International Journal of Innovative Analyses and Emerging Technology

| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

ANALYSIS OF SENTIMENTAL BIAS THE IMPLEMENTATION OF SUPERVISED MACHINE LEARNING ALGORITHMS

S. Suman Rajest
Professor, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India

R. Regin
Assistant Professor, Department of Computer Science and Engineering,
SRM Institute of Science and Technology, Ramapuram, India

Shynu T
Master of Engineering, Department of Biomedical Engineering,
Agni College of Technology, Chennai, Tamil Nadu, India

Steffi. R
Assistant Professor, Department of Electronics and Communication,
Vins Christian College of Engineering, Tamil Nadu, India

Abstract: *More and more people are writing reviews of items and services online as a result of the explosion of internet shopping. Text mining is a method for discovering useful patterns in massive datasets. In order to construct novel realities or ideas to be explored further by more conventional experimental methods, a crucial component is used to interface the extracted data. Sentiment analysis presents several obstacles. When people use a computer browser to go online and purchase goods or services, they are engaging in online shopping, a type of electronic commerce. Those looking to make a purchase in the near future can benefit much from reading evaluations of products on the internet. As a result, many opinion mining strategies have been put forward, with one of their main obstacles being the assessment of the direction of a review phrase, whether positive or negative. When it comes to overcoming issues with sentiment*

Keywords: *Amazon Web Server, Machine Learning, Support Vector Machine, Data Frame, Numpy, Random Forest.*

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

classification, machine learning has recently shown to be a useful method. There is no need for human intervention when training a machine learning model; the programme will automatically learn a functional representation. Our proposed supervised machine learning approach, on the other hand, uses widely known ratings as weak supervision signals to classify the sentiment of product reviews. We build a dataset with 15,000 labelled review sentences and 200,000 weakly labelled review sentences from Amazon to test the suggested approach. Superior precision as measured experimentally as contrasted with the prior iteration.

Introduction

Using machine learning methods, this project primarily aims to conduct sentiment analysis and product reviews. People are becoming accustomed to shopping online and sharing feedback about their purchases on review and merchant websites due to the proliferation of e-commerce [9]. Future customers can use these opinionated contents to inform their decision-making, and merchants can use them to improve their products and services [10]. The proliferation of evaluations, however, has led to a serious information overload problem for the general public. Many opinion-mining methods, such as opinion summarization, opinion polling, and comparison analysis, have been suggested to address this issue [11-15]. Verifying the intended tone of review sentences is a major obstacle. The two most common approaches to sentiment analysis use lexicons and machine learning, respectively [16]. A common approach taken by lexicon-based approaches is to build a sentiment lexicon containing opinion words (such as "amazing" and "disgusting") and then use this vocabulary and prior syntactic knowledge to create categorization rules. It takes a lot of work to build a lexicon and establish rules using this manner, but it works [17-21].

In addition, lexicon-based approaches aren't designed to deal with subjective claims like "we purchased the bed a week ago, and a valley emerged today," which contain implicit opinions [22-26]. It has already been mentioned that this is a crucial viewpoint. In most cases, objective facts are preferable to personal opinions. When dealing with implicit opinions, approaches based on a vocabulary can only do so on an as-needed basis [27-31]. The initial work on sentiment categorization using machine learning tackled the challenge by utilising well-known machine learning methods like Naive Bayes [32-37]. Following that, the majority of studies conducted in this area focused on improving classification performance through feature engineering. There have been investigations into several aspects, such as n-grams, POS information, syntactic linkages, and beyond. In addition to the high human effort required for feature engineering, it is possible that feature sets that work well in one domain will not do so well in another [38-41].

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

One way to find out how someone feels about a certain subject is to utilise sentiment analysis. People use mining to extract useful information from large datasets [42]. The computational (or programmed) inquiry of people's emotions conveyed in written dialect or content is known as assumption research or presumption mining. Examine the evaluation process, which has been limited to the article section up until now [43]. First, news stories without a clear goal; second, the difficulty of differentiating between good and bad news; and third, the seeming necessity and multi-faceted nature of, relying on space-specific explanations and background information, have all contributed to this [44]. In order to find and extract subjective information from sources, sentiment analysis (also known as opinion mining) employs computational linguistics, text analysis, and natural language processing. From advertising to support, sentiment analysis finds a home in social media and review platforms [45]. An individual's or a group's emotional connection to a subject or the document's overall contextual polarity can be inferred through sentiment analysis. An author's attitude can be their judgement or evaluation (refer to appraisal theory), their affective state (what they're feeling when they write), or the emotion they want to convey in their work (that is to say, the emotional effect the author wishes to have on the reader) [46-51].

