



Exploring The Synergy Of Artificial Intelligence In Data Science And Analytics – A Review

¹Neetu

¹Assistant Professor

¹Department of Computer Science

¹Government College, Gharaunda, Haryana, India

Abstract: Data science is the study of extracting, collecting, gathering, representing, and safeguarding data for use in technical problems or for business goals. The ultimate goal of artificial intelligence is to automate all human activities currently requiring human intelligence. The main challenge is creating a mechanism that functions exactly like the human brain. Data science deals with complicated problems analytically. The handling of vast amount of data is the main challenge in data science. In this paper, we study about the various artificial intelligence methods for data science and data analytics that replicates the working of human brain in data processing and creates patterns, reduces it if possible and produces the accurate results. This paper discusses the applications, hardware, software and AI techniques.

Index Terms - Data Science, Deep Learning, Machine Learning, Artificial Intelligence, Data Analytics.

I. INTRODUCTION

Data science uses automated machine learning techniques to examine and mine massive volumes of data. Data science is a field that mostly developed out of necessity for practical applications rather than as a research area. It has evolved over time from being utilized in the relatively narrow sector of estimations to being a thorough closeness in every aspect of research and industry. The following applications of data science include detecting credit card fraud, movie recommendations, spotting fake news, smart inquiry Chabot, classifying breast cancer, and predicting rainfall and self-driving cars. [1][2] As the amount of unstructured data we produce increases, information science simply becomes irrelevant. By allowing teams to share code, results, and reports, a data science stage reduces excess and promotes advancement. By untangling the board and using open source tools, structures, and framework, it eliminates bottlenecks in the flow of work[3].

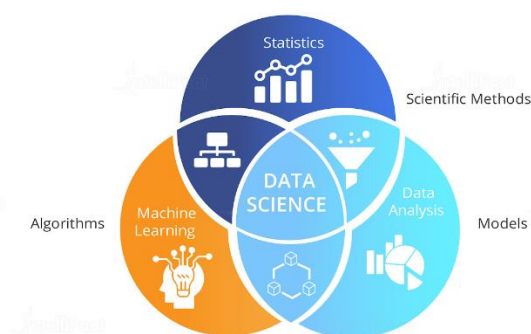


Fig 1: Representing Data Science and Analytics with Machine Learning

II. OPEN ISSUES FOR DATA SCIENCE

Data science involves analysing vast amounts of data and includes data extraction. The internet of things (IoT), cloud computing, and quantum computing are three broad categories into which the research problems pertaining to enormous data examination are grouped.

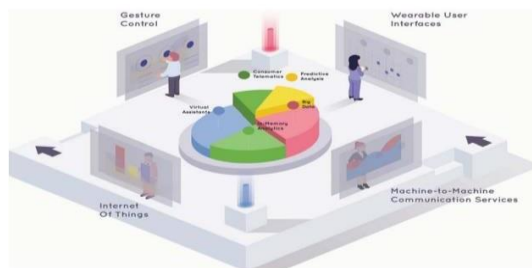


Fig 2 : Emerging Technologies in Data Science

A. *IoT for Data Science*

Machines are eventually joining the show to operate countless autonomous devices using web-based techniques and create the Internet of Things (IoT). In this way, mechanical assemblies are evolving into online users, much like how people interact with web applications. Researchers are considering the Internet of Things for its most exciting opportunities and challenges. For the creation of data, frameworks, and correspondence in the future, it has a critical financial and societal influence.

B. *Cloud Computing for Data Science*

Computer platforms wrapped in virtualization software enable platforms to behave like a real PC while having the adaptability of specific features like processors, plate space, memory, and working framework. Massive Data and cloud computing improvements are made with the goal of establishing flexible and on-demand access to resources and information. Cloud Computing organizes massive amounts of data by allowing on-demand access to manageable intellectual resources via computerised processes.

