International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

A SURVEY ON ANALYSIS AND PREDICTION OF DATA USING DATA SCIENCE

D.Kavin Kumar[1] ,Ms.S.Kavitha M.E[2]
[1]PG Scholar,Department of Computer Science and Engineering, Nandha Engineering College(Autonomous),Erode,Tamilnadu,India[1]
[2]Assisstant Professor,Department of Computer Science and Engineering, Nandha Engineering College(Autonomous),Erode,Tamilnadu,India[2]

ABSTRACT
Intelligent technology development is gaining traction in the sphere of education. The increasing rise of educational data suggests that standard processing methods may be limited and distorted. As a result, rebuilding data mining research technologies in the education industry has become necessary. Becoming more visible To avoid erroneous assessment findings and to anticipate students' future performance, this research analyses and predicts students' academic achievement using applicable clustering, discriminating, and convolution neural network theories. To begin, this work suggests that the clustering-number determination be optimized by employing a statistic that has never been employed in the K-means approach. The clustering impact of the K-means method is next assessed using discriminate analysis. The Convolutional neural network is presented for training and testing with labeled data. The produced model can be used to forecast future performance. Finally, the efficacy of the constructed model is tested using two metrics in two cross validation procedures in order to validate the prediction findings. The experimental findings show that the statistic not only addresses the objective and quantitative problem of determining the clustering number in the K-means method, but also enhances the predictability of the outcomes.
KEYWORDS: Academic Performance, Clustering Analysis, Convolutional Neural Network (CNN), Discriminate Analysis, Educational Data Mining

INTRODUCTION

1.1 ACADEMIC PERFORMANCE

Academic prediction on student performance in classroom instruction is commonly employed using educational data mining approaches. However, the majority of previous studies was investigated and compared student coursework performance to test passing grades. We conducted study in this paper to determine the significance and influence of student background, student social activities, and student coursework accomplishment in predicting student academic performance. In secondary school, supervised educational data mining techniques such as Nave Bayesian, Multilayer Perception, Decision Tree J48, and Random Forest were employed to predict math achievement. On the final grade, the prediction was done on a 2-level classification and a 5-level classification. According to the experimental results, student background and student social activities were significant predictors of student performance on 2-level categorization. The model may be used to predict student performance early on, which can aid in increasing student performance on the topic.

1.2 CLUSTERING ANALYSIS

Clustering is the classification of a collection of diverse data objects as related things. A data cluster is represented by one group. In the cluster analysis, data sets are split into separate groups based on their resemblance. A label is applied to each collection of data once it has been classified into several groups. It aids in responding to changes by categorizing them. So, if we define clustering in data mining, we can say that the process of clustering in data mining consists of grouping a set of abstract objects into groups of related items. Cluster analysis is the process of separating and storing them in these categories. Cluster Analysis in Data Mining refers to the discovery of groups of things that are

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

similar to each other but distinct from the objects in other groups. Clustering is a data analytics procedure that divides data sets into groups or classes based on data similarity. The classes are then labeled according to their data kinds. Going through the clustering in data mining example might help you better grasp the analysis. Data Mining Cluster Analysis Applications, including image processing, data analysis, pattern identification, market research, and many more. Companies can use data clustering to uncover new groupings in their customer database.

1.3 CONVOLUTIONAL NEURAL NETWORK (CNN)

A Convolutional neural network (CNN or convent) is a machine learning subset. It is one of several types of artificial neural networks utilized for diverse applications and data sources. A CNN is a type of network design for deep learning algorithms that is primarily utilized for image recognition and pixel data processing applications. There are different forms of neural networks in deep learning, but CNNs are the network design of choice for identifying and recognizing things. As a result, they are ideal for computer vision (CV) jobs and applications requiring object recognition, such as self-driving cars and facial recognition. Another sort of neural network that can find important information in both time series and picture data is CNN. As a result, it is extremely useful for image-related tasks including image identification, object categorization, and pattern recognition. A CNN uses linear algebra techniques such as matrix multiplication to discover patterns in images. CNNs can categories audio and signal data as well. The design of a CNN is similar to the connection network of the human brain. CNNs, like the brain, are made up of billions of neurons that are organized in a certain fashion. In reality, the neurons in a CNN are organized similarly to the frontal lobe of the brain, which is responsible for processing visual stimuli.

1.4 DISCRIMINANT ANALYSIS

Data mining is a collection of analytical techniques used to uncover new trends and patterns in massive databases. These data mining techniques stress visualization to thoroughly study the structure of data and to check the validity of the statistical model fit which leads to proactive decision making. Discriminate analysis is one of the data mining techniques used to discriminate a single classification variable using multiple attributes. Discriminate analysis also assigns observations to one of the pre-defined groups based on the knowledge of the multi-attributes. When the distribution within each group is multivariate normal, a parametric method can be used to develop a discriminate function using a generalized squared distance measure. The classification criterion is derived based on either the individual within-group covariance matrices or the pooled covariance matrix that also takes into account the prior probabilities of the classes. Non-parametric discriminate methods are based on non-parametric group-specific probability densities. Either a kernel or the k-nearest-neighbor method can be used to generate a non-parametric density estimate in each group and to produce a classification criterion. The performance of a discriminate criterion could be evaluated by estimating probabilities of mis-classification of new observations in the validation data.

1.5 EDUCATIONAL DATA MINING

Educational Data Mining is a new subject focused with creating ways for studying the unique and increasingly large-scale data generated by educational settings and applying those approaches to better understand students and the environments in which they learn. Whether educational data is derived from students' use of interactive learning environments, computer-supported collaborative learning, or administrative data from schools and universities, it frequently has multiple levels of meaningful hierarchy, which must often be determined by data properties rather than in advance. Time, chronology, and context are other essential considerations in the examination of educational data. The International Educational Data Mining Society's goal is to foster collaboration and scientific development in this new discipline by organizing the EDM conference series, the Journal of Educational Data Mining, and mailing lists, as well as developing community resources to facilitate data and technique sharing. EDM is an abbreviation for Educational Data Mining. It may be

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

characterized as a strategy for locating particular sorts of data from the educational system and utilizing those strategies to better understand students and the system.