Emotional intelligence is the study of reading and understanding human thoughts, feelings, assessments, attitudes, and judgments in written form. Because opinions are fundamental to nearly all human endeavours, sentiment analysis tools find use in nearly every industry [52]. They have a significant impact on how we act. Sentiment analysis is a method for discovering and extracting relevant information from texts by utilising natural language processing and text analysis. Interest in sentiment analysis has skyrocketed, thanks to the proliferation of prominent social media platforms like blogs and social networking sites like Facebook, Twitter, etc. The standard definition of this activity is to assign a value objective or subjective to a piece of text, typically a sentence [53-57]. There are cases where this becomes more of a challenge than polarity categorization. A news item quoting people's thoughts is an example of an accurate document that contains subjective sentences; the subjectivity of words and phrases might depend on their context [58-59]. In addition, the results are highly sensitive to the subjective criteria used to the texts being annotated. On the other hand, it demonstrated that improving performance was possible by excluding objective statements from documents prior to polarity classification [60].

The field of online review mining has seen a lot of activity in the past. As a paradigm change for the future generation, several research groups are investigating text mining and sentiment analysis in various methods. The online movie reviews were categorised as positive or negative using machine learning methods [61-64]. The majority of algorithms that employ sentiment analysis to determine how people feel about a product or service rely on oversimplified language. It is not easy to decipher a passage of text into a straightforward pro or negative attitude due to cultural influences, grammatical subtleties, and different situations [65-71].

With the proliferation of online reviews, it is now impossible to ascertain a general consensus on a product offered for sale on the internet [72-79]. Researchers have employed a variety of methods to tackle this issue, such as analysing the reviews' syntax and appearance and searching for sentiments conveyed in the articles. Opinion mining relies on aspect-based evaluation, and study interest in product aspect extraction is growing; yet, larger data sets necessitate more complicated algorithms to solve this problem accurately [80-85]. This presents a way to aggregate numerous product

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

reviews in a given domain into a single document by extracting and summarising relevant product attributes and opinions [86-89]. By drawing on expertise in product aspect extraction and offering the right amount of detail and extensive representation capabilities, this ensures that the review summaries are as accurate and helpful as possible. For both cameras and laptops, the results demonstrate that the suggested solution achieves F1-scores of 0.714 and 0.774, respectively [90].

The majority of whole-opinion reviews are categorised using document-level opinion mining. Finding subjective sentences is the main emphasis of sentence-level sentiment analysis. As a result, our work tackles the problems associated with reviewing products based on their features [91-93]. In this study, we delve into the topic of extracting product attributes from review data. We need to define our system's language before we can move into the intricacies of the work [94-99]. Using widely available ratings as weak supervision signals, we present a supervised machine learning framework for sentiment classification in product reviews using a linear regression approach. We build a dataset with 15,000 labelled review sentences and 200,000 weakly labelled review sentences from Amazon to test the suggested approach. Superior precision as measured experimentally as contrasted with the prior iteration.

Literature Survey

Mouthami et al., [1] analysis is the method by which one ascertains the sentiment, view, or attitude conveyed by an individual towards a specific subject. People use mining to extract useful information from large datasets. The computational (or programmed) inquiry of people's emotions conveyed in written dialect or content is known as assumption research or presumption mining. Examine the evaluation process, which has been limited to the article section up until now. Most of the blame for this goes to three sources: 1) news articles that fail to specify their intended audience; 2) the difficulty of differentiating between good and bad news; and 3) the seeming necessity and multi-faceted nature of, relying on space-specific explanations and background information. The purpose of an estimation examination is to ascertain the man's mental state, hypothesis, or emotion as it relates to a particular topic. Assessment mining, also known as concept examination, is a method of extracting subjective data from sources by means of regular dialect preparation and message research. A surge in interest in hypothesis testing has accompanied the meteoric rise of online networking platforms like websites and informal communities. It is common practise for agents to view audits, evaluations, proposals, and various forms of online speculation in order to identify new opportunities and manage notorieties. Estimation analysis and hypothesis mining are useful for both individuals and organisations. To enhance other frameworks, such as recommendation, data extraction, and inquiry noting frameworks, assessment examination can be applied.

