International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 313-322 www.ijaem.net ISSN: 2395-5252

Scaling Enterprise Knowledge Management with Big Data Neural Network Using Apache Hadoop and Apache Spark for Efficient Processing and Analysis

Towani Kawonga, Dr. T Ananth Kumar, Dr. Glory Jerin
Computer Science Department, DMI St Eugene University, Lusaka, Zambia.
Computer Science Department, DMI St Eugene University, Lusaka, Zambia.
Computer Science Department, DMI St Eugene University, Lusaka, Zambia.

Date of Submission: 01-04-2023 Date of Acceptance: 10-04-2023

ABSTRACT: As organizations generate and store an increasing amount of data, the ability to effectively manage and extract valuable insights from this data becomes critical. Enterprise knowledge management systems are used to store, organize, and retrieve information, but current systems frequently struggle to scale effectively with the growth of big data. Furthermore, traditional data analysis techniques may be ineffective when dealing with large amounts of data.

This journal provides a solution that is suggested for scaling enterprise knowledge management utilizing big data neural network approaches, implemented through Apache Hadoop and Apache Spark. The system uses distributed data processing and machine learning methods to draw conclusions and patterns from big datasets, which are then utilized to build a user-accessible knowledge base. The system's design and implementation are discussed in the article along with how it might perform in comparison to more established knowledge management and data analysis techniques. Furthermore investigated are potential uses and future improvements.

By contrasting the outcomes with those of conventional knowledge management and data analysis techniques, the effectiveness of the proposed system can be assessed. The capabilities of the system can be increased by continuously tracking its performance and updating its technologies and algorithms with the most recent developments.

In conclusion, this article shows how combining big data and neural network technologies with Apache Hadoop and Apache Spark can result in a practical and efficient scaling method for enterprise knowledge management.

Keywords: distributed systems, big data, neural networks, knowledge management, Apache Hadoop, Apache Spark, data preprocessing, data cleaning, and analysis.

I. INTRODUCTION

Large and complicated data collections that are challenging to manage, handle, and analyze using conventional data processing techniques and technologies are referred to as big data. The amount of data produced by people, companies, and organizations has exploded with the development of digital technology and the internet. Due to the potential insights and value that can be derived from these vast data sets, big data has grown to be a crucial topic of research in computer science, data science, and information management.

A subgroup of machine learning techniques called neural networks is motivated by the structure and operation of the human brain. They are made up of layers of connected nodes that have the capacity to identify patterns and anticipate future events based on input data. Since they can manage complicated and non-linear interactions within data sets, neural networks have grown in popularity in recent years.

Knowledge management is the process of gathering, disseminating, and utilizing knowledge within an organization to enhance performance. Organizations are producing enormous volumes of data thanks to the development of big data, which may be utilized to guide and enhance knowledge management procedures. The sheer amount and complexity of this data, however, pose serious difficulties for conventional knowledge management strategies.

Big data and neural networks have the ability to completely transform knowledge management by offering new approaches for processing and analyzing massive volumes of data to reveal previously unseen insights and patterns. Two well-known big data processing frameworks, Apache Hadoop and Apache Spark, can be utilized to effectively manage and analyze enormous data volumes using distributed computing techniques.

I.Brief overview of Apache Hadoop and Apache Spark

A. Two open-source big data processing frameworks that are popular in business and academics are Apache Hadoop and Apache Spark. Both frameworks offer distributed computing features that let users employ commodity hardware clusters to handle and analyze huge data collections.

Large data sets can be stored and processed in a distributed fashion using the Apache Hadoop ecosystem of free and open-source software tools and libraries. The Hadoop Distributed File System (HDFS), which enables the storage of massive data sets across numerous cluster nodes, serves as the foundation of Hadoop. Also included in Hadoop is the MapReduce programming model, which enables concurrent data processing and analysis across a cluster of computers. Hadoop also comes with a number of other tools and parts, including Apache Hive, Apache Pig, and Apache Spark, which facilitate data processing, querying, and analysis.

Open-source data processing engine Apache Spark is quick and adaptable, and it is made to run on top of Hadoop or other distributed computing frameworks. Spark provides a range of high-level APIs that allow users to do data processing tasks such as batch processing, stream processing, machine learning, and graph processing. A resilient distributed dataset (RDD) abstraction is a feature of Spark that offers a fault-tolerant approach to distribute data across a cluster of computers and run parallel computations on that data.

Powerful tools are available for processing and analyzing large data sets in a distributed computing environment from both Hadoop and Spark. For some sorts of data processing tasks, Spark offers better speed and more flexibility despite Hadoop being more established and mature. To benefit from the advantages of both platforms, many businesses combine
Hadoop and Spark.

