

Machine Learning in Agriculture

Subjects: **Computer Science, Information Systems | Automation & Control Systems**

Contributor: Ania Cravero

Big Data and ML have appeared as high-performance informatics technologies for creating new opportunities to unravel, quantify and understand data-intensive processes in the environment of farm operations

big data

machine learning

architecture

adaptation

agriculture

systematic literature review

1. Big Data and ML in Agriculture

[Figure 1](#) and [Figure 2](#) show a graph representing an XY scatter diagram. The size of each bubble is proportional to the number of papers that are in the pair of categories that correspond to the coordinates in which the bubble is located.

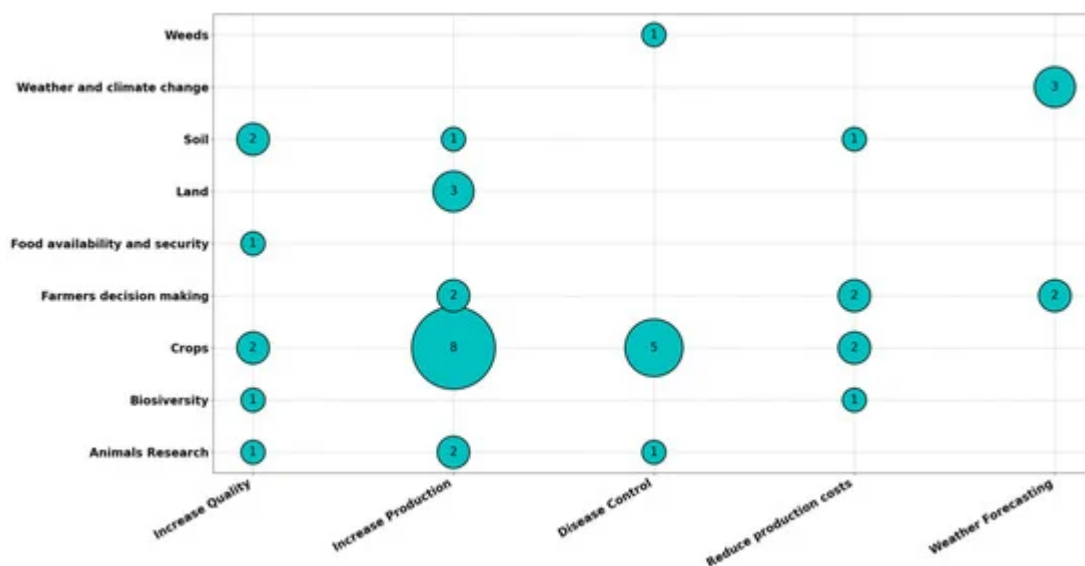


Figure 1. Agriculture types versus the types of problems encountered.

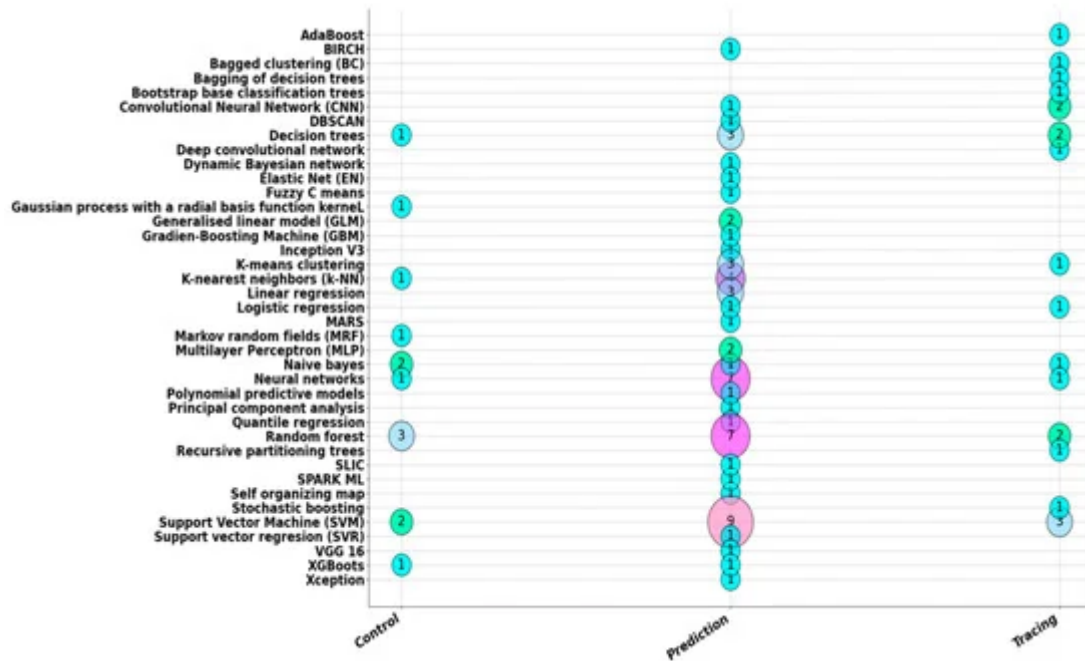


Figure 2. ML techniques used versus the application domains.

