e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

A COMPRESSIVE REVIEW MODELS FOR BIG DATA ANALYTICS RELIES ON ARTIFICIAL INTELLIGENCE

N. Manimozhi[*1], R. Suganya[*2], Dr. P. Senthil Pandian[*3], G. Suguna[*4], R. Sakthi Devi[*5], S. Ramya[*6]

[*1,2,4,5,6]Assistant Professor, Department Of Computer Science And Applications, Farouk Educational Trust, Perambai, Villupuram District, Tamilnadu – 605110, India.

[*3]HOD-Department Of Computer Science And Applications, Farouk Educational Trust, Perambai, Villupuram District, Tamilnadu – 605110, India.

ABSTRACT

Big Data Analytics poses a grand challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse, complex, and of a massive scale. Big data analytics must also be team effort cutting across academic institutions, government and society and industry and by researchers from multiple disciplines including computer science and engineering, health, data science and social and policy areas.

Keywords: Big Data, Ai Models, Neural Network, Supervised Machine Learning.

I. INTRODUCTION

AI and big data can work together to achieve more. First, data is fed into the AI engine, making the AI smarter. Next, less human intervention is needed for the AI to run properly. And finally, the less AI needs people to run it, the closer society comes to realizing the full potential of this ongoing AI/big data cycle. That evolution will require the involvement of human beings who are trained in data analytics and AI algorithm programming.

The ultimate goals of AI are as follows:

- Reasoning
- Automated learning and scheduling
- Machine learning
- Natural language processing (the ability to understand human speech as it is spoken)
- Computer vision (the ability to extract accurate information from an image or series of images)
- Robotics
- General intelligence

For these AI fields to mature, their AI algorithms will require massive amounts of data. Natural language processing, for example, will not be possible without millions of samplings of human speech, recorded and broken down into a format that AI engines can more easily process.

Big data will continue to grow larger as AI becomes a more viable option for automating more tasks — and AI will become a larger field as more data is available for learning and analysis.

II. BIG DATA ANALYTIC TECHNIQUES AND REVIEW

Athmaja, Hanumanthappa & Kavitha (2017) presented a systematic literature-based review of the big data analytics approaches according to the machine learning mechanisms. However, no categorization is provided for reviewing related studies in the present paper. Moreover, the non-functional features of the studies have not been investigated. The authors do not provide any systematic procedure for gathering the related studies. Ghani et al. (2019) have reviewed the existing big social media analytics approaches in five classes: artificial neural networks, fuzzy systems, swarm intelligence, evolutionary computation, and deep learning. The authors assessed the reviewed techniques based on their quality metrics. However, there is no systematic procedure to select articles related to this field.

Another work provided by Sivarajah et al. (2017) for the big data analysis techniques. The authors categorized these techniques into three main groups, including descriptive, predictive, and prescriptive analytics. However, there are some gaps in analyzing the qualitative parameters, and the study selection process.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