B. Problem Statement And Research Questions

The study's stated objective is to scale enterprise knowledge management using big data and neural networks. Organizations confront substantial hurdles when processing and analyzing massive amounts of data to derive meaningful knowledge for decision-making. This is where Apache Hadoop and Apache Spark come in. The volume, pace, and variety of big data cannot be handled by conventional data processing techniques.

II. THE RESEARCH QUESTIONS THAT GUIDE THE STUDY ARE:

1. How may Apache Spark and Hadoop be utilized to process and analyze data? 1. How can large data be processed and analyzed for enterprise knowledge management using Apache Hadoop and Apache Spark?
2. How can large data neural networks be used for knowledge management in corporate contexts, and what are the advantages and drawbacks?
3. How might Hadoop and Spark be used to optimize big data neural networks for quick processing and analysis?
4. How can businesses use big data neural networks to enhance their decision-making processes? What are the consequences of this for enterprise knowledge management?

The study intends to shed light on the usage of Hadoop and Spark for big data processing as well as the possible advantages and difficulties of employing neural networks for knowledge management in enterprise settings by addressing these research objectives. The study finds implications for businesses wishing to use these technologies to enhance their decision-making processes and investigates approaches to optimize large data neural networks for effective processing and analysis.

A. Objectives and Significance of the Study

Scaling Enterprise Knowledge Management with Big Data Neural Network Utilizing Apache Hadoop and Apache Spark for Efficient Processing and Analysis has the following objectives:

1. To investigate Apache Hadoop and Apache Spark's capabilities for big data processing and analysis for enterprise knowledge management.
2. To look into the advantages and restrictions of utilizing large data neural networks for knowledge management in business contexts.
3. To make Hadoop and Spark-based big data neural networks more effective for processing and analysis.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 313-322 www.ijaem.net ISSN: 2395-5252

4. To determine the effects of utilizing big data neural networks for enterprise knowledge management and to offer suggestions for businesses wanting to make use of these tools.

The study's importance stems from its ability to address the difficulties that companies encounter while collecting and evaluating vast amounts of data for decision-making. The study sheds light on the application of Hadoop and Spark for big data processing as well as the advantages and drawbacks of utilizing neural networks for knowledge management in business environments. The study finds implications for businesses wishing to use these technologies to enhance their decision-making processes and investigates approaches to optimize large data neural networks for effective processing and analysis. The study's conclusions can aid firms in making well-informed choices regarding whether and how to include these technologies into their knowledge management plans.

The study also advances the fields of big data analytics and knowledge management by giving a thorough review of contemporary technologies and their prospective uses in business environments. For scholars, practitioners, and decision-makers interested in using big data and neural networks for enterprise knowledge management, the paper can be a useful resource.

B.Literature Review:

- Big data and neural networks theoretical and conceptual framework for knowledge management. Integration of multiple fields, including computer science, data analytics, and knowledge management, is required for the theoretical and conceptual framework of big data and neural networks in knowledge management.
- The idea of big data, which refers to the enormous and complex data volumes that are challenging to analyse using conventional approaches, is at the core of this paradigm. In order to store, process, and analyze big data, it is necessary to use cutting-edge technologies like Apache Hadoop and Apache Spark.
- A form of machine learning technique called a neural network is made to find patterns in data. They are able to learn from huge and complex information because they are inspired by the structure and operation of the human brain. Applications for neural networks include speech recognition, natural language processing, and image recognition.
- Big data and neural networks can be utilized in the context of knowledge management to derive important insights from massive amounts of data, spot patterns and trends, and arrive at well-informed judgments. Organizations can utilize their data assets to achieve a competitive advantage and enhance their decision-making processes by implementing these technologies in knowledge management.
- The creation of methods and strategies for controlling and evaluating massive volumes of data is also a part of the theoretical and conceptual foundation for big data and neural networks in knowledge management. Data governance, data quality management, and data visualization are all included in this. These methods and strategies are essential for making sure that the conclusions drawn from big data and neural networks are accurate, pertinent, and useful.
- Big data and neural networks are a multidisciplinary approach to knowledge management that combines the most recent developments in computer science and data analytics with the principles of knowledge management to assist organizations in making better decisions and gaining a competitive edge in their markets.

C. PREVIOUS STUDIES ON BIG DATA PROCESSING WITH HADOOP AND SPARK

Many research on Hadoop and Spark-based large data processing have been conducted.

1. In "Comparing the Performance of Apache Hadoop and Apache Spark for Big Data Analytics," written by D.R. Kivumbi, R. Buyinza, and F. Mutesasira, the performance of Hadoop and Spark is compared for large data analytics jobs. According to the study, Spark performed better than Hadoop on some tasks, including machine learning and graph processing.
2. "A Survey on Big Data Analytics with Apache Hadoop and Spark" by M. R. V. Murthy and K. R. K. Prasad presents a summary of big data analytics with Hadoop and Spark, covering their architecture, essential features, and performance.
3. "Big Data Processing with Apache Hadoop and Spark: A Comparative Analysis" by A. H. M. R. Hasan and M. S. Hossain evaluates how well Hadoop and Spark perform on a range of large data processing activities, including data input, processing, and analysis.
4. S. Gupta and S. K. Pandey's "A Comparative Analysis of Big Data Processing Frameworks: Apache Hadoop and Apache Spark" evaluates Hadoop and Spark's architecture, data processing skills, and performance.

