techniques in order to achieve better food processing techniques with the help of meta heuristics and genetic algorithms. Better optimization techniques are important for increasing food quality and reduce losses.
- Large scale remote sensing and crop mapping data with the collaboration of governments and institutions will be essential for monitoring the impact of environmental sustainability. This is in respect to achieve and measure the productivity.
- Better advanced and full-fledged scientific simulations and models could be for climate phenomenon will be able to provide policy makers, farmers and agriculturists with more help in decision making.
- High precision in farming can be obtained with the help of better screening methods which offer high quantitative analytics between plants and its environment.
- Using self-operating robots in the field of agriculture would certainly revolutionize this domain and its total productivity. It can be used to identify weed and remove them or identify pest and abolish them.
- Advancing precisely in genetic engineering would make it achievable for crop and farm animal mutation. This will be more acceptable to the end users because it is more or less like crop breeding. This would also definitely contribute in the area of epigenetics.

The aforementioned points would also produce large amount of big data which can be used by future agriculturalists and researchers to find solutions for the challenges in the field.

VI. Comparison table of Big data analytic techniques

TABLE I
BIG DATA ANALYTIC TECHNIQUES [20]

Area in Agriculture	Sources of Big data	Analytic Techniques implemented
Weather and climate data	Geo-spatial data weather stations Historical information via websites remote sensing data	Statistical analysis Machine learning(K-means, random forest) Map-reduce GIS analysis
Animals	Ground sensors heat data feed intake milk production data optical sensors	Neural networks scalable vector machines Decision trees
Crops and Vegetation	Historical datasets satellite data Ground senors are devices Government produced data	K-means clustering Fourier transform scalable vector machines wavelet filtering
Land	Historical data Geo-spatial(Radar/satellite) data Airplane data remote sensing aperture data	Random forest K-means Image processing ndvi vegetation indices
Weeds	Historical data Drone image data Airplane image Field sensors data Web based digital libraries	Logistic regression Neural networks Image processing
Soil	Moisture data Ground sensors salinity data from sensors Historical database optical data	Neural networks K-means clustering algorithm
Biodiversity	Historical data Government websites Geo-spatial data SER databases	Naive Bayes statistical modelling
Food Security	Historical data Survey data remote sensors crop growth datasets Geo-spatial data	geo-spatial analysis simulation neural networks statistical modelling image processing
Farmers	Historical data Government websites Social media weather stations optical sensors	Big data benchmarking web services mobile apps
Remote Sensing	Drone data satellite data weather datasets Web based data Geospatial data	Cloud computing with Map reduce analytics decision support system Geo-spatial analysis computer vision Artificial intelligence
Insurance and Finance	Government data website data Human generated data Private institutions (Banks) optical sensors	Predictive analytics Statistical Modelling cloud technologies

VII. CONCLUSION

In the review paper, the notion of introducing big data to agriculture has been discussed with its driving factors and what visible constructive changes have been observed, after its initiation to the field as compared to the traditional methods. Given the variety of data required for performing big data analytics we also explore the various sources, methods and techniques for obtaining data. An agronomy-based framework for performing agricultural analytics has also been explained. Also, with the recent development in the field of technology, platforms and tools. Integrating them with big data analytics and implementing them in the field of agriculture has also been introduced, explained and reviewed in detail. Finally, a few challenges have been discussed, with possible solutions to overcome it.

Following that, there is a representation of comparison table of various agricultural areas with its data sources and the techniques that can be applied to it. Nevertheless, the increase in big data and big data analytics with its adoption to open standards have immense potential to boost more research and development towards smart agriculture. Having said that we do not want to challenge the existing and traditional agricultural niches implemented by the farmers and agriculturists. But instead motivate and enhance the ongoing methods for producing higher quality products, generate more revenue for the business and plan sustainably without affecting the natural resources [21].