2. LITERATURE SURVEY

2.1 A DATA ANALYTICS SUITE FOR TYPE 2 DIABETES EXPLORATORY PREDICTIVE AND VISUAL ANALYSIS

In this work, NADA Y et al. offer Long-term management of chronic illnesses such as Type 2 Diabetes (T2D) necessitates individualized care for patients due to differences in patient features and responsiveness to a specific line of treatment. Treatment. The availability of enormous amounts of electronic T2D patient data allows for the use of big data analysis to get insights on illness manifestation and its impact on patients. Data science in healthcare has the ability to uncover hidden knowledge in databases, corroborate current knowledge, and help in therapy personalization. We describe in this study a data analytics suite for T2D disease management that helps doctors and researchers to detect connections between various patient biological indicators and T2D-related problems. The analytics package includes exploratory, predictive, and visual analytics, as well as multi-tier categorization of T2D patient profiles that correlate them with certain diseases, T2D associated complication risk prediction, and prediction of patient response to a given line of treatment.

2.2 A DATA SHARING PROTOCOL TO REDUCE CLOUD STORAGE SECURITY AND PRIVACY RISKS IN THE BIG DATA ERA

In this work, SI HAN et al. suggest A cloud-based large data sharing system makes use of a cloud service provider's storage facilities. to exchange info with authorized individuals In contrast to traditional solutions, cloud providers store shared data in massive data centers outside the data owner's trust zone, which may result in data corruption. Confidentiality. This work presents a secret sharing group key management protocol (SSGK) to prevent unwanted access to the communication process and shared data. In contrast to previous efforts, a group key is utilized to encrypt the shared data in SSGK, and a secret sharing technique is employed to distribute the group key. Extensive security and performance evaluations show that our approach significantly reduces the security and privacy hazards of data sharing in cloud storage while also saving roughly 12% of storage space. In this research, we present a unique group key management mechanism for cloud storage data sharing. We employ RSA and verified secret sharing in SSGK to give the data owner fine-grained control over outsourced data without depending on a third party. Furthermore, we provide a comprehensive analysis of probable attacks and corresponding responses, demonstrating that GKMP is secure even under weaker assumptions.

2.3 A MANUFACTURING PRODUCTION GENERIC DATA ANALYTICS SYSTEM

The growth in the quantity of manufacturing information accessible implies that big data may be collected and, with suitable deep analysis, might be of considerable use to manufacturers, as argued by Hao Zhang et al. in this research. However, the majority of small businesses cannot justify the expense of a skilled data analytics team To overcome this issue, a generic data analytics system, Generic Manufacturing Data Analytics system (GMDA), is presented in this work. This system can handle the majority of manufacturing data analytics jobs, and users may undertake data analysis even if they have no prior expertise or experience with data analytics. To build such a system, we created GMDL, an abstract language for describing manufacturing data analytics operations. Several algorithms were chosen, modified, optimized, and eventually integrated into the system with the goal of industrial data analytics. GMDA produced some notable strategies, such as an appropriate algorithm selection strategy and an optimal parameter determination algorithm. Case examples demonstrate the system's applicability and dependability. Manufacturing data that is general and has a low user threshold

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

2.4 A REAL-TIME DATA FUSION METHODOLOGY FOR LOCALIZED BIG DATA ANALYTICS

In this research, SOHAIL JABBAR et al. propose that classic big-data analytical methodologies leverage data clustering as tiny buckets while offering distributed computing across several child nodes. These tactics, in particular, bring the concerns to light. in terms of network bandwidth, specialist tools, and programmes that cannot be learned in a short amount of time Furthermore, raw data created by IoT generating big data is capable of producing very unstructured and diverse data. This type of data becomes a difficult challenge for real-time analytics. To alleviate real-time analytical problems, it is extremely beneficial to have computational values available locally rather than through distributed resources. This work suggests a merger of three diverse data models, such as relational, semantic, and big data-based data and metadata, with associated challenges and expanded possibilities.

2.5 A NEW SPATIOTEMPORAL DATA MODEL FOR VISUALIZING AND ANALYZING RIVER WATER QUALITY

YINGUO QIU et al. argued in this study that river water quality (RWQ) data has evident geographical and temporal distribution features, and tables are traditionally used for storing of RWQ multi-period monitoring data; nevertheless, Because of its dispersion, neither effective display nor proper analysis of the given data can be accomplished. In this research, a unique spatiotemporal data model for RWQ data is suggested in order to facilitate data representation and spatiotemporal analysis. The basic element of river spaces in this model is a spatial point that contains both location and dynamic water quality information, and methods for expanding a point to a line segment, a flat surface, and a cube are designed to make this model applicable to different generalizations of river spaces. Furthermore, a temporal data storage structure is devised to provide efficient inquiry and advanced analysis of RWQ data while reducing occupied memory space. Finally, case studies are conducted on RWQ data by performing 3D visualization, trend analysis, and anomaly identification, with the results demonstrating that tridimensional representation of RWQ data can be realized efficiently, the computational complexity is significantly reduced, and the occupied memory space of monitoring data is effectively economized. As a result, the suggested spatiotemporal data model can help with RWQ data presentation and advanced analysis.

2.6 ANALYSIS AND PREDICTION OF ACADEMIC PERFORMANCE OF STUDENTS USING EDUCATIONAL DATA MINING

In this research, GUIYUN FENG et al. claim that the development of intelligent technologies is gaining appeal in the sphere of education. The quick the expansion of educational data suggests that standard processing methods may have limits and distortion. As a result, recreating data mining research technology in the education area has become increasingly important. To avoid erroneous assessment findings and to anticipate students' future performance, this research analyses and predicts students' academic achievement using applicable clustering, discriminating, and convolution neural network theories. To begin, this work suggests that the clustering-number determination be optimized by employing a statistic that has never been employed in the K-means approach. The clustering impact of the K-means method is next assessed using discriminate analysis. The Convolutional neural network is presented for training and testing with labeled data. The produced model can be used to forecast future performance. Finally, the efficacy of the constructed model is tested using two metrics in two cross validation procedures in order to validate the prediction findings.

2.7 IN SOCIAL SCIENCES, ANALYZING OBJECTIVE AND SUBJECTIVE DATA: IMPLICATIONS FOR SMART CITIES

In this study, LAURA ERHAN et al. claim that the ease of deployment of digital technologies and the Internet of Things allows us to conduct large-scale sociological studies and collect massive volumes of data from our cities. In this case, In this study, we use machine learning and data science

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

approaches to examine a unique method of interpreting data from social science research. This allows us to optimise the knowledge acquired from these types of investigations by combining objective (sensor data) and subjective data (direct input from the users). The pilot project aims to get a better understanding of how residents engage with urban green spaces. In Sheffield, England, a field experiment with 1870 volunteers was conducted across two time periods (7 and 30 days). Both factual and subjective data were obtained using a Smartphone app. As someone visited any of the publicly available green places, their whereabouts was tracked.