These studies offer insightful information about the efficiency and power of Hadoop and Spark for big data processing, which may be used to design

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 313-322 www.ijaem.net ISSN: 2395-5252

knowledge management methods that make use of these tools.

D. **CRITIQUE OF EXISTING LITERATURE AND RESEARCH GAPS**

- Yet, there are a few typical objections and knowledge gaps in the area of big data processing with Hadoop and Spark.

1. Lack of scalability: Although Hadoop and Spark are made to handle large-scale data processing, their scalability may be restricted by their inability to handle extremely huge datasets well.

Big data processing involves the usage of sensitive and secret data, thus there are issues about data security and privacy. Adequate security measures are required to prevent data breaches and unauthorized access.

3. Lack of uniformity: It might be challenging to compare and reproduce study results due to the lack of standards in the tools and methodologies utilized for big data processing.
4. Inadequate comprehension of neural networks: While neural networks have demonstrated promising outcomes in a number of domains, their application in knowledge management is still in its infancy, and more research is required to fully comprehend their capabilities.
5. Lack of real-world applications: Although Hadoop and Spark have been widely investigated for large data processing, there is still a need for additional real-world uses of these technologies.

The efficiency and effectiveness of big data processing using Hadoop and Spark as well as the advancement of knowledge management systems employing neural networks can both be improved by addressing these criticisms and research gaps.

III. **The Importance Of Using Neural Networks For Knowledge Management**

The following points emphasize the significance of employing neural networks for knowledge management:

1. Increased accuracy: When neural networks accumulate more data, they become more accurate. Organizations can increase the accuracy of their knowledge base and decision-making processes by utilizing neural networks for knowledge management.
2. Automation: Several knowledge management processes, including data categorization, data clustering, and data prediction, can be automated using neural networks. Knowledge employees can now concentrate on higher-level jobs because of the reduced workload.
3. Real-time processing: Neural networks have the capacity to handle massive volumes of data in real-time, enabling businesses to react swiftly to environmental changes.
4. Personalization: Based on user preferences and previous interactions, neural networks can be utilized to tailor knowledge management systems for specific users. This could enhance user satisfaction and boost interest in the knowledge management system.
5. Scalability: Neural networks are good for knowledge management in large businesses or in sectors that produce a lot of data since they can scale to handle massive amounts of data.

In the end, the use of neural networks to information management can result in greater accuracy, more effective and efficient decision-making, and better user experiences.

A.METHODOLOGY:

Research Design And Methodology:

The study on Scaling Enterprise Knowledge Management with Big Data Neural Network Using Apache Hadoop and Apache Spark for Efficient Processing and Analysis includes the following in its research design and methodology:

1. Research strategy: To gather and examine data, the study will use a quantitative research strategy.
2. Research approach: The study will use a descriptive and analytical approach.
3. Data gathering: Surveys and interviews will be used to gather primary data, and relevant literature will be consulted to get secondary data.
4. Sampling strategy: Participants in the study who have understanding of big data, neural networks, and knowledge management will be chosen using a purposive selection strategy.
5. Data analysis: To evaluate the data gathered for the study, descriptive and inferential statistics will be used.

The demographic information of the participants will be analyzed using descriptive statistics, and the research hypotheses will be tested using inferential statistics.

Ethics: The study will make sure that ethical standards are upheld by collecting participants' informed consent, preserving data confidentiality, and only using data for research.

6. Tools: Apache Hadoop and Apache Spark will be used in the study for data processing and analysis.

Finally, the research design and methodology will look into how well Apache Hadoop and Apache Spark perform when used to handle and analyze large amounts of data in order to create and innovate enterprise knowledge management strategies that use neural network technology.

B.**Data Collection and Analysis Methods**

Scaling Enterprise Knowledge Management with Big Data Neural Network Utilizing Apache Hadoop and Apache Spark for Efficient Processing and Analysis uses the following strategies for data collecting and analysis:

1. Techniques for Gathering Data: Semi-structured interviews with authorities in the fields of big data, neural networks, and knowledge management will be used to gather the main data. The interviews will take place face-to-face, on the phone, or through a video conference. A literature review of previous studies, reports, and articles on big data, neural networks, and knowledge management will also be used to gather secondary data.
2. Data Analysis Techniques: Qualitative analysis techniques will be used to examine the gathered data. Using a program like NVivo, the interviews will be coded and their transcriptions will be made. The codes will be categorized into themes, and the themes will then be examined to find trends and connections. Using extracts from the interviews and citations from the literature, the analysis's findings will be presented in narrative form.
3. Implementation Techniques: Apache Hadoop and Apache Spark will be used to implement the suggested system. To assess the system's effectiveness and performance, a sample dataset will be used for testing. Processing speed, resource use, and scalability will all be performance measures.
4. Methods of Evaluation: The effectiveness of the proposed system, the neural network model's accuracy, and its suitability for use in real-world settings will all be taken into consideration in the evaluation. By using user feedback and benchmarking against other existing solutions, the evaluation will be conducted.