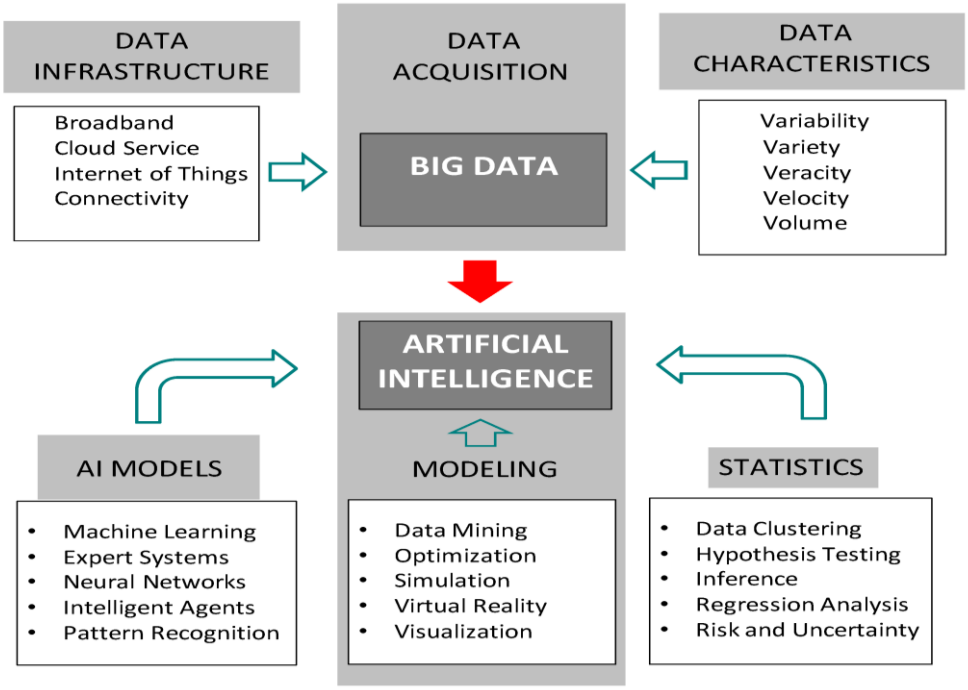

Figure 1: AI Models.

III. TYPES OF AI MODELS

There are various types of AI models, defined by the means used to create them. Three approaches used in data science are supervised learning, unsupervised learning, and semi-supervised learning models.

Supervised Machine Learning Models

AI models can be built using supervised machine learning. These models are trained by people, often ones with specific subject matter expertise, typically referred to as subject matter experts or SMEs. SMEs review new data points and label them. While training data, they might mark it as "responsive" or "non-responsive". They might tag it as relating to any number of issues, such as "Contains offensive language" or "Privileged". Models learn from the training the SMEs provide in real-time, and use that learning to find more similar content.

AI models built with supervised machine learning often are used to perform predictive analyses. They look to the past, assessing decisions made by the SMEs about the documents they have reviewed. Using artificial neural networks loosely designed after the human brain, the models use those assessments to attempt to predict the future, forecasting the decisions SMEs might make about the remaining documents. A common use of supervised machine learning in eDiscovery is TAR, or technology assisted review.

Unsupervised Machine Learning Models

AI models also can be developed with the help of unsupervised machine learning, an approach that incorporates more automation. These models are trained by software, sometimes using a process that mimics the training provided by people. They categorize your input data or identify patterns or trends without the need for initial human training.

AI models built using unsupervised machine learning typically are turned to for descriptive analyses. They might be used to summarize content. They could be used to classify content, with the classified content then displayed using a visual tool such as a cluster wheel. They could be used to extract rules about content.

Semi-Supervised Machine Learning Models

This subset of machine learning is often described as a middle ground between supervised and unsupervised machine learning and combines aspects of the two other approaches. SMEs label a small amount of data to start training a model. That partially-trained model is pointed to a larger body of data that the model then labels, a process referred to as "pseudo-labelling".

The results of the two approaches are combined and used to create a model that might be used for descriptive or predictive purposes.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

Expert system:

An expert systems development project balances potential savings from the proposed system against the cost. The team members develop a prototype system to test assumptions about how to encode the knowledge of experts. Next, they develop a full-scale system, focusing mainly on the addition of a very large number of rules. The complexity of the entire system grows with the number of rules, so the comprehensibility of the system may be threatened. Generally, the system is pruned to achieve simplicity and power. The system is tested by a range of experts within the organization against the performance criteria established earlier. Once tested, the system is integrated into the data flow and work patterns of the organization.

Example for Expert: System Organizational Intelligence: Case-Based Reasoning

Expert systems primarily capture the tacit knowledge of individual experts, but organizations also have collective knowledge and expertise that they have built up over the years. This organizational knowledge can be captured and stored using case-based reasoning. In case-based reasoning (CBR), descriptions of past experiences of human specialists, represented as cases, are stored in a database for later retrieval when the user encounters a new case with similar parameters. The system searches for stored cases with problem characteristics similar to the new one, finds the closest fit, and applies the solutions of the old case to the new case. Successful solutions are tagged to the new case and both are stored together with the other cases in the knowledge base. Unsuccessful solutions also are appended to the case database along with explanations as to why the solutions did not wor**k**