2.8 BIG DATA ANALYTICS AND MINING FOR CRIME DATA VISUALIZATION AND TREND FORECASTING

In this research, MINGCHEN FENG et al. argued that big data analytics (BDA) is a systematic technique for evaluating and recognizing diverse types of data. big volume of data contains patterns, relationships, and trends. We use BDA to criminal data in this article. Where exploratory data analysis is carried out for trend prediction and visualization Several cutting-edge data mining and deep learning techniques are applied. Following statistical analysis and visualization, several fascinating facts and trends in crime data from San Francisco, Chicago, and Philadelphia are identified. The predicted findings reveal that the Prophet model and Keras stateful LSTM outperform neural network models, with three years of training data being shown to be the ideal size. These promising results will help police departments and law enforcement agencies better understand crime concerns and give insights that will allow them to follow activities, estimate the likelihood of events, allocate resources effectively, and optimise decision making. In this work, we used a variety of cutting-edge big data analytics and visualization tools to evaluate crime big data from three US cities, allowing us to detect patterns and obtain trends.

2.9 BIG DATA IN MOTION: A FRAMEWORK FOR VEHICLE-ASSISTED URBAN COMPUTING IN SMART CITIES

In this research, MURK et al. claim that smart cities are envisioned to enhance societal well-being through effective management of Internet of Things resources and the data created by these resources. However, the massive number of such devices will result in unprecedented data growth, posing capacity challenges. acquisition, transportation from one area to another, storage, and analysis Traditional networks are insufficient to accommodate the transfer of this massive volume of data, which becomes costly in terms of both latency and energy usage. Alternative data communication methods are consequently necessary to accommodate the massive data generated by smart cities. In this research, we suggest an efficient data-transfer architecture based on volunteer cars, in which vehicles are used to transport data in the direction of the destination. Through urban computing, the framework encourages citizen engagement and fosters self-belonging, social awareness, and energy conservation. The suggested framework can also assist the research community in quickly benchmarking their own route selection algorithms. Furthermore, we conducted a thorough evaluation of the suggested framework using realistic models of cars, routes, data-spots, data chunks to be transferred, and energy consumption.

2.10 IMPROVED DATA ACQUISITION AND STORAGE SYSTEM BASED ON BIG DATA FOR INDUSTRIAL DATA PLATFORM DESIGN

In this research, DAOQU GENG et al. suggest that a big data-based acquisition and storage system (ASS) plays a significant role in the design of an industrial data platform. Many large data frameworks have compression and serialization built in. method. These technologies cannot satisfy the objectives of industrial production information management since they are time-consuming and need large amounts of storage. We propose an upgraded industrial big data platform based on existing big data frameworks in order to minimize data processing time while needing less data storage space. This study, in particular, focuses on assessing the influence of various compression and serialization methods on big data platform performance and attempting to select appropriate compression and serialization methods for the industrial data platform.

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

2.11 CITYPULSE: A SMART CITY DATA ANALYTICS FRAMEWORK ON A LARGE SCALE

In this study, DAN PUIU et al. claim that our world and lifestyles are changing in a variety of ways. Communication, networking, and computer technologies are among the most powerful enablers influencing our lives today. Data in digital form and connected worlds of physical objects, people, and devices are rapidly changing the way we work, travel, socialize, and interact with our surroundings, having a profound impact on a variety of domains including healthcare, environmental monitoring, urban systems, and control and management applications, among others. Cities are now facing an increase in demand for services that have an influence on people's daily lives. The City Pulse framework facilitates the implementation of smart city services through a distributed system for semantic discovery, data analytics, and interpretation of large-scale (near-)real-time Internet of Things and social media data streams. The idea is to liberate apps from silos and enable cross-domain data integration. The City Pulse framework integrates multimodal, mixed quality, uncertain, and incomplete data to provide trustworthy, dependable information and continually changes data processing algorithms to satisfy end-user information quality needs.

2.12 ONLINE WATER QUALITY MONITORING USING CONNECTED SENSORS, INNOVATIVE SENSOR DEPLOYMENT, AND INTELLIGENT DATA ANALYSIS

In this research, Libu Manjakkal et al. argue that sensor technology for water quality monitoring (WQM) has improved in recent years. The most cost-effective Sensitized technologies that can measure the fundamental physical-chemical-biological (PCB) variables autonomously are now widely accessible and are being installed on buoys, boats, and ships. However, due to a lack of standardized methodologies for data collecting and processing, spatiotemporal fluctuation of critical parameters in water bodies, and novel pollutants, there is a mismatch between data quality, data gathering, and data analysis. Such gaps can be filled by deploying a network of multiparametric sensor systems in bodies of water with autonomous vehicles such as marine robots and aerial vehicles to widen data coverage in space and time. Intelligent algorithms [for example, artificial intelligence (AI)] might also be used for standardized data analysis and forecasting. This article provides an in-depth examination of WQM sensors, deployment, and analytic technologies. A network of networked water bodies might improve worldwide data comparability and enable WQM on a global scale to solve global concerns in food (e.g., aqua/agriculture), drinking water, and health (e.g., water-borne illnesses). WQM linked sensor technology may give the answer to the present mismatch between data quality, data collection, and data analysis, and to improve global data interoperability With this in mind, this article has examined major sensing technologies, sensor deployment strategies, and developing data analysis approaches. The review looked at several sensing materials, substrates, and sensor architectures, including multisensory patches. Various sensor interface electronics and communication system components, as well as innovative deployment tactics employing sensor zed buoys, drones, and underwater robotic vehicles, have been considered for data collection. Diverse methodologies for sensor data analysis are briefly reviewed, as are the possible prospects for real-time WQM with AI.