Figure 1 presents the number of papers by agriculture heading versus the type of problem it solves.

We found few papers describing the use of Big Data and ML for weeds, food availability and security and biodiversity problems. Greater advances are observed in the Crops area to solve production problems and diseases. ML and Big Data are being used to analyze a large volume of historical data from sensors, satellite images, photos, climate studies and data from expert farmers, using cloud tools. Quality and cost reduction of production seem less important. These research topics are likely to increase in the coming years.

On the other hand, Figure 2 shows the number of papers by ML technique versus application domains in agriculture. It is observed that the most used techniques are Random Forest, Support Vector Machine and Neural Networks.

2. Technologies Used in Big Data and ML Architectures for Agriculture

Figure 3 presents a cloud of concepts about the technologies used in Big Data and ML architectures in agriculture. Technologies for data acquisition are used in data collection systems for the environment, soil, plant data, animals; through sensors, satellite images and videos. On the other hand, data from experts, data from other systems and the cloud are collected.

The adaptations of ML in Big Data for agriculture have been implemented to solve problems of processing a volume of data and aspects of variety. The cloud largely solves the problem thanks to the different applications and resources available (i.e., treatment of satellite images and videos).

The speed of data ingestion is still a challenge considering data sources such as sensors, drones, cameras and other systems. Regarding data processing, speed depends on factors such as data volume and selected technologies. The cloud has technologies, such as Data Lake, to process data and store it in different areas according to the level of processing or detail. The use of these Data Lake is a challenge for developers since there are not enough trained human resources for their implementation.

Data Lake is defined as “a large, heterogeneous, raw data storage system, powered by multiple data sources and allowing users to explore, extract and analyze data” [1]. The positive aspects of its users consider the improvement in data quality since it is provided by a set of metadata; data control and traceability according to data governance policies; the integration of data of all types and formats; the organization of data logically and physically. Data Lake integrates with Big Data to store and control the data in its different states, raw data, semi-processed and processed (value data), thus improving the speed of processing in the analysis for ML.

Another technological challenge is data visualization since platforms and applications for agriculture still have traditional designs. It is necessary to incorporate the use of Business Intelligence tools and intelligent platforms, such as those of Microsoft, IBM (Watson) and those of the cloud, to display data dynamically, allowing decision-makers to manipulate data and reports according to the needs of the moment in real-time [2]. In the context of Big Data, traditional data visualization methods cannot meet the needs of data visualization and analysis. How to handle this data and extract its potential value has become increasingly important for competition and organizational development [3].

[Figure 4](#) presents a framework with the technological challenges of the use of ML and Big Data in agriculture.

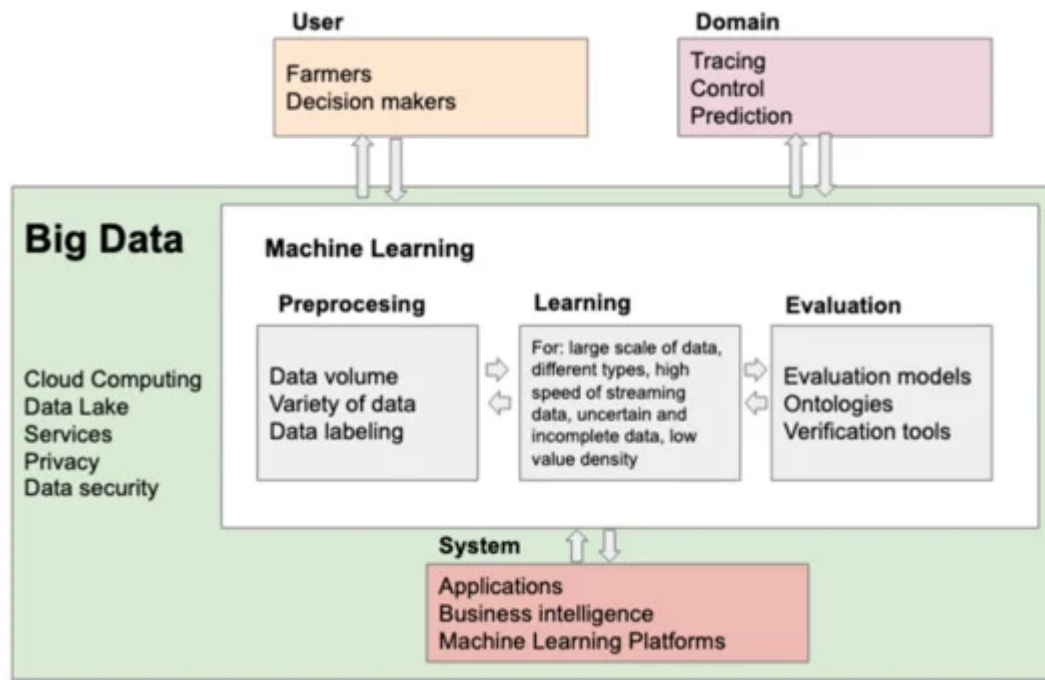


Figure 4. A framework of Machine Learning and Big Data challenges in agriculture.