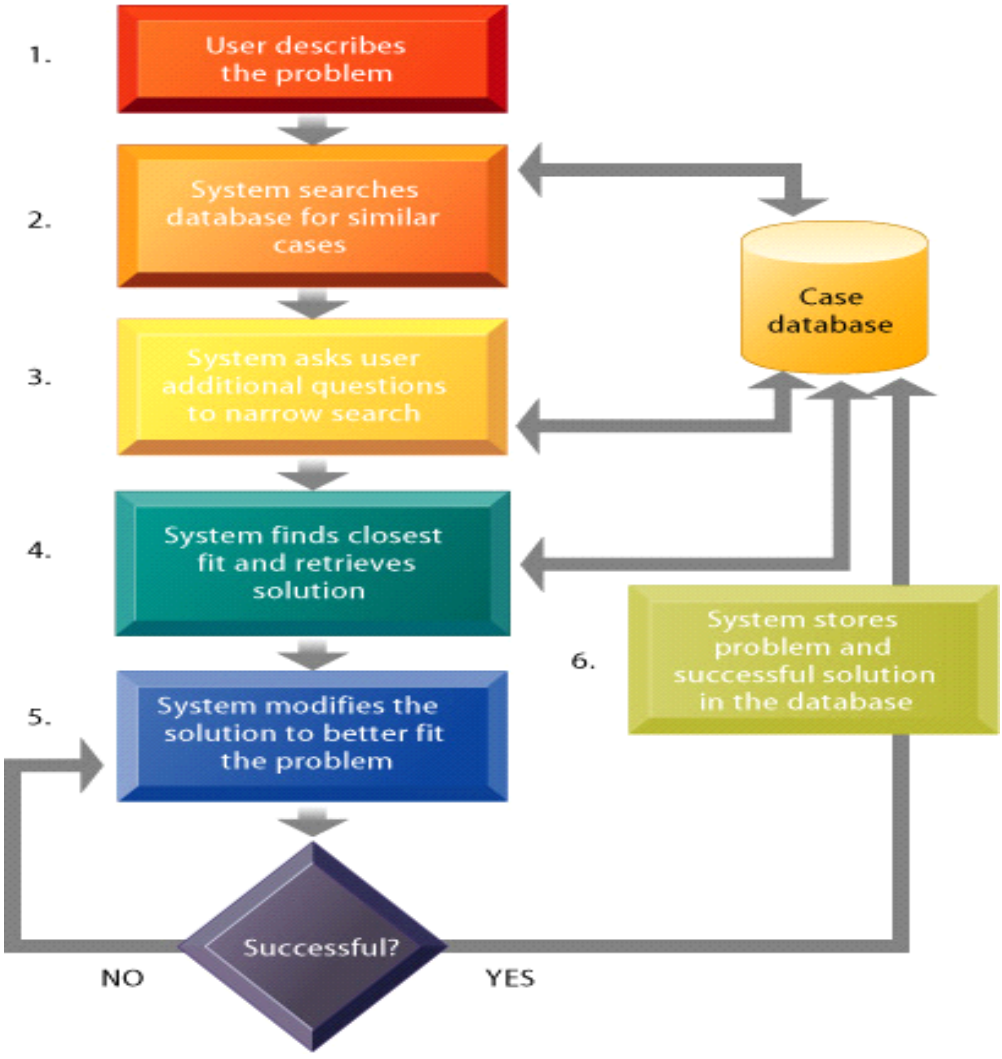

Figure 2: How case-based reasoning works.

Case-based reasoning represents knowledge as a database of past cases and their solutions. The system uses a six-step process to generate solutions to new problems encountered by the user.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

Neural networks:

Neural networks are used for modeling complex, poorly understood problems for which large amounts of data have been collected. They are especially useful for finding patterns and relationships in massive amounts of data that would be too complicated and difficult for a human being to analyze. Neural networks discover this knowledge by using hardware and software that emulate the processing patterns of the biological brain. Neural networks "learn" patterns from large quantities of data by sifting through data, searching for relationships, building models, and correcting over and over again the model's own mistakes.

A neural net has a large number of sensing and processing nodes that continuously interact with each other. Figure 3 represents one type of neural network comprising an input layer, an output layer, and a hidden processing layer. Humans "train" the network by feeding it a set of training data for which the inputs produce a known set of outputs or conclusions. This helps the computer learn the correct solution by example. As the computer is fed more data, each case is compared with the known outcome. If it differs, a correction is calculated and applied to the nodes in the hidden processing layer.