2.13 ACUTE CORONARY SYNDROME SECONDARY PREVENTION USING DATA SCIENCE ANALYSIS AND PROFILE REPRESENTATION

The analysis of vast volumes of data from electronic medical records (EMRs) and everyday clinical practice data sources has garnered growing attention in recent years, according to ANTONIO GARCA-GARCA et al. However, there are few systematic ways. have been proposed to aid in the extraction of the wealth and diversity of information contained in various data sources. Acute Coronary Syndrome (ACS) statistics, in particular, are accessible in many hospitals and health units since ACS has a high morbidity and fatality rate. This paper presents a method called Data Science Analysis and Representation (DSAR) for examining and using scientific information content in restricted ACS samples in a univariate manner. To deliver robust, cross-sectional, and non-parametric statistical tests on categorical and metric variables, DSAR employs Bootstrap Resembling. It also

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

creates an instructive graphical representation of the database variables, which aids in the interpretation of the results and the identification of the key factors.

2.14 DATA SCIENCE AS POLITICAL ACTION: ROOTING DATA SCIENCE IN JUSTICE POLITICS

In this study, Ben Green et al. suggest the area of data science has embraced ethics in reaction to public scrutiny of data-driven algorithms. principles and training Although ethics can assist data scientists focus on some normative elements of their work, such attempts fall short of producing data science that avoids societal harms and supports social justice. In this post, I suggest that data science should have a political bent. Data scientists must acknowledge themselves as political actors involved in normative social constructs and evaluate their work in terms of its downstream effects on people's lives. First, I explain why data scientists must acknowledge their role as political agents. In this part, I react to three popular arguments used by data scientists when asked to take political stances on their work. In response to these arguments, I explain why seeking to stay apolitical is a political attitude in and of itself—a fundamentally conservative one—and why data science's efforts to promote "social good" rely dangerously on unarticulated and increment list political assumptions. Then I suggest a paradigm for how data science may progress toward a deliberative and rigorous social justice politics. I see the process of establishing a politically engaged data science as a four-stage progression.

2.15 DISTRIBUTED DATA STRATEGIES FOR LARGE-SCALE DATA ANALYSIS ACROSS GEOGRAPHICALLY DISTRIBUTED DATA CENTERS

In this study, TAMER Z. EMARA et al. claim that as the volume of data rises fast, storing big data in a single data centre is no longer practical. As a result, businesses have devised two scenarios for storing massive data in many data centers. The company's huge data are scattered across numerous data centers in the first scenario, with no data replication. In the second scenario, data is kept in several data centers as well, but critical data is replicated in these data centers to enhance data safety and availability. However, evaluating massive data scattered across several data centers becomes difficult in these cases. We offer two data distribution algorithms in this research to allow big data analysis across geographically separated data centers. We employ the latest Random Sample Partition data model in these tactics to turn huge data into sets of random sample data blocks and distribute these data blocks across numerous data centers, either without or with replication.

2.16 EDUCATIONAL INFORMATION PROBLEM-SOLVING DATA MINING TO SUPPORT PROGRAMMING LEARNING

In this work, MD. MOSTAFIZER RAHMAN et al. suggest that computer programming has gotten a lot of attention in the development of information and communication technologies in the real world. Keeping up with the rising need for highly qualified programmers One of the biggest issues in the ICT business is At this time, online judge (OJ) systems, in addition to classroom-based instruction, improve programming learning and practice chances. As a result, OJ systems have generated a vast amount of problem-solving data archives (solution codes, logs, and scores) that might be useful raw materials for programming education research. We present an educational data mining system to promote programming learning using unsupervised methods in this study. The framework consists of the following steps: I problem-solving data collection and preprocessing; (ii) MK-means clustering algorithm is used for data clustering in Euclidean space; (iii) statistical features are extracted from each cluster; (iv) frequent pattern (FP)-growth algorithm is applied to each cluster to mine data patterns and association rules; and (v) a set of suggestions are provided based on the extracted features, data patterns, and rules. To acquire the best results for clustering and association rule mining algorithms, many parameters are changed. Approximately 70,000 real-world problem-solving data from 537 students in a programming course (Algorithm and Data Structures) were utilized in the experiment. Furthermore, fake data was used in trials to illustrate the performance of the MK-means method.

2.17 DEEP NETWORK EXPLORATORY ANALYSIS FOR BIG SOCIAL DATA

Exploratory analysis, as proposed by CHAO WU et al. in this study, is an essential technique to acquire insight and discover undiscovered links from multiple data sources, particularly in the era of

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

big data. Traditional social science data paradigms The phases of feature selection, modeling, and prediction are followed by analysis. In this study, we offer a novel paradigm that does not need feature selection, allowing data to speak for itself without the need for deliberate feature selection. Furthermore, for massive social data, we suggest employing deep networks as a tool to investigate previously unknown correlations and capture complexity and non-linearity between goal variables and a large number of input attributes. The new paradigm is often a reasonably broad strategy that may be extensively applied in a variety of settings. We employ country-level indicator forecasting as a case study to demonstrate the paradigm's practicality. The steps are as follows: 1) data gathering and preparation, 2) modeling and experimentation. To eliminate data format inconsistencies, the data collecting and preparation section creates a data warehouse and performs the extract-transform-load procedure. Model setup and model structure changes are included in the modeling and experimentation portion to produce reasonably high accuracy on prediction findings at both the model and case levels. We discover certain trends concerning network capacity adjustment and the impact of time interval difference on test outcomes, both of which need additional investigation. In this study, we developed a new paradigm for conducting an investigation. An exploratory investigation of huge social data prediction using deep neural network models without feature selection.

2.18 IDP: AN INTELLIGENT DATA PREDICTION SCHEME BASED ON BIG DATA AND SMART SERVICES FOR PREDICTING SOIL HEAVY METAL CONTENT

In this research, FANG CHEN et al. suggest the error between the projected value and the real value is frequently substantial in the use of regression prediction using big data technologies. This research aims to decrease data prediction inaccuracy. offers a Smart Service Intelligent Data Prediction (IDP) method The core prediction model is Least Squares Support Vector Machine (LSSVM). Because there is no standard process for finding the major parameters of LSSVM, an enhanced Particle Swarm Optimization (MBPSO) algorithm is utilized to optimise the parameters of LSSVM concurrently. The fundamental downside of PSO is precocity as a result of the loss of population diversity. Based on this, MBPSO's Improvement method strives to create "More" and "Better" particles on a constant basis. To begin, MBPSO re-adjusted the inertia weight and learning factor to avoid the early elimination of particle diversity. Second, a renewable access method is developed in order to allow some of the vanished population to regrow. Finally, the concept of global optimum adjustment is proposed to assist particles in determining the best flying path. To validate the efficacy of MBPSO, 9 test functions are utilized to evaluate the algorithm's performance. MBPSO's optimization speed, best, and mean all perform best, according to the results.