References and Footnotes

References

[1] D. Team, "Data Science in Agriculture - Advancing Together and Benefiting Farmers," DataFlair, 27-Sep-2019. [Online]. Available: urlhttps://dataflair.training/blogs/data-science-in-agriculture/. [Accessed: 02-Dec-2019]

[2] "Agriculture is the new bee for Big Data analytics!," Irish Tech News. [Online]. Available: https://irishtechnews.ie/agriculture-is-the-new-bee-for-big-data-analytics/. [Accessed: 02-Dec-2019].

[3] M. N. Islam Sarker, M. Wu, B. Chanthamith, S. Yusufzada, D. Li, and J. Zhang, "Big Data Driven Smart Agriculture: Pathway for Sustainable Development," in 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2019, pp. 60–65, doi: 10.1109/ICAIBD.2019.8836982.

[4] S. Rajeswari, K. Suthendran, and K. Rajakumar, "A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics," in 2017 International Conference on Intelligent Computing and Control (I2C2), 2017, pp. 1–5, doi: 10.1109/I2C2.2017.8321902.

[5] K. Charvat et al., "Advanced Visualisation of Big Data for Agriculture as Part of Databio Development," in IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 415–418, doi: 10.1109/IGARSS.2018.8517556.

[6] R. V. McCarthy, M. M. McCarthy, W. Ceccucci, and L. Halawi, "Introduction to Predictive Analytics," in Applying Predictive Analytics, Cham: Springer International Publishing, 2019, pp. 1–25.

[7] A. Agarwal*1, "Educational Data Sets And Techniques Of Recommender Systems: A Survey," Oct. 2017, doi: 10.5281/ZENODO.1036288.

[8] "Data Mining Techniques," ZenTut. [Online]. Available: https://www.zentut.com/data-mining/data-mining-techniques/. [Accessed: 06-Jan-2020].

[9] R. H. Ishwaran, U. B. Kogalur, and J. S. Rao, "spikeslab: Prediction and Variable Selection Using Spike and Slab," 2010, doi: 10.32614/rj-2010-018.

[10] M. Peixeiro, "The Complete Guide to Time Series Analysis and Forecasting," Medium, 07-Aug-2019. [Online]. Available: https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775. [Accessed: 06-Jan-2020].

[11] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms," in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697349.

[12] M. Kumar and M. Nagar, "Big data analytics in agriculture and distribution channel," in 2017 International Conference on Computing Methodologies and Communication (ICCMC), 2017, pp. 384–387, doi: 10.1109/ICCMC.2017.8282714.

[13] S. Sharma, G. Rathee, and H. Saini, "Big Data Analytics for Crop Prediction Mode Using Optimization Technique," in 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2018, pp. 760–764, doi: 10.1109/PDGC.2018.8746001.

[14] S. V. Bhosale, R. A. Thombare, P. G. Dhemey, and A. N. Chaudhari, "Crop Yield Prediction Using Data Analytics and Hybrid Approach," in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1–5, doi: 10.1109/ICCUBEA.2018.8697806.

[15] S. Rajeswari, K. Suthendran, and K. Rajakumar, "A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics," in 2017 International Conference on Intelligent Computing and Control (I2C2), 2017, pp. 1–5, doi: 10.1109/I2C2.2017.8321902.

[16] K. L. Ponce-Guevara et al., "GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial)," in 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), 2017, pp. 1–6, doi: 10.1109/ETCM.2017.8247519.

[17] P. Shah, D. Hiremath, and S. Chaudhary, "Towards development of spark based agricultural information system including geo-spatial data," in 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 3476–3481, doi: 10.1109/BigData.2017.8258336.

[18] S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, "Big Data in Smart Farming – A review," Agricultural Systems, vol. 153, pp. 69–80, May 2017, doi: 10.1016/j.agsy.2017.01.023.

[19] Y. Cheng, Q. Zhang, and Z. Ye, "Research on the Application of Agricultural Big Data Processing with Hadoop and Spark," in 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2019, pp. 274–278, doi: 10.1109/ICAICA.2019.8873519.

[20] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," Computers and Electronics in Agriculture, vol. 143, pp. 23–37, Dec. 2017, doi: 10.1016/j.compag.2017.09.037.

[21] Sanjay Bhatikar Ph.D., PMP, "Big Data and Agriculture," 13:11:22 UTC.