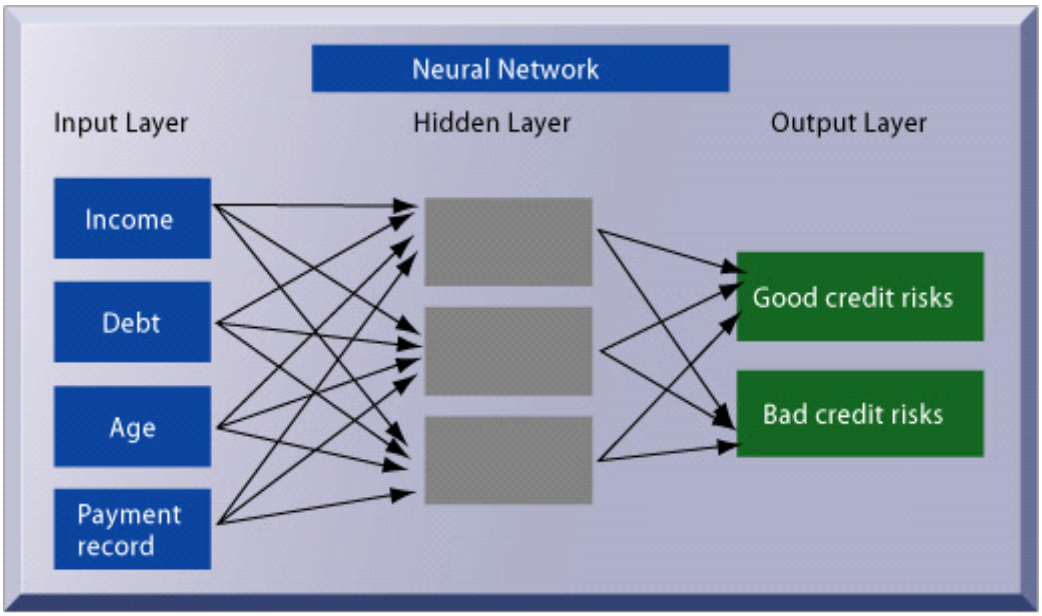

Figure 3: How a neural network works

Intelligent Agents:

Intelligent agent technology can help businesses navigate through large amounts of data to locate only information that is considered important and, in some cases, act on that information on behalf of the user. Intelligent agents are software programs that work in the background without direct human intervention to carry out specific, repetitive, and predictable tasks for an individual user, business process, or software application. The agent uses a limited built-in or learned knowledge base to accomplish tasks or make decisions on the user's behalf. Intelligent agents can be programmed to make decisions based on the user's personal preferences—for example, to delete junk e-mail, schedule appointments, or travel over interconnected networks to find the cheapest airfare to California. The agent can be likened to a personal digital assistant collaborating with the user in the same work environment. It can help the user by performing tasks on the user's behalf, training or teaching the user, hiding the complexity of difficult tasks, helping the user collaborate with other users, or monitoring events and procedures.

Pattern Recognition:

Pattern recognition is defined as the study of how machines can observe the environment, learn to distinguish various patterns of interest from their background, and make logical decisions about the categories of the patterns. During recognition, the given objects are assigned to a specific category. Because it is a constantly evolving and broad field, there exist several definitions of Pattern Recognition, as it is a constantly evolving and broad field.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

An early definition of pattern recognition defines it as "a classification of input data via extraction of important features from a lot of noisy data" (1978, Thomas Gonzalez). Another definition describes pattern recognition as "a scientific discipline whose aim is the classification of the objects into a lot of categories or classes. Pattern recognition is also an integral part of most machine intelligence systems built for decision making" (2003, Sergio Theodoridis). In general, pattern recognition can be described as an information reduction, information mapping, or information labeling process. In computer science, pattern recognition refers to the process of matching information already stored in a database with incoming data based on their attributes.

Pattern Recognition and Artificial Intelligence (AI)

Artificial Intelligence (AI) refers to the simulation of human intelligence, where machines are programmed to think like humans and mimic their actions. Most prominently, fields of artificial intelligence aim to enable machines to solve complex human recognition tasks, such as recognizing faces or objects. Accordingly, pattern recognition is a branch of Artificial Intelligence.