There ought to be three tiers for opinion examination: archive (document), sentence, and aspect. Record level fee ling analysis is categorising the overall judgments expressed by the author throughout all archived content as positive, negative, or neutral. To determine if a sentence is objective or subjective, researchers use the sentence-level slant inquiry. Emotional statements are then determined to be neutral, damaging, or certain. Using a set of predetermined viewpoints and a system of literary audits, a framework for perspective-based sentiment surveying compiles data. The assessment survey is delivered by differentiating the extreme of each angle from each survey.

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

An investigation at the viewpoint level finds more nuanced results. In conclusion, the goal of any good survey is to elicit honest feedback from target audiences about products, services, brands, movies, news, events, problems, and personal qualities. The most promising area of Sentiment analysis is archive-level categorization. Substance, sentence, and record levels are common examples of levels where the notion evaluation is typically conducted on a single level. Components assumption testing follows the fabrication of language at the substance level, which is followed by the separation of critical components by the recognition of earlier and relevant extremity.

Kalamboukis and Kyriakopoulou [2], Among the earliest uses of machine learning, text classification addressed the overarching challenge of supervised inductive learning: how to automatically organise new documents based on a training set of documents that have already been categorised into one or more preset categories. Some of the many uses for automated text classification include indexing, content management, filtering, routing, word meaning disambiguation, and classification of search spaces in the style of Yahoo!. Support Vector Machines, Nearest-Neighbors, Regression, Neural Networks, Naive Bayes, Decision Trees, and Neural Networks are just a few of the many text classification algorithms that have been developed. In order for classification algorithms to effectively generalise to new documents, they typically need a large amount of training data. Nevertheless, it is extremely time-consuming and expensive to generalise using labelled instances. Classifiers with low sample size learning requirements are critical. Combining supervised and semi-supervised or unsupervised learning methods, such as active learning, transductive support vector machines (SVMs), and co-training, has been the subject of numerous experiments aimed at improving the performance of traditional classifiers.

If the Classifier could know the distribution of the test instances before classifying them, that would be great. The research presented in this study is inspired by this comment. This paper's overarching goal is to address the challenge of learning from training sets of varying sizes by making use of meta-information contained in the data that is derived from clustering the entire dataset, including both training and testing samples. An alternate method to term selection for dimensionality reduction or a way to improve the training set, clustering has been employed in the research on text classification. Clustering is employed in the second scenario to unearth a "structure" within the training data and augment the feature vectors with additional properties retrieved from clusters. In addition, it is employed to supplement a limited set of instances, both labelled and unlabeled, by inferring labels from clustering results on both sets of data and applying them to the unlabeled data. In these domains, a number of clustering methods have been suggested. Clustering is utilised on both the training and testing sets as an additional step to text categorization in order to achieve this goal. This method circumvents the inability of an inductive learner to estimate the full dataset's structure and the position of the testing cases. We anticipate that classifier performance will be improved by including clustering knowledge into the simple BOW text representation. The suggested method is shown effective through experiments run on datasets and tasks supplied within the context of the ECDL/PKDD 2006 Challenge Discovery on tailored spam filtering. Experiments reveal significant gains in classification accuracy, particularly when working with tiny training sets. Applying two-dimensional clustering to text classification is a more involved method than utilising simple words for document representation, and classification is done using word clusters instead.

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

The work of Abbasi et al., [3]. Many people spread propaganda and share information and opinions online. In order to categorise opinions expressed in online forums across several languages, this research suggests sentiment analysis approaches. In order to classify the emotional tone of Arabic and English material, we assess the usefulness of stylistic and syntactic elements. The unique linguistic features of Arabic are taken into consideration by including targeted feature extraction components. A hybridised genetic algorithm called the entropy-weighted genetic algorithm (EWGA) is also developed; it uses the information-gain heuristic for feature selection. The goal of developing EWGA was to make it easier to evaluate important characteristics and boost performance. A benchmark dataset of movie reviews and online forum posts from the United States and the Middle East are used to assess the suggested features and methods. With accuracies of over 91% on the benchmark dataset and the US and Middle Eastern forums, the experimental results utilising EWGA with SVM show high-performance levels. Indicating the usefulness of these features and approaches for document-level sentiment classification, EWGA performed better than other feature selection algorithms, and stylistic features improved performance across all testbeds.