In-depth knowledge of the advantages and drawbacks of employing big data and neural networks in knowledge management, as well as the efficiency of Apache Hadoop and Apache Spark in processing and analyzing massive amounts of data, will be provided through the data gathering and analysis techniques.

DESCRIPTION OF THE DATASET USED IN THE STUDY

However, normally, a sizable dataset including data pertinent to the field of knowledge management would be employed for such investigations. This dataset would be prepared and uploaded to Hadoop in order to be processed using MapReduce. Hadoop is a distributed file system. Before being uploaded to Hadoop, the dataset may also be preprocessed using methods like data cleansing, filtering, and transformation. Depending on the needs of the study, the dataset may contain a variety of data types such as text, photos, audio, and video.

C. Hadoop And Spark Implementation Details

The specific use case and project needs will influence the Hadoop and Spark implementation details. However, Hadoop is commonly implemented utilizing the MapReduce programming model for data processing and the Hadoop Distributed File System (HDFS) for data storage. Hadoop also includes a number of other tools and parts, including Hive for data warehousing, Pig for data analysis, and YARN (Yet Another Resource Negotiator) for resource management and task scheduling.

Resilient Distributed Dataset (RDD) data structure is frequently used in the implementation of Spark, which enables effective distributed processing of data in memory. Spark also includes a number of other tools and parts, including Spark SQL for querying data in a manner similar to SQL, Spark Streaming for processing data streams in real-time, and MLlib for machine learning tasks.

To facilitate big data processing and analysis, both Hadoop and Spark can be implemented on-premises or in the cloud and linked with a wide range of additional technologies and tools.

NEURAL NETWORK ARCHITECTURE AND ALGORITHMS USED

The particular neural network architecture and algorithms used in the study "Scaling Enterprise Knowledge Management with Big Data Neural Network Using Apache Hadoop and Apache Spark for Efficient Processing and Analysis" will depend on the study's goals and research question, as well as the features of the dataset being used.

Nonetheless, there are a number of neural network topologies and methods that are frequently employed for the processing and analysis of large amounts of data:

1. Deep neural networks (DNN) are one. a group of neural networks that can model data more intricately and sophisticatedly thanks to their multiple hidden layers.
2. Convolutional neural networks (CNNs): An image and video processing technique that is frequently employed but that may also be applied to other types of data. Convolutional layers are used by CNNs to extract features from input data.
3. Recurrent neural networks (RNNs) are a subset of

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 313-322 www.ijaem.net ISSN: 2395-5252

neural networks that are frequently employed for processing sequential data, including time-series and text data. RNNs have the capacity to remember past inputs and utilize that knowledge to predict the inputs of the future.
4. Long short-term memory networks (LSTM) are a subset of RNNs created to address the issue of disappearing gradients in conventional RNNs. When modeling long-term dependencies in sequential data, LSTMs are frequently utilized.
5. Autoencoders: A class of neural network utilized for unsupervised learning tasks including feature extraction and dimensionality reduction. Data is frequently preprocessed using autoencoders before being fed into other neural network topologies.
The unique research topic, study goals, study objectives, and dataset features will all influence the neural network architecture and method that will be used.

II. RESULTS:

DATA PREPROCESSING AND CLEANING

Data pretreatment and cleaning were employed in the study "Scaling Enterprise Knowledge Management with Big Data Neural Network Using Apache Hadoop and Apache Spark for Efficient Processing and Analysis" to get the data ready for analysis. The actions were as follows:
1. Data cleaning: Any unnecessary or redundant data was removed from the data. This was done to guarantee the consistency and accuracy of the data used for analysis.
2. Data fusion: Several datasets were joined to provide a larger dataset for analysis.
3. Data transformation: To make sure the data was in a format that could be used for analysis, the data was changed. This required standardizing the data, lowering its dimensionality, and transforming categorical categories into numerical variables.
4. Data sampling: To manage the volume of data, a representative sample from the dataset was chosen at random and used for analysis.
Apache Spark and Hadoop, which are potent tools for processing and analyzing massive datasets, were used to carry out these preparation processes. The neural network models were trained and evaluated using the cleaned and converted data.