From the analyzed papers we find several challenges that must be solved to use ML in Big Data in the area of agriculture. The main ones are the following:

(a) Redundancy in the data: It is also known as data duplication and leads to inconsistent data that can be detrimental to the ML based system. There are techniques to identify duplicates in a given data set [4] but these traditional methods are not effective in the case of Big Data. In the cases analyzed, we have found redundancy problems when using various data sources, such as sensors, satellite images and expert judgment.

(b) Data discretization: is the process of translating quantitative data into qualitative data that results in a non-overlapping division of continuous domain. In the case of agriculture, the challenge remains to use adequate data visualization systems for this. In most cases graphs of the data are shown and the profile of the users as farmers is not considered.

(c) Data labeling: In this regard, data labeling tools have not been used to improve understanding of the data obtained from data sources. This challenge is important so that Big Data can be used with new technologies such as Data Lake.

(d) Privacy and data security: quality, which has always been a key issue in agricultural management information systems but is more challenging with volume of real-time data. On the other hand, data privacy and security have not been mentioned in the analyzed papers.

(e) Interoperability due to the growing number of platforms and services: There is still little concern about using standards and formats that allow the exchange of information, in addition to interoperability.

(f) Bandwidth for data transmission: It is still a problem due to the increasing volume of data [\[5\]](#).

(g) Fault tolerance: It is still a challenge because data processing is done in batches, faults are possible and the process or information must be recovered [\[5\]](#).

(h) Using ML tools in the cloud: Some papers describe the challenge of using ML tools in the cloud. Some examples are: Microsoft Azure Machine Learning, now part of Cortana Intelligence Suite [\[6\]](#); Google Cloud ML Platform [\[7\]](#); Amazon ML [\[8\]](#); and IBM Watson Analytics [\[9\]](#). Because these services are backed by powerful cloud providers, they offer not only scalability but also integration with other cloud platform services. With Big Data, this results in high network traffic and can even become unfeasible due to time or bandwidth requirements. Because these ML services are proprietary, information about their underlying technologies is very limited [\[10\]](#).

References

1. Khine, P.P.; Wang, Z.S. Data lake: A new ideology in big data era. ITM Web Conf. 2018, 17, 03025.
2. Hirve, S.; Reddy, C.H.P. A Survey on Visualization Techniques Used for Big Data Analytics. Adv. Intell. Syst. Comput. 2019, 447–459.
3. Jun, S. Business Intelligence Visualization Technology and Its Application in Enterprise Management. In Proceedings of the 2020 2nd International Conference on Big Data Engineering and Technology; Association for Computing Machinery (ACM), Singapore, 3–5 January 2020; pp. 45–48.
4. Chen, Q.; Zobel, J.; Verspoor, K. Evaluation of a machine learning duplicate detection method for bioinformatics data-bases. In Proceedings of the ACM Ninth International Workshop on Data and Text Mining in Biomedical Informatics, Melbourne, Australia, 23 October 2015; pp. 4–12.
5. Sassi, I.; Ouaftouh, S.; Anter, S. Adaptation of Classical Machine Learning Algorithms to Big Data Context: Problems and Challenges: Case Study: Hidden Markov Models Under Spark. In Proceedings of the 2019 1st International Conference on Smart Systems and Data Science (ICSSD), Rabat, Morocco, 3–4 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–7.
6. Barga, R.; Fontama, V.; Tok, W.H. Cortana Analytics. Predict. Anal. Microsoft Azure Mach. Learn. 2015, 279–283.
7. Google. Google Cloud Machine Learning. 2016. Available online: (accessed on 15 November 2016).
8. A.W.S. Amazon. Machine Learning. 2016. Available online: (accessed on 7 June 2016).
9. IBM. IBM Watson Ecosystem Program. 2014. Available online: (accessed on 8 January 2014).

10. Padhi, B.K.; Nayak, S.; Biswal, B. Machine Learning for Big Data Processing: A Literature Review. *Int. J. Innov. Res. Technol.* 2018, 5, 359–368.

Retrieved from <https://encyclopedia.pub/entry/history/show/18743>