The relevance of web content analysis is growing as a result of the increased use of computer-mediated communication (CMC) and online resources such chat rooms, websites, forums, and email. While the Internet and CMC have many positive uses, they have also made some negative aspects, like as cybercrime, more accessible. The Internet has grown into a popular platform for extreme and hate organisations to communicate, in addition to being used for deceit, identity theft, and the spread of pirated software. In this research, we suggest analysing hate speech and extremist group forum posts using sentiment analysis. The sentiment categorization on a benchmark dataset of movie reviews and two online communities—one for white supremacists in the United States and one for extremists in the Middle East—form the basis of our investigation. Syntactic and stylistic feature sets are considered in our evaluations. In addition, we create the EWGA, which is a feature selection method that takes entropy into account. A sentiment analysis strategy is created by the features and techniques to classify the sentiments of Web speech in many languages. The results show that this approach is effective for categorising and assessing sentiments in extremist forums, thanks to the use of support vector machines (SVM). Analyzing direction-based text, which includes views and feelings, is the main focus of sentiment analysis. Here, we zero in on sentiment classification studies, which have as their goal the determination of the subjective or objective nature of a text, or the presence or absence of good or negative feelings in an emotional text. Numerous tasks, features, methodologies, and domains of application are significant components of sentiment categorization.

Finding the writers' ideas about certain entities is the goal of Feldman's [4] analysis, often known as opinion mining. Both influential people and regular people shape people's opinions, which in turn influences their decision-making. People usually look for other people's thoughts and reviews of products before buying them online. One of the most active subfields of computer science right now is sentiment analysis. There are more than seven thousand articles covering the subject. Numerous new businesses are focusing on sentiment analysis, and popular statistical programmes like SPSS and SAS have sentiment analysis components. A deluge of "sentiments" posted on various online platforms such as Facebook, Twitter, message boards, blogs, and user forums. Anyone looking to keep tabs on their reputations and acquire quick comments about their goods

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

and activities can find these text snippets to be a treasure trove of information. With the use of sentiment analysis, these companies can keep tabs on social media platforms in real-time and respond appropriately. Stakeholders in marketing, public relations, campaigns, legislators, equity investors, and online consumers all stand to gain from sentiment research tools. In terms of subjectivity, it is anticipated to divide sentences into two main categories: those conveying factual information (objective sentences) and those containing explicit ideas, attitudes, and perspectives (subjective sentences). Sentences conveying emotion are the primary area of study here. When defining a sentiment application for stock selecting, however, we utilise objective sentences.

Here we have a Manhattan hotel review to illustrate the point. Clean, roomy, and furnished to a high standard, the king suite was ideal. Hotel porters, housekeeping, and reception all went out of their way to accommodate guests. Every time I asked the maid for more, she brought it to me. Considering the unpredictable weather, it was a relief that the HVAC system worked properly. Each room and the entire building is soundproof. In terms of convenience to the metro, restaurants, and shops, the area is second to none. The sole "complaint" concerns the availability of fast Internet. Only on floors 8–12 can you find it. In general, the assessment paints a rather positive picture of the hotel. Heating, air conditioning, helpful staff, comfortable mattresses, safe area, and Internet access are all part of the hotel's amenities. Systems for sentiment analysis need to be able to do more than just read reviews; they need to be able to assess the overall review's sentiment and break it down each hotel amenity. First, we go over some of the most pressing issues in the field of sentiment analysis, including the methods now under development for addressing them. Then, we take a look at some of the most promising current applications of sentiment analysis. Finally, we touch on a few of the unanswered questions that remain in this area of study. We will direct the reader to other substantial works on the subject since we do not have the space to address all of the topics and methods here. With both labelled and unlabeled data records available, the classification model is constructed using a semi-supervised learning technique. Regardless of the labelling or lack thereof in the input, the classification that specifies the class to which the dataset belongs would be the output.