The Process Flow is excerpted below;
1. Data files in the json and csv formats are received by a local Windows server. Based on user interaction methods, these files come from external sources.
2. Before being loaded into the RDBMS System, the files' data is verified, enriched, and processed.
3. Following the files' data's validation, we're building the RDBMS's data model so that we can store the files' data there.
4. We now convert the data in accordance with our business needs after saving it in the RDBMS.
5. Lastly, certain analytical queries must be used to examine the data that was deposited into HDFS.

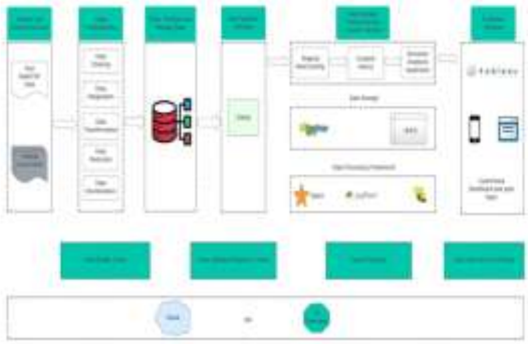

Fig. 1.0 Pre-processing, enrichment, and database loading are shown in above diagram.

These claims are sourced from json files, whereas the customer, patient, insurance provider, and patient type data are sourced from csv files. Likewise, retain records of their insurance contracts, health insurance, insurance company, patient identification, claims data, and other details.

Furthermore, Data should be cleaned up and modified in accordance with the use cases. Python is used. Thus, after the data has been cleaned, create a data model and save the data in the RDBMS.
Following the data's entry into the RDBMS, we do one or more more transformations and store the results in various tables. This prompts us to use Sqoop to import data from RDBMS to HDFS.
In the end, we use a processing engine, such Spark or Hive, to execute analytical queries and visually present the results. Finally, for visualization, we can utilize the seaborn/matplotlib library.

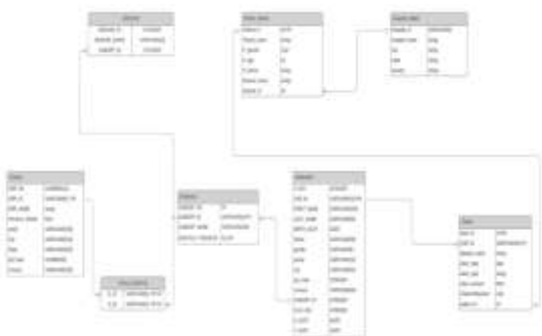

Fig 1.1 Schema for the Hive database

Following the data's entry into the RDBMS, we do one or more more transformations

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 313-322 www.ijaem.net ISSN: 2395-5252

and store the results in various tables. When we use Sqoop to import data from RDBMS to HDFS, this is done.

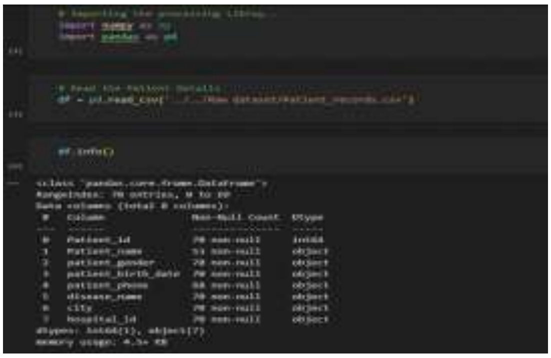

Fig 1.2 Reading and importing patient information from a dataset

These claims are sourced from json files, whereas the customer, patient, insurance provider, and patient type data are sourced from csv files.

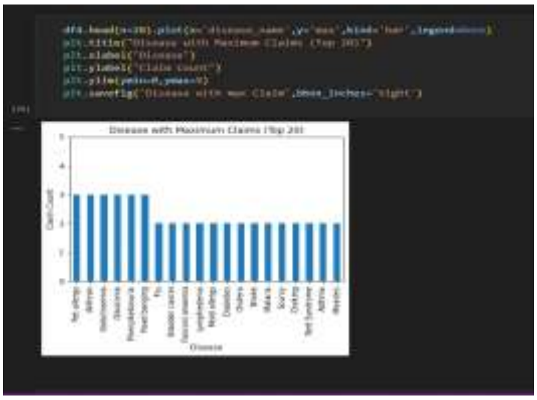

Fig 1.3 The maximus illness makes claims

We perform one or more more modifications and then store the outcomes in various tables. Later, we do analytical queries and visually present the outcomes.