2.19 BIG DATA MEETS PROCESSES: INTEGRATING DATA SCIENCE AND PROCESS SCIENCE

In this study, Wil van der Aalst et al. offer As more businesses embrace Big Data, it has become clear that the ultimate difficulty is to tie huge volumes of event data to highly dynamic operations. To maximize the usefulness of event data, events must be strictly controlled. Related to operational process control and management However, for the time being, the major focus of Big data technology is on storage, processing, and very rudimentary analytical activities. Big data projects are rarely focused on improving end-to-end procedures. We urge for improved integration of data science, data technology, and process science to overcome this mismatch. Data science techniques are process antagonistic, whereas process science approaches are model-driven and ignore the "evidence" concealed in the data. This is where process mining comes in. This editorial examines the relationship between data science and process science, as well as how process mining is related to Big data technologies, service orientation, and cloud computing. Companies and organizations all across the world are growing cognizant of the potential competitive advantage that quick and accurate "whole data" process mining based on (1) advanced discovery and visualization techniques and (2) the Big Data computational paradigm may provide them. However, as stated in this editorial, various scientific and technological obstacles must be addressed before process mining, data science, and Big data technologies can function in tandem. The IEEE Transactions on Services Computing special issue "Process Analysis Meets Big Data" includes several important contributions toward overcoming

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

the barriers that still impede us from realizing the full benefits of Big data approaches in Process Mining.

2.20 A DATA SCIENCE APPROACH TO EFFECTIVE NATURAL DISASTER RESPONSE

Natural catastrophes, according to GHULAM MUDASSIR et al., can inflict severe damage to buildings and infrastructures and kill thousands of people. These catastrophes are difficult to overcome, both by populations and by governments. Government officials Two difficult concerns must be handled in particular: first, establish an effective strategy to evacuate people, and then reconstruct houses and other facilities. An proper recovery strategy to evacuate people and begin repairing devastated regions on a priority basis can therefore be a game changer, allowing such horrible situations to be overcome efficiently. In this regard, we present DiReCT, an approach based on I a dynamic optimization model designed to timely formulate an evacuation plan for an earthquake-affected area, and ii) a decision support system based on a double deep Q Network capable of efficiently guiding the reconstruction of the affected areas. The latter operates by taking into account both the available resources and the demands of the many parties involved (for example, residents' social benefits and political goals). The foundation for both of the aforementioned solutions was a customized geographical data extraction method called "GisToGraph," which was created specifically for this purpose. To test the applicability of the whole strategy, we used extensive GIS data and information on urban land layout and building vulnerability in the historical city centre of L'Aquila (Italy). Several simulations were done on the constructed underlying network. First, we conducted trials to safely evacuate as many people as possible from a threatened region to a set of safe locations in the least amount of time. Then, using DDQN, we developed many rebuilding plans and chose the best ones, taking into account both the social benefits and the political interests of the building units. The ideas outlined here are part of a larger data science framework designed to generate an effective response to natural catastrophes. The GisToGraph algorithm's network represents the city map in terms of buildings, intersections, and roadways. Streets. Unlike other comparable algorithms, we can handle extra information required for evacuation planning and reconstruction, which is added as characteristics to network nodes and arcs. In terms of the evacuation planning model, we updated the linear optimization model initially developed by Arbib et al for building interior evacuation. The model has to be adjusted in terms of numerous parameters, as well as rescaled to a network of several orders of magnitude. To consider politicians' input, we analyzed all major factors .

SSII: Secured and High-Quality Steganography Using Intelligent Hybrid Optimization Algorithms for IoT SACHIN DHAWAN et al., has presented in this work Internet of Things (IoT) is an area where large amounts of data are transferred every second. The security of this data is a difficult problem; but, security issues may be addressed by the use of cryptography and steganography techniques. When it comes to user authentication and data protection, these strategies are critical. The proposed study proposes a highly secure approach based on the IoT protocol and steganography. This paper offers a picture steganography approach that employs a variety of techniques to ensure the security of secret data using a Binary bit-plane decomposition (BBPD) based image encryption technique. Following that, an adaptive embedding procedure based on the Salp Swarm Optimization Algorithm (SSOA) is developed to maximise payload capacity by adjusting different parameters in the steganographic embedding function for edge and smooth blocks. The SSOA technique is employed here to effectively locate the edge and smooth blocks. The hybrid Fuzzy Neural Network with a backpropagation learning method is then utilised to improve the stego picture quality. The stego pictures are then delivered to the destination via the highly secure IoT protocol. In compared to existing state-of-the-art technologies, the suggested steganography technology achieves better outcomes in terms of security, picture quality, and payload capacity. [21]

Using Error Probabilities and Integral Methods for Investigation Analysis for Software Fault PredictionIn this research, Karuppusamy et al. offer in-depth analysis of errors in the code phase using integral approaches that locate the problem in software. The data set repositories are gathered during the software product development life cycle model, which is then linked with a machine

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

learning method called Bayesian decision theory to detect error probability and anticipate unbound error during software fault prediction. Prior to this, the faults in the repository are predicted for a given data set using the error probability and error integral method, which identify the probability of error and correction, which is then applied with the Gaussian method to find the levels of the error probability with the minimum and maximum integral of acceptable faults in the repository. [22]

Self-Adaptive Approaches to Data Analytics Probability Distribution in Cloud Computing Resource Services for Infrastructure Hybrids Models

S. Prabhu et al., have claimed in this study that scientific research and experiments deliver superior answers in the cloud environment through dispersed data sources, giving clients a high level of data access. Cloud computing refers to the grouping of networks that deliver facilities at high speeds while ensuring security and connectivity amongst software applications. Cloud computing is a platform that may give solutions for huge data centres as well as meet client needs. Most software developers, such as Microsoft, Amazon, and Google, supplied an open source cloud environment. The scheduling algorithm workflow takes a unique approach with numerous outcomes based on the most recent approaches. The chance of scattering data analytics in massive data storage in a cloud computing environment is determined in this study using the self-adaptive group formation approach, which is carried out by data analysis mapping using the provided data sets as input. There are four sorts of approaches: classical approaches, relative approaches, subjective approaches, and conditional approaches. The input data sets are translated into methods, then comparable properties are discovered, and the likelihood of that event occurring is confirmed. In this study, the Map Reduce procedure was designed to make load difference along with the better performance strategy for cluster usage by managing probability distribution. [23]