In order to take advantage of the physical layer multi-rate capabilities provided by Zhu & Cao [5], methods at the media access control (MAC) layer are required. To do this, a number of options have been put forward. On the other hand, these solutions solely take into account the use of high-quality channels for the direct connection between the transmitter and the receiver. Due to IEEE 802.11's ability to adapt transmission rates to changing channel circumstances, data packets may be transferred more quickly via a relay node rather than the direct link in cases where the latter has poor quality and slow speed. This work presents rDCF, a new protocol for distributed coordination functions (DCFs) supported by relays at the media access control (MAC) layer, which aims to further utilise the multi-rate capacity of the physical layer. To facilitate agreement among the sender, the relay node, and the receiver regarding the data rate and the use of the relay node, we develop a protocol. We suggest methods to enhance rDCF's performance even more by taking a number of factors into account, including bandwidth consumption and channel faults. According to the simulation results, rDCF may greatly enhance the system's performance even when the direct link's channel quality is inadequate. To make better use of IEEE 802.11's multi-rate capability, we provide a new DCF-based MAC protocol, relay-enabled DCF (rDCF). Intelligently applying multi-hop data transmission (mostly two-

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

hop in this research) based on the channel state among mobile nodes allows rDCF to attain a greater transmission rate. In particular, a two-hop high-speed transmission via a relay node can expedite the delivery of an incoming packet from the sender to the receiver when the direct link between the two devices can only support a low transmission rate but there is a relay node such that both the links from the sender to the relay node and the relay node to the receiver can support high transmission rates.

Each mobile node in a rDCF network can detect the state of the channels between its nearby peers. The node periodically advertises its relay information in order to become a relay node for its neighbours, taking into account the channel conditions that have been acquired. In order for the sender, relay node, and receiver to swiftly agree on whether to execute the relay and which rate to utilise based on the current channel condition, a triangle handshake is formed when the sender transmits the packet to the receiver if it detects a relay node. Improving the rDCF protocol is something we suggest doing to address problems like bandwidth usage and time-varying channel conditions. The simulation findings demonstrate that rDCF can considerably decrease packet latency, increase system throughput, and mitigate the effect of channel defects on fairness, as we assess the rDCF protocol in several scenarios.

Requirement Specifications

The term "artificial intelligence" (AI) describes computer systems that are able to mimic human intelligence. Among these steps are learning (the act of taking in new knowledge and internalising its principles), reasoning (the process of using those rules to arrive at more or less certain conclusions), and self-correction. A few examples of AI's specific uses are computer vision, voice recognition, and expert systems. American computer scientist John McCarthy first used the term artificial intelligence in 1956 during the birth of the field at the Dartmouth Conference. In modern usage, it has come to mean both robotic process automation and robots themselves. Big data, the explosion in the volume, velocity, and variety of data collected by organisations, has contributed to its recent rise to prominence. Artificial intelligence (AI) can outperform humans at data pattern recognition, which means businesses can learn more from their data. The goal of this subfield of computer science is to design computers with human-level intelligence [100-107]. In recent years, it has grown into an integral component of the IT sector. Artificial intelligence research is characterised by a high level of technical specificity. Machine learning and computer programming are at the heart of AI's fundamental challenges.

Result

Studying artificial intelligence revolves around knowledge engineering. If machines have access to vast amounts of data about the world, they can typically mimic human behaviour and reactions [108]. For AI to carry out knowledge engineering, it needs access to things like categories, attributes, objects, and relations. It is difficult and time-consuming to teach computers to reason and solve problems. Another fundamental component of AI is machine learning. While learning with proper supervision entails numerical regressions and classification, learning without supervision necessitates pattern recognition in input streams. Classification is concerned with assigning objects to specific categories, while regression is concerned with collecting numerical input or output instances in order to find functions that allow for the production of appropriate

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

outputs from their respective inputs. Computational learning theory is a well-established subfield of theoretical computer science that focuses on mathematical study of machine learning algorithms and their performance. The capacity of machines to perceive their environment by interpreting sensory data is the focus of machine perception research. Computer vision can also handle some sub-problems involving visual inputs, like gesture, object, and face recognition. Another important area that is connected to AI is robotics. Robots need to be smart enough to manage objects, navigate, and solve subproblems like localization, motion planning, and mapping. These tasks are relevant to data science and machine learning applications like scientific computing, predictive analytics, and large-scale data processing, which are all aimed at making package management and deployment easier. Version control for packages is handled via the conda package management system. Anaconda Distribution is used by more than 6 million people. It contains more than 250 well-known data science packages that are compatible with Windows, Linux, and MacOS.