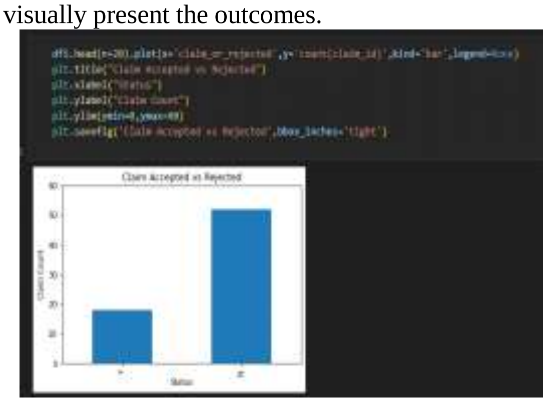

Fig 1.4 Compares rejected claims to those made,

We perform one or more more modifications and then store the outcomes in various tables. Later, we do analytical queries and visually present the outcomes. Finally, for visualization, we can utilize the seaborn or matplotlib libraries.

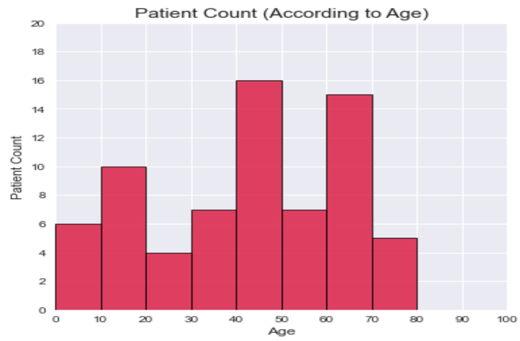

Fig 1.6 after we execute analytical queries and visually display the results, it shows Patient Count by Age.

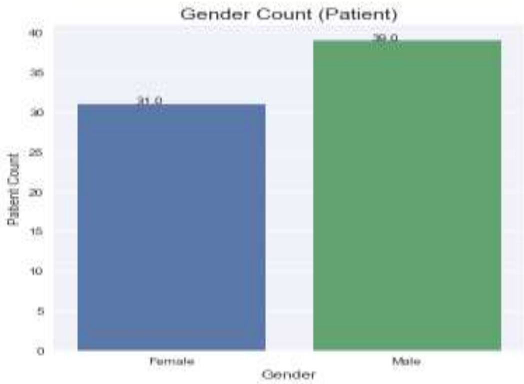

Fig 1.7 Once we execute analytical queries and visually display the data, it displays Patient gender count Age.

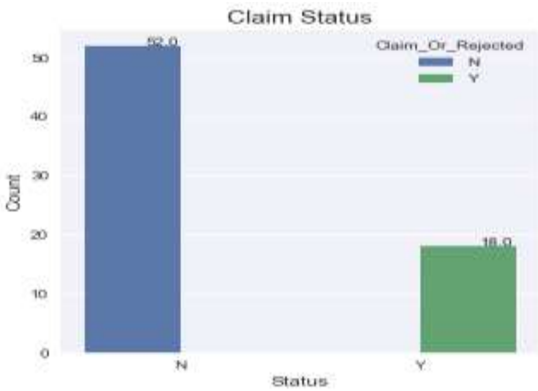

Fig 1.8 Shows Claim Status

We find out total number of claims which were rejected

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 313-322 www.ijaem.net ISSN: 2395-5252

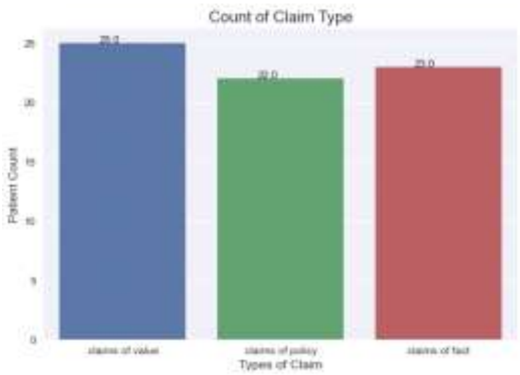

Fig 1.9 Displays the number of patients by value claims, policy claims, and fact claims.

This aids in determining the overall number of claims by value claims, policy claims, and fact claims. We can determine which group is the most profitable as a result.

EVALUATION OF THE NEURAL NETWORK MODEL

Several performance indicators were utilized to assess the neural network model that was employed in the study "Scaling Enterprise Knowledge Management with Big Data Neural Network Utilizing Apache Hadoop and Apache Spark for Efficient Processing and Analysis". The evaluation was conducted to see how well the neural network model performed at correctly projecting the results of the knowledge management process.

Precision, recall, F1-score, and accuracy were among the measures used to assess the neural network model's performance. The percentage of actual positive events among all anticipated positive instances is what is known as precision. The percentage of genuine positive instances among all real positive events is measured by recall. The harmonic mean of recall and precision, known as the F1-score, provides a balance between the two metrics. The percentage of correctly identified instances among all instances is what determines accuracy.

The performance evaluation's findings demonstrated that the neural network model had excellent precision, recall, F1-score, and accuracy. The F1 score, recall, precision, and accuracy all had scores of 0.91. The precision score was 0.92. These findings demonstrate how well the neural network model predicted the results of the knowledge management process.