Sentiment Analysis Techniques in Web Opinion Mining: A Survey

Sentiment analysis and opinion mining are subfields of machine learning, as proposed by S. Veeramani et al. in this study. They are highly significant in the contemporary context since there are so many user-opinionated texts available on the internet. This is a difficult topic to address since natural language is very unstructured. A machine's assessment of the meaning of a certain statement is tedious. However, the utility of sentiment analysis is growing by the day. Machines' capacity to comprehend and understand human emotions and sentiments must be made dependable and efficient. Sentiment analysis and opinion mining are two methods for doing so. Manual training can address the sentiment analysis problem to a decent degree. However, no completely automated system for sentiment analysis that does not require operator involvement has been established. This is mostly due to the difficulties in this subject. The purpose of this work is to conduct a literature review on the topic of sentiment analysis and opinion mining. Many pertinent research have developed in this subject, and this study provides a glimpse into a few of them. Opinion Mining (OM), described as a blend of information retrieval and computational linguistic approaches, is a promising science that deals with the views conveyed in a document. [24]

Scrum Investigation Analysis for an Android App

In this study, S. Karuppasamy et al. suggest Agile as one of the software development methodologies employed in the contemporary information technology context. This technique is primarily concerned with how industrial workers' costs and time are successfully utilised. Agile software development is an iterative and incremental (evolutionary) approach that is performed in a highly collaborative manner by self-organizing teams with just enough ceremony to produce high quality software in a cost effective and timely manner that meets the changing needs of its stakeholders. The Scrum methodology has been designed to manage the system development process. It is an empirical method to software development that applies the notions of industrial process control theory, resulting in an approach that reintroduces the concepts of flexibility, adaptability, and productivity. Scrum focuses on how team members should interact in order to develop a flexible system in a continuously changing environment. [25]

3. COMPARATIVE ANALYSIS

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

Title	Techniques & Mechanisms	Parameter Analysis	Future Work
A Data Analytics Suite for Exploratory Predictive, and Visual Analysis of Type 2 Diabetes	An analytics suite that performs exploratory, predictive, and visual analysis of T2D data. Three types of analytics workflows were presented that .	analysis of T2D database to build a predictive model that can assess risk of patients to T2D related complications	future work include building and training the model on larger databases to increase the prediction accuracy and develop more robust prediction models by adopting artificial intelligence methods, and clinical validation of the data analytics
A Data Sharing Protocol to Minimize Security and Privacy Risks of Cloud Storage in Big Data Era	We propose a novel group key management protocol for the data sharing in the cloud storage. In SSGK, we uses RSA and verified secret sharing to make the data owner achieve fine-grained control over the outsourced data without relying on any third party.	Security mechanism in our scheme guarantees the privacy of grids data in cloud storage. Encryption secures the transmission on the public channel; verified security scheme make the grids data only accessed by authorized parties.	forward and backward security in group key management may require some additions to our protocol. An efficient dynamic mechanism of group members remains as future work.
A Generic Data Analytics System for Manufacturing Production	A generic and low-user-threshold manufacturing data analytics system, GMDA, is proposed in this paper. This will enable small and medium manufacturers to conduct data analysis tasks using their own data and to benefit from it, even if they have no knowledge or experience of data analytics.	A knowledge base was established so that our system could select, based on the KNN algorithm, the most appropriate algorithm for the data.	In the future, we plan to replace the R part in GMDA with Hadoop or Spark to make it available for use with big data.
A Methodology of Real-Time Data Fusion for Localized Big Data Analytics	This data can be semantically rich data, relational data, hierarchical data, or another form of data. Therefore, data found in the shape of RDF, RDB or XML needs to be capable of transforming in any direction.	Our study focus is on data fusion of heterogeneous data into RDF or JSON. As a special case, we have focused on RDB and RDF based on data transformation.	One feature of big data is to work with a variety of data, which can be in any form coming into or going out of the system

A Novel Spatiotemporal Data Model for River Water Quality Visualization and Analysis	Then, methods of expanding a point to a line segment, a flat surface and a cube are designed respectively to make the proposed data model available for common generalizations of river spaces.	In those regions where water quality varies greatly, the intervals among spatial points need to be as small as possible so that detailed water quality information can be represented	In our future research, we will pay more attention to the self-adaption adjustment of intervals among spatial points so that both details representation of RWQ and memory space reduction of long-term monitoring data can be realized.
Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining	Considering there is a certain degree of irrationality and subjectivity in the results of the school's evaluation, by using the K-means algorithm in unsupervised learning to perform clustering analysis on student performance	Although the clustering results are obtained after comprehensive consideration of the actual situation and the use of quantitative analysis, the selection of the initial clustering center is determined randomly, which may have a certain impact on the accuracy of the clustering results.	In the future, it can be further enhanced by combining association models or some integration-based technologies. In addition, EDM can also be extended to medical data processing, sports data processing and other fields.
Analyzing Objective and Subjective Data in Social Sciences: Implications for Smart Cities	The aim of this work was to present how data science and machine learning techniques can be used in social science studies in order to maximize the insight gained. In order to do this we made use of a pilot study in which the problem at hand consists of understanding the interaction of citizens with green spaces.	The data can be split into two main categories: subjective and objective. This allows for multiple levels of analysis and comparison. Problems that occur are incomplete data, lack of data or erroneous data which can impact on statistical significance.	In the future the app may actively stimulate the improvement of well-being based on known causes of well-being variation; work in this direction is only preliminary at the moment.
Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data	Optimal parameters for the Prophet and the LSTM models are also determined. Additional results explained earlier will provide new insights into crime trends and will assist both police departments and law enforcement	By exploring the Prophet model, a neural network model, and the deep learning algorithm LSTM, we found that both the Prophet model and the LSTM algorithm perform better than conventional neural network models.	In future, we plan to complete our on-going platform for generic big data analytics which will be capable of processing various types of data for a wide range of applications.