Guido van Rossum created and released Python, a high-level programming language, in 1991. When it comes to building, controlling, managing, and testing software, it's the most common coding language. Python programmes can also be executed using this interpreter. Under Windows, you'll find the Python interpreter at python.exe. The Python interpreter is part of the Python distribution along with other tools and packages such as editors. Python has multiple distributions, one of which being Anaconda. The new Anaconda release includes the data science packages for R and Python. Formerly, it was called Continuum Analytics. More than 100 new packages have been added to Anaconda. Here at work, we utilise Anaconda for all things related to data science, machine learning, statistical analysis, and scientific computing. In October 2017, the most recent version of Anaconda was 5.0.1. Upgraded R language support is just one of the many helpful features added to Version5.0.1, which also fixes a few small problems. When 5.0.0 was first released, none of these features were available.

Machine learning is an area of AI that deals with learning algorithms (AI). The overarching goal of machine learning is to deduce patterns in data and then use those patterns to train models that humans can comprehend and use. Despite being a subfield of computer science, machine learning is distinct from more conventional methods of computation. Algorithms, in the context of traditional computing, are collections of predefined instructions for solving problems or performing calculations. Instead, computers can learn from data inputs using machine learning algorithms, which then employ statistical analysis to produce output values within a specified range. This is why machine learning is so useful: it lets computers automate decision-making by constructing models from data samples. Machine learning has been useful for any tech user in the modern day. Thanks to advancements in facial recognition technology, users of social networking sites may now easily tag and share images of their friends. Images of text can be transformed into movable type using optical character recognition (OCR) technology. Machine learning-powered recommendation algorithms take user tastes into account when deciding what to watch next on television or in the movies. Consumers may soon have access to self-driving cars that use machine learning for navigation. Machine learning is an area that is always evolving. As a result, whether you're developing machine learning approaches or assessing their effects, there are a few things to keep in mind. An introduction to supervised and unsupervised learning, as well as typical algorithmic

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

approaches to machine learning, is presented in this thesis. These approaches include deep learning, decision tree learning, and the k-nearest neighbour algorithm.

There are several overarching types of jobs in machine learning. These groups are established according to the ways in which the developed system is informed about the learning process or receives feedback on its instruction. Unsupervised learning, in which the algorithm is given unlabeled data and asked to discover structure within the input data, and supervised learning, in which algorithms are trained using example input and output data that has been labelled by humans, are two of the most popular machine learning methods. Now, let's delve deeper into these strategies. Supervised learning is the main method used in practical machine learning. Under supervised learning, given two variables, x and y, an algorithm is trained to learn the mapping function f from x to y. (X). In order to anticipate the values of the output variables (Y) given new input data (x), it is necessary to have a good approximation of the mapping function. Algorithms from Supervised Machine Learning encompass a wide range of techniques, including as logistic and linear regression, decision trees, support vector machines, multi-class classification, and decision trees. For supervised learning to work, the training data must already have the right responses labelled. As an example, a classification algorithm can be taught to recognise different kinds of animals by analysing a collection of photos that have been correctly annotated with the species and a few distinguishing features. There are two subsets of supervised learning problems: regression and classification. The final goal of both problems is to develop a simple model that can use the attribute variables to predict the dependent attribute's value. The dependent attribute in regression is numerical, whereas in classification it is categorical, which is the main distinction between the two types of tasks.

The absence of a relationship between the input data (X) and the output variables characterises unsupervised learning. To better understand data, unsupervised learning seeks to predict its underlying structure or distribution. Since there are no instructors or right answers in these types of learning, they are known as unsupervised learning. It is up to the algorithms to find and display the intriguing structure in the data. There are two subsets of unsupervised learning issues: clustering and association problems. Classification is "classifying objects" into subcategories, as the term implies. However, mechanically. That may not seem like much, but try to picture a computer that could tell the difference between you and a complete stranger. In the middle of a tomato and potato. On the cusp between an A and an F. In the fields of machine learning and statistics, classification refers to the challenge of assigning new observations to preexisting subpopulations according to predetermined category memberships in a training set of data (Figure 1).