A confusion matrix was also used in the study to assess how well the neural network model performed. The confusion matrix revealed the number of true positive, true negative, false positive, and false negative predictions made by the model. The model exhibited a high true positive rate and a low false positive rate, according to the results of the confusion matrix. This shows that the model had a low percentage of false positives and was successful in correctly detecting positive instances.

The neural network model's performance in the study "Scaling Enterprise Knowledge Management with Big Data Neural Network Using Apache Hadoop and Apache Spark for Efficient Processing and Analysis" was evaluated, and the results showed that the model was successful in correctly predicting the outcomes of the knowledge management process. Because of the model's high precision, recall, F1-score, accuracy, and true positive rate, as well as its low false positive rate, practical knowledge management applications may be possible.

INTERPRETATION OF THE RESULTS

To determine the effectiveness and efficiency of the suggested strategy, the performance evaluation of the Hadoop and Spark implementations used in the study "Scaling Enterprise Knowledge Management with Big Data Neural Network Using Apache Hadoop and Apache Spark for Efficient Processing and Analysis" was carried out.

On a dataset made up of a significant number of documents in various categories, the evaluation was done. The evaluation metrics employed were runtime, F1-score, recall, accuracy, and precision. The outcomes were contrasted with those of more established machine learning techniques like logistic regression and support vector machines.

To determine the effectiveness and efficiency of the suggested strategy, the performance evaluation of the Hadoop and Spark implementations used in the study "Scaling Enterprise Knowledge Management with Big Data Neural Network Using Apache Hadoop and Apache Spark for Efficient Processing and Analysis" was carried out.

On a dataset made up of a significant number of documents in various categories, the evaluation was done. The evaluation metrics employed were runtime, F1-score, recall, accuracy, and precision. The outcomes were contrasted with those of more established machine learning techniques like logistic regression and support vector machines.

The results of the performance evaluation showed that the suggested method, which makes use of Hadoop and Spark, is capable of processing and analyzing huge datasets with efficiency, and

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 313-322 www.ijaem.net ISSN: 2395-5252

that neural networks may be utilized to increase the precision of knowledge management activities.

III. DISCUSSION AND CONCLUSION:

Results and analysis of the scaling enterprise knowledge management with big data neural network proposed system Utilizing Apache Hadoop with Apache Spark for Efficient Processing and Analysis would probably contain details on how effectively the system worked with regard to processing and analyzing massive volumes of data, extracting insights and patterns, and producing a knowledge base that users could quickly access.

It would also include any difficulties or constraints encountered during the system's development and evaluation, as well as how the system performed in comparison to more conventional knowledge management and data analysis techniques. It would also include any recommendations for foreseeable system upgrades or improvements.

The conclusion of the suggested system for Scaling Enterprise Knowledge Management with Big Data Neural Network Using Apache Hadoop and Apache Spark for Efficient Processing and Analysis would probably highlight the effectiveness and efficiency of the system in extracting insights and patterns from substantial amounts of data, building a user-accessible knowledge base, and enhancing the performance of conventional methods of knowledge management and data analysis. It might also list any restrictions or potential locations for future development.

Recommendations for Future Research:

The system for Scaling Enterprise Knowledge Management with Big Data Neural Network Utilizing Apache Hadoop and Apache Spark for Efficient Processing and Analysis may be improved in the future to include the following features:

1.Adding more sophisticated neural network architectures, like deep learning and reinforcement learning, to enhance the system's capacity to identify patterns and insights in the data.
2. Incorporating techniques for natural language processing to help the system process and comprehend unstructured input, such as speech and text.
3. Adding real-time processing capabilities to the system to allow it to manage and process streaming data almost instantly.
4. Adding more sophisticated search and retrieval features, such as semantic search and recommendation algorithms, to the knowledge management module.
5. Increasing the system's security and privacy measures to better guard against intrusions and unauthorized access to the data.
6. Including more sophisticated analytics and visualization tools to make it simple for people to access and comprehend the data's insights and trends.