C. *Quantum Computing for Data Science*

A genuine quantum computer would have been able to tackle the current enormous data problems, which are especially difficult for conventional computers, if one existed today. It may soon be able to develop quantum computers without the usual specialised difficulties. A method for dealing with the quantum mechanics to process the data is provided by quantum figuring. Also, it frequently gets caught up and captured by the wonders of other areas. It is because qubits behave in a quantum manner.

III. DATA SCIENCE AND DATA ANALYTICS

The primary goal of data science when it was first introduced to the industrial level was to discover more precise and trustworthy solutions than those attained via business analysis. An analyst needs to possess the sought-after skill sets of data processing, predictive modelling, and visualisation to operate in the data science industry. There is a wide range of uses for data science. It is used in a variety of industries, including finance, transportation, and healthcare. Many businesses use data science to aid in their production, improve on better informed decisions, and produce unique items that are tailored to the needs of customers. Data science is evolving at an exponential space due to the availability of tools, technologies and resources. It addresses many real word problems with optimistic solution.

A. *Machine Learning*

From the early 1950s, there has been machine learning technology. Machine learning is a result of a data-driven method that was developed in 1990. There was a change in emphasis towards information retrieval and natural language search from 1995 to 2005. The neural network, which was first used in 1957 for the first neural network computers, made a resurgence in 2005. Machine learning is one of those technologies that has both successful and unsuccessful applications, but there is a chance that in the next 2 to 5 years, this technology will become widely used. There should be increase in some of the aspects affecting the machine learning business, such as infrastructure and technical competence, to maintain its growth.

B. *Deep Learning*

Alexey Grigoryevich Ivakhnenko and Valentin Grigor' evich Lapa introduced deep learning in 1965 [4]. They used many models with complex equations and polynomial functions that was statistically examined. A technique for identifying and mapping related or similar data was created in 1995. For recurrent neural networks, long short term memory is established in 1997. With the advent of processors with quick computing speeds in the late 1990s, the efficiency of GPUs for image processing increased, eventually multiplying computation speed by a factor of about 1000. Currently the processing of Big Data is dependent on Deep Learning. Currently AI and Deep Learning growing and more advanced ideas are coming up.

C. *Artificial Intelligence*

The first workshop on artificial intelligence was held in 1956 at Dartmouth College [5]. The lack of computer hardware resources for computations was an issue back then. After the government insisted, business and the government contributed billions of dollars to the development of AI during the 1980s. Once funds and interest were provided to advance the subject of artificial intelligence, it saw a boom from 2000 to 2010. Machine learning became an effective solution to a variety of issues in business and society after the creation of robust computer technology.

IV. ARTIFICIAL INTELLIGENCE IN DATA ANALYTICS

In many regions of the world, data analytics and artificial intelligence are popular research topics. Developing nations are spending millions to become more prominent in the world. A tiny number of them are funding unclassified artificial intelligence research.

A. *Machine Learning methods*

Support vector machines are supervised learning models that examine the data used in regression and classification studies. They work well for categorising variables in a binary fashion and do not need that there be a linear relationship between them. The idea behind the Naive Bayes classifier is that it is a probabilistic classifier based on the Bayes theorem with assumptions about the independence of the features. The aggregate, conditional probabilities of various qualities are calculated.

B. *Deep Learning methods*

The back propagation algorithm is used to compute the gradients required for a network's weight calculation. Both stochastic and batch learning methods are available. During deep learning training, the units are dropped at random using the dropout strategy. It mostly serves to lower the amount of parameters. According to the skip-gram technique, two vocabulary terms are only comparable if they are used in the same context.

V. RESOURCES USED IN DATA ANALYTICS

A. *Hardware Resources*

NVIDIA or Intel hardware GPU components are used by numerous researchers, data scientists, and research companies for training, testing, and accelerating their algorithms for deep learning. GPU effectively and quickly completes parallel computing processes for data, pictures, movies, and graphics. The new Jetson Xavier computer from Nvidia is equipped with a variety of processing components despite its small size. By creating effective core hardware components, the Intel AI Chip-Neural network processor offers flexible support for all deep learning primitives. It scales numerous processing nodes with minimal electricity and maximises computation efficiency.