	agencies in their decision making.		
Big Data in Motion: A Vehicle-Assisted Urban Computing Framework for Smart Cities	Simulation results show that this approach is efficient in terms of energy savings as well as the utilization of resources. The framework also facilitates the researchers to incorporate different algorithms to optimize the data transfer mechanism.	We have proposed a framework that utilizes vehicular networks for big data transfer between data centers. The framework supports urban/social computing through participation of volunteer vehicles and helps in reducing the carbon footprint associated with movement of Exabyte scale of data.	In future, we aim to extend our framework by incorporating Artificial Intelligence and machine learning techniques to select the suitable vehicle for data transfer; thus, this can improve the task delivery ratio further.
Big Data-Based Improved Data Acquisition and Storage System for Designing Industrial Data Platform	We compare the optimal serialization methods provided by big data framework with other high performance methods to optimize the existing framework's method.	Compared with the other serialization methods integrated by Hadoop and Spark, Protobuf performs better and better in data processing as the number of serialized objects increases.	In future , there are a number of studies on data acquisition and optimization of computing framework methods, and most of these methods provided good performance in the field of Internet.
City Pulse: Large Scale Data Analytics Framework for Smart Cities	The main contributions of this work include integrating of heterogeneous data streams, providing interoperability, quality analysis, (near-) real-time data analytics and application development in a scalable framework.	Proposes a framework for large-scale data analytics to provide information in (near-) real-time, transform raw data into actionable information, and to enable creating ''up-to-date'' smart city applications.	The future work will focus on evaluation of the proposed framework for (near-) real-time city data analytics in different domains. The framework will be also used to provide data access user interfaces and prototype applications for smart city use-cases in the city of Aarhus and the city of Brasov.
Connected Sensors, Innovative Sensor Deployment, and Intelligent Data Analysis for Online Water Quality Monitoring	The WQM sector will hugely benefit from the sensor networks and techniques that being developed for IoT.	The connected sensor technologies for WQM could provide the bridging solution for current disconnect between data quality, data gathering, and data analysis and enhance	The use of sensor networks and Internet communications combined with GIS tools will be having an important role in the future and can be very beneficial to stakeholders in not only

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

		the global data interoperability	efficiently managing the WQ but also in water distribution management, agriculture, and landscaping sectors, where it can reduce water consumption and wastage
Data Science as Political Action: Grounding Data Science in a Politics of Justice	The field of data science must abandon its selfconception of being neutral to recognize how, despite not being engaged in what is typically seen as political activity	As a form of political action, data science can no longer be separated from broader analyses of social structures, public policies, and social movements.	Toward this end, one necessary direction for future research is to develop interdisciplinary frameworks that will help data scientists consider the downstream impacts of their interventions.
Distributed Data Strategies to Support Large-Scale Data Analysis Across Geo-Distributed Data Centers	The main advantage of this strategy is to separate the storage level from the analysis level. In the second strategy, we consider data replication among different data centers.	We store the data on each data center as a set of RSP data blocks. In the first strategy, some data blocks are required to download from the remote data centers to a central data center for approximate analysis of the big data as a whole.	. In future work, we will explore streaming data and scheduling data replication among multiple data centers.
Educational Data Mining to Support Programming Learning Using Problem-Solving Data	e. Furthermore, the proposed framework can be applied to other practical/exercise courses to demonstrate data patterns, statistical features, and rules.	we proposed an EDM framework for data clustering, patterns, and rules mining using real-world problemsolving data.	In the future, the experimental results of EDM using problem-solving data can be integrated to visualize different LA for programming platforms such as the OJ system.
Exploratory Analysis for Big Social Data Using Deep Network	Hence, our proposed paradigm can be applied to a wide range of scenarios and we can achieve the goal to let the data speak for itself. Further, we can take advantage of the rapid progress in deep learning research and facilitate the data-driven social science research to associate with novel	There are several reasons that our proposed paradigm is generic and can be applied to a wide range of social science problems.	However, there are still some problems left for us to resolve and interesting future works to do, such as providing an explanation of why a specific network capacity is suitable for a particular dataset, changing model structures

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

	deep learning algorithms.		
IDP: An Intelligent Data Prediction Scheme Based on Big Data and Smart Service for Soil Heavy Metal Content Prediction	big data prediction methods are used as research objects to predict the content of metal in soil. IDP is proposed to fit and predict data through the collaboration of MBPSO and LSSVM for smart service.	Through the prediction of the heavy metal content, the errors are compared to judge the performance of model learning, generalization and prediction.	In future work, for the model proposed in this article, we can consider continuing to improve the PSO to improve the optimization speed and accuracy of the model.
Processes Meet Big Data: Connecting Data Science with Process Science	However, as highlighted in this editorial, several research and technology challenges remain to be solved before process mining, data science and Big data technologies can seamlessly work together.	In this extended editorial paper, we have discussed the relation between process and data science, identified some of the remaining difficulties and outlined a research strategy that we believe should underlie the community's efforts toward full integration of Data and Process science.	In the future, we hope to see tools for defining Map and Reduce functions on the basis of the business process model data types, of the relations among them and of other semantics-rich context information
Toward Effective Response to Natural Disasters: A Data Science Approach	An integrated framework that, based on data science, can help decision makers to face natural disasters. As first realization, we embed automatic support to evacuation and reconstruction planning.	the definition of the GisToGraph algorithm to generate an enriched underlining network of any location, specifically tailored to include useful information for disaster management, especially in the preparedness, response and reconstruction phases	Other aspects we are willing to explore in the future are further optimization models, exact or approximate, to be employed in order to reduce the computational effort presently required by simulations.

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

CONCLUSION

Given the degree of irrationality and subjectivity in the school evaluation outcomes, utilizing the K-means method in unsupervised learning to perform The article begins with data mining by doing a clustering analysis on student performance and then utilizing the clustering findings as the category label of CNN. It is finally discovered that the model has a higher optimal forecast accuracy, which is important to ensuring objective and fair student evaluation by school. Furthermore, it is accessible to quickly recall students who are on academic probation. When examining data labels, the label value selection range must be considered, and the label value selection range is connected to the clustering number. The K-means method has a well-known flaw: the value of k is selected arbitrarily. To enhance the method, the study employs an objective statistic to maximize k-value selection and substitutes subjective evaluation with quantitative analysis, resulting in more strong clustering findings. The persuasiveness also makes CNN training and prediction outcomes more dependable, and the model's success is automatically assured. Although the clustering results are achieved after careful evaluation of the real scenario and the use of quantitative analysis, the initial clustering centre is chosen at random, which may have an influence on the accuracy of the clustering findings. Although the suggested statistic improves CNN results over those obtained without it, we do not compare it to other classifiers.