IV. GLOBAL ARTIFICIAL INTELLIGENCE IN BIG DATA ANALYTICS

1. The report also analyzes how different forms of AI may be best used for problem-solving. The report also evaluates the market for AI in IoT networks and systems. The report provides forecasting for unit growth and revenue for both analytics and IoT from 2022 to 2027. The global AI market is predicted to snowball in the next few years, reaching a $190.61 billion market value in 2025.
2. The wearable AI market size is predicted to reach $180 billion by 2025.
3. The forecasted AI annual growth rate between 2020 and 2027 is 33.2%.
4. The global AI chip market revenue is expected to reach $83.25 billion by 2027.
5. Between 2018 and 2025, the Asia-Pacific region will experience the highest compound annual growth rate.
6. By 2030, China will be the world leader in AI technology, with 26.1% of the global market share.
7. A lack of trained and experienced staff is an expected restriction in the AI market's growth.
8. In 2019, the machine learning application industry received $37 billion of funding in the U.S.
9. By 2030, AI will lead to an estimated $15.7 trillion, or 26% increase in global GDP.

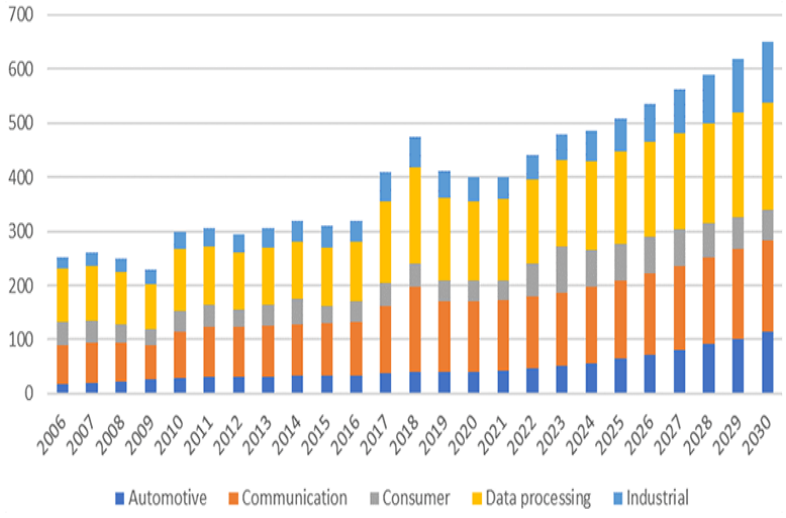

Figure 4: Global Artificial intelligence in Big Data analytics

V. CONCLUSION

In the Real-world examples of AI models are all around us. In fact, we encounter them in many aspects of our lives. In the healthcare industry, AI models help doctors and other medical staff diagnose pediatric diseases. They let companies identify opinion spam on e-commerce, social media, and similar sites. They are used to summarize communications, write articles, and detect credit card fraud. Google uses AI algorithms to match riders with carpool drivers; detect and filter out email spam; and assist with composing messages. Microsoft's AI Builder provides AI models used to process forms, extract insight from product reviews, and automate inventory taking. AWS is a managed service that provides developers and data scientists with the ability to

build, train, and deploy models for uses such as predictive maintenance, computer vision, and predicting consumer behavior. And Netflix, Amazon, and YouTube use AI systems as a means of service optimization by using big data machine learning to provide more relevant content recommendations.

VI. REFERENCES

[1] "Big Data Technologies: A Survey." Journal of King Saud University - Computer and Information Sciences 30, no. 4 (2018): 431–48. https://doi.org/10.1016/j.jksuci.2017.06.001. CITE.

[2] Mützel, Sophie. "Facing Big Data: Making Sociology Relevant, Facing Big Data: Making Sociology Relevant." Big Data & Society 2, no. 2 (2015): 2053951715599179. https://doi.org/10.1177/2053951715599179. CITE.

[3] Chen, Min, Shiwen Mao, and Yunhao Liu. "Big Data: A Survey." Mobile Networks and Applications 19, no. 2 (2014): 171–209. https://doi.org/10.1007/s11036-013-0489-0. CITE.

[4] Alkurd, Abualhaol & Yanikomeroglu (2020) Alkurd R, Abualhaol I, Yanikomeroglu H. Big-data-driven and AI-based framework to enable personalization in wireless networks. IEEE Communications Magazine. 2020;58(3):18–24.

[5] Bengio, Courville & Vincent (2012) Bengio Y, Courville AC, Vincent P. Unsupervised feature learning and deep learning: a review and new perspectives. 20121206.5538.