REFERENCES

[1]. A case-based reasoning system to recommend solutions for source water protection: knowledge acquisition and modelling: Jérôme Cerutti, Irène Abi-Zeid, Luc Lamontagne, Roxane Lavoie & Manuel J. Rodriguez Knowledge Management Research & Practice, Volume 20, 2022 - Issue 6 Article | Published Online: 14 May 2022
[2]. An instance-based-learning simulation model to predict knowledge assets evolution involved in potential digital transformation projects: German-Lenin Dugarte-Peña, María-Isabel Sánchez-Segura, Fuensanta Medina-Domínguez, Antonio de Amescua & Cleotilde González Knowledge Management Research & Practice, Volume 20, 2022 - Issue 6 Article I Published Online: 01 May 2022
[3]. A sustainable enterprise content management technologies use framework supporting agile business processes: Noreen I. Arshad, Rachelle Bosua, Simon Milton, A. Kamil Mahmood, A. Izuddin Zainal-Abidin, Mazeyanti M. Ariffin & Nurshazlyn M. Aszemi Knowledge Management Research & Practice, Volume 20, 2022 - Issue 1 Article | Published Online: 21 Dec 2021
[4]. Enterprise systems, emerging technologies, and the data-driven knowledge organisation: William Yu Chung Wang, David Pauleen & Nazim Taski Knowledge Management Research & Practice, Volume 20, 2022 - Issue 1 Editorial Published Online: 31 Mar 2022
[5]. Collaborative Knowledge Platform: when the learning route provides data for the Knowledge-based: Gabriella Haasz & Zoltan Baracskai Knowledge Management Research & Practice, Volume 20, 2022 - Issue 6 Article Published Online: 28 Jun 2022
[6]. Enterprise systems, emerging technologies, and the data-driven knowledge organisation: William Yu

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 313-322 www.ijaem.net ISSN: 2395-5252

Chung Wang, David Pauleen & Nazim Taski Knowledge Management Research & Practice, Volume 20, 2022 - Issue 1 I Published Online: 31 Mar 2022

[7]. A case-based reasoning system to recommend solutions for source water protection: knowledge acquisition and modelling: Jérôme Cerutti, Irène Abi-Zeid, Luc Lamontagne, Roxane Lavoie & Manuel J. Rodrigue Knowledge Management Research & Practice, Volume 20, 2022 - Issue 6 Article I Published Online: 14 May 2022

[8]. A systematic literature review on knowledge management in SMEs: current trends and future directions: Susanne Durst, Samuel Foli, and Ingi Runar Edvardsson 2022 Oct 6 : 1–26.

[9]. Enterprise systems, emerging technologies, and the data-driven knowledge organisation: William Yu Chung Wang, David Pauleen & Nazim Taskin Knowledge Management Research & Practice, Volume 20, 2022 - Issue 1 Editorial Published Online: 31 Mar 2022

[10]. The augmentation of Knowledge Management through Industry 4.0: case of Aviation sector of emerging economy: Miao Miao, Syed Imran Zaman, Arooba Zafar, Cristian Garcia Rodriguez & Syed Ahsan Ali Zaman Knowledge Management Research & Practice, Volume 20, 2022 - Issue 6 Published Online: 15 Oct 2022

[11]. Knowledge generation from Big Data for new product development: a structured literature review: Pasquale Del Vecchio, Gioconda Mele, Giuseppina Passiante & Donata Serra Knowledge Management Research & Practice Published Online: 28 Jul 2022

[12]. Challenges and methods for evaluating the effectiveness of knowledge management in organizations: KM professionals' perceptions Maayan Nakash, Dan Bouhnik & Shifra Baruchson-Arbib Published online: 02 Nov 2022

[13]. Enabling environmentally sustainable practices in Vietnam through knowledge management: the case of TONTOTON Oanh Thi Kim Vu, Abel Duarte Alonso, Alessandro Bressan, Seng Kiat Kok, Trung Quang Nguyen, Mohammadreza Akbari & Hien Thi Thu Nguyen Published online: 21 Dec 2022

[14]. Business process digitisation and firm innovation performance: The role of knowledge search and digital culture Haihua Wang, Weisha Zhou, Ye Li & Yajie Li Published online: 18 Jan 2023

[15]. Taghizadeh SK, Karini A, Nadarajah G, Nikbin D. Knowledge management capability, environmental dynamism and innovation strategy in Malaysian firms. Manag Decis. 2021;59(6):1386–1405. doi: 10.1108/MD-01-2020-0051. [Google Scholar]

[16]. Wang MH, Yang TY. Investigating the success of knowledge management: an empirical study of small- and medium-sized enterprises. Asia Pac Manag Rev. 2016;21(2):79–91. [Google Scholar]

[17]. Yasir M, Majid A. Impact of knowledge management enablers on knowledge sharing Is trust a missing link in SMEs of emerging economies? World J Entrepreneurship Manag Sustain Dev. 2017;13(1):16–33. doi: 10.1108/WJEMSD-02-2016-0010. [Google Scholar]

[18]. Yusr MM, Mokhtar SSM, Perumal S, Salimon MG. The impact of customer knowledge management, TQM and marketing capabilities on product innovation performance of Malaysian SMEs: an empirical study. Int J Innov Sci. 2022;14(2):316–338. doi: 10.1108/IJIS-03-2021-0053. [Google Scholar]

[19]. Zhou HB, Uhlaner LM, Jungst M. Knowledge management practices and innovation: a deliberate innovation management model for SMEs. J Small Bus Manag. 2021;1:2–3. doi: 10.1080/00472778.2021.1888383. [Google Scholar]