B. *Software Resources*

The software and equipment used for deep learning and artificial intelligence are described.

Pylearn2: a modular, extensible, and built for machine learning algorithms library that offers support for GPU and CNN.

Torch : is an open source machine learning framework and AI tool. Lua is the used scripting language. One of the most well-liked tools because to its adaptability and computational effectiveness. Reimplementation logic is not supported, which is the disadvantage. The newest open source framework is torchnet, and it is effective and reusable.

Theano: It is a scripting language that uses the Python library to effectively evaluate mathematical expressions. Multidimensional arrays are involved. Theano Because it uses mathematical formulas, it does data-intensive tasks more quickly than the CPU. Using to train deep learning networks for huge data sets [6].

Caffe : It is an open source framework that supports a number of libraries for AI deployment architecture, including MATLAB, CUDA, C++, and Python.

Cuda-convnet: Rapid Neural Network is compatible with Python, C++, and CUDA. Convolutional neural networks are used in it.

VI. APPLICATION OF DATA SCIENCE

Data science is a field that primarily developed out of necessity, as opposed to being an area that was being explored. It has evolved over time from being used in the relatively narrow realm of measurements and research to being a ubiquitous presence in every sphere of science and business.

1. Significant amounts of collected and separated data can be used to spot patterns in the data, which can then be utilised to compile perceptual models. This is the premise of the field of machine learning, which is based on numerous estimations that are claimed to "learn" and data acknowledgment figuring. All things considered, machine learning techniques are employed to create perceptual models across several disciplines.
2. Information science is used to identify fraud using data taken from analyst logs. Instances found in client actions can be used to cut off extortion cases and malicious insiders. Information mining and machine learning algorithms are heavily used by banks and other financial institutions to detect fraud.
3. Computer vision, the study of teaching computers to "see," uses picture information and learning calculations to acquire and decode images and make decisions in a similar way. It makes use of data from picture and video analysis. Applications for autonomous driving, self-sufficient transportation, and human-PC collaboration use this.
4. Modern NLP methods utilize tremendous measures of literary information from corpora of records to factually show etymological information, and utilize these models to accomplish undertakings like machine translation, parsing, characteristics dialect age and notion analysis. [7]

VII. CONCLUSION

In the current scenario, artificial intelligence is becoming more prevalent and we are making progress in that direction. It's also critical to recognise the technology's versatility in minimising hazards while maintaining safety. Analytics prediction is currently being done using machine learning techniques. When it comes to providing reliable results, deep learning is thought to offer more advantages than traditional machine learning methods. The intricacy and difficulties involved with this field grow as we advance in it.

REFERENCES :

- [1] C. Dichev and D. Dicheva, "Towards Data Science Literacy," *Procedia Comput. Sci.*, vol. 108, no. June, pp. 2151–2160, 2017, doi: 10.1016/j.procs.2017.05.240.
- [2] S. Shreshtha, A. Singh, S. Sahdev, M. Singha, and S. Rajput, "A Deep Dissertation of Data Science: Related Issues and its Applications," *Proc. - 2019 Amity Int. Conf. Artif. Intell. AICAI 2019*, pp. 939–942, 2019, doi: 10.1109/AICAI.2019.8701415.
- [3] C. Alonso-Fernández, A. Calvo-Morata, M. Freire, I. MartínezOrtiz, and B. Fernández-Manjón, "Applications of data science to game learning analytics data: A systematic literature review," *Comput. Educ.*, vol. 141, no. June, p. 103612, 2019, doi: 10.1016/j.compedu.2019.103612.
- [4] A brief history of deep learning- Data versity.
- [5] J Vincent, "Nvidia launches AI computer to give autonomous robots better brains" *the verge*, 2018.
- [6] Welcome – Theano 1.0.0 documentation. <http://deeplearning.net/software/theano/>.
- [7] Manjunath Jogin, Mohana, Madhulika, Meghana R K "Feature extraction using Convolution Neural Networks (CNN) and Deep Learning" 2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,(RTEICT) 2018, India.