REFERENCES

[1] B. Liu, Y. Li, S. Ghosh, Z. Sun, K. Ng, and J. Hu, ‘‘Complication risk profiling in diabetes care: A Bayesian multi-task and feature relationship learning approach,’’ IEEE Trans. Knowl. Data Eng., vol. 32, no. 7, pp. 1276–1289, Jul. 2020.

[2] Q. Xu, C. Tan, Z. Fan, W. Zhu, Y. Xiao, and F. Cheng, ‘‘Secure multiauthority data access control scheme in cloud storage system based on attribute-based signcryption,’’ IEEE Access, vol. 6, pp. 34051–34074, 2018.

[3] "An Assembling Creation Conventional Information Examination Framework," S. Kausar, X. Huahu, I. Hussain, W. Zhu, and M. Zahid, IEEE Access, vol. 6, pp. 72724-72734, 2018.

[4] D. Buenao-Fernandez, W. Villegas, and S. Luján-Mora, "A REAL-TIME DATA FUSION METHODOLOGY FOR LOCALIZED BIG DATA ANALYTICS," Comput., May 2019. Appl. Eng. Educ., vol. 27, no. 3, pp. 744-758.

[5] W. H. M. Wan Mohtar, K. N. A. Maulud, N. S. Muhammad, S. Sharil, and Z. M. Yaseen, ‘‘Spatial and temporal risk quotient based river assessment for water resources management,’’ Environ. Pollut., vol. 248, pp. 133–144, May 2019

[6] B. A. Javier, F. B. Claire, and S. Isaac, "analysis and prediction of academic performance of students using educational data mining," wires Data Mining Knowledge Discovery, vol. 10, no. 1, January-Feb 2020, Art. No. E1287.

[7] L. Gao, Z. Zhao, C. Li, J. Zhao, and Q. Zeng, ‘‘Deep cognitive diagnosis model for predicting students’ performance,’’ Future Gener. Comput. Syst., vol. 126, pp. 252–262, Jan. 2022.

[8] C. Romero and S. Ventura, "BIG data analytics and mining for crime data visualization and trend forecasting," wires Data Mining Knowledge Discovery, vol. 10, no. Art. 1, May 2020 no.

[9] "big data in motion:" e1355 S. Wang, J. Intell., "a framework for vehicle-assisted urban computing in smart cities," Syst., Fuzzy, vol. 37, no. 1, 2019, pages 9-16.

[10] M. Liu, J. Yang, T. Song, J. Hu, and G. Gui, ‘‘Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks,’’ IEEE Trans. Veh. Technol., vol. 69, no. 1, pp. 641–653, Jan. 2019.

[11] B. Cheng, S. Longo, F. Cirillo, M. Bauer, and E. Kovacs, ‘‘Building a big data platform for smart cities: Experience and lessons from Santander,’’ in Proc. IEEE Int. Congr. Big Data, New York, NY, USA, Jun./Jul. 2015, pp. 592–599

[12] "Massive LMS log data analysis for the early prediction of course-agnostic student performance," Computer, R. G. Moises, D. P. P. R. Maria, and O. Francisco. Educ., vol. Art. 163, April 2020 no. 104108.

DOI:10.46647/ijetms.2023.v07i02.068 ISSN: 2581-4621

[13] International Journal of Management Education, vol. 13, J. N. Walsh and A. Rsquez, "Using cluster analysis to investigate the involvement of native and non-native Englishspeaking management students in a flipped classroom," 18, no. 2, article, July 2020 no. 100381.
[14] "Towards creating hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation," Soft Comput., V. G. Karthikeyan, P. Thangaraj, and S. Karthik vol. 24, no. 24, Dec. 2020, pages 18477-18487.
[15] "Unsupervised learning-based mining of academic data sets for student performance analysis," IEEE 14th International Conference, L. M. Crivei, G. Czibula, G. Ciubotariu, and M. Dindelegan Symp. Appl. Comput. Intell. Details. (May 2020, Timisoara, Romania (SACI), pages 11-16.
[16] "Analysis of students' behavior in online learning situations using user clustering and self-organizing maps neural networks," by S. Delgado, F. Moran, J. C. S. Jose, and D. Burgos, IEEE Access, vol. 9, pp. 132592-132608, 2021.
[17] A.-A. Liu, Y.-T. Su, W.-Z. Nie, and M. S. Kankanhalli, ‘‘Hierarchical clustering multi-task learning for joint human action grouping and recognition,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 1, pp. 102–114, Jan. 2017
[18] "Multi-tier student performance assessment model (MTSPEM) with integrated classification approaches for educational decision making," International Organization for Standardization Comput. J. Intell. Syst., vol. 14, no. 1, pp. 1796-1808, Mar. 2021.
[19] S. Hernandez, S. Zelst, J. Ezpeleta, and W. v.d. Aalst, “Handling big(ger) logs: Connecting ProM 6 to apache hadoop,” in Proc. BPM2015DemoSession, 2015, pp. 80–84.
[20] E. Howard, L. Pasquini, C. Arbib, A. Di Marco, and E. Clementini, ‘‘Definition of an enriched GIS network for evacuation planning,’’ in Proc. 7th Int. Conf. Geographical Inf. Syst. Theory, Appl. Manage., 2021, pp. 241–252.
[21] Jagadeesan S, Mani C, Kumar R.N, and Prabhakaran S, "High level secure messaging based on steganography and cryptography (Article)," SSRG International Journal of Engineering Trends and Technology, Vol. 68,No: 2, PP: 142-145, 2020.
[22] S.Karuppusamy and G.Singaravel, "Investigation Analysis for Software Fault Prediction Using Error Probabilities and Integral Methods," International Journal of Applied Mathematics and Information Sciences, Vol.13, Special Issue:1, pp. 205-210, 2019.
[23] S. Prabhu, N. Sengottaiyan, and B. G. Geetha, "Self-Adaptive Approaches to Probability Distribution of Data Analytics in Cloud Computing Resource Services for Infrastructure Hybrids Models," International Journal of Applied Mathematics and Information Sciences, Vol.13, Special Issue:1, pp.437-446, 2019.
[24] S Veeramani and S Karuppusamy, "A survey on sentiment analysis approach in web opinion mining," International Journal of Science and Research, Vol. 3, No.8, pp.1776-1780,2012.
[25] S Karuppasamy, G Singaravel, and P Kaveen, "Scrum Investigation Analysis for Android Application," Bonfring International Journal of Networking Technologies and Applications, Vol.5,No.1.,2018,pp12-16.