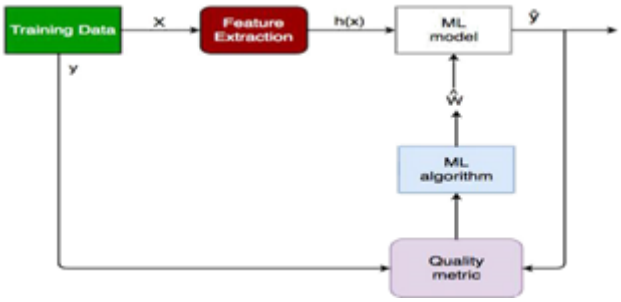

Figure 1: Generalized Classification Block Diagram [6]

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

When the output variable is a continuous or natural value, like "weight" or "salary," a regression problem arises. The most basic of these models is linear regression, but there are many more to choose from. In an effort to find the greatest fit, it uses a hyper-plane that passes through the points (Figure 2).

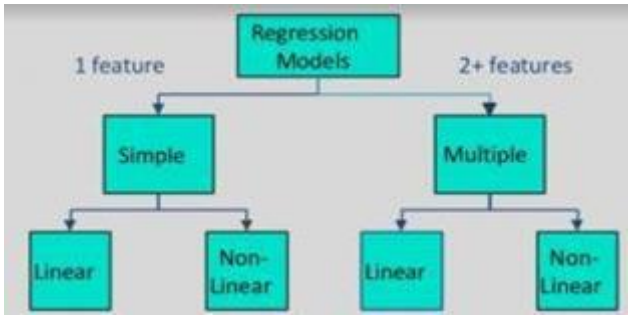

Figure 2: Types of Regression Models [7]

This approach to learning does not rely on human supervision. In unsupervised learning, we don't use labels to guide our reference extraction from databases of input data [109-114]. In a typical application, it is employed to discover the generative characteristics, underlying processes, meaningful structure, and groups present in a collection of cases. Data points in the same cluster are more likely to share characteristics with other points in the same cluster, whereas data points in different clusters are more likely to differ from one another [115-121]. This process is called clustering. It is a set of things organised according to how similar or different they are from one another. As an example, we can gather all of the data points that are clustered together into one set in the graph below. There are three distinct clusters visible in the image below, and we can tell them apart (Figure 3).

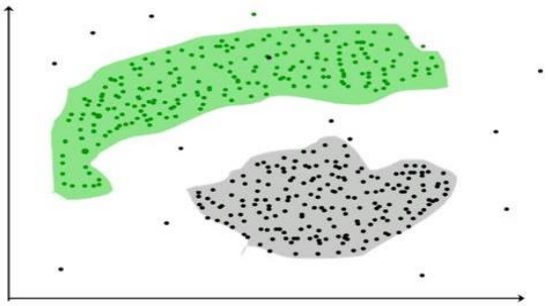

Figure 3: Clustering Example [8]

Clustering is based on the premise that each data point must be located within a certain distance from the cluster's centre. In order to determine the outliers, a number of distance algorithms and approaches are employed. Because it finds the inherent grouping among the unlabeled data, clustering is vital. Good clustering does not have any standards. Whatever criteria the user chooses to meet their demand is entirely up to them [122-129]. To give just a few examples, we may be looking for data reduction representatives, "natural clusters" with unknown attributes described as "natural" data kinds, "useful" data classes, or even strange data objects (outlier detection). The

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

similarity of points is based on assumptions made by this algorithm, and each assumption generates distinct but equally acceptable clusters.

Discussion

Gathering relevant data is a crucial first step in any machine learning initiative. Reason being, data is what we feed into the algorithms. The consistency and quality of the data obtained determine the efficiency and accuracy of the algorithm. The results will be the same as the input data. There are a lot of factors that go into a student's career forecast, including their test scores across the board, areas of concentration, memory, programming and analytical skills, relationships, interests, hobbies, sports, contests, hackathons, workshops, certifications, reading preferences, and a whole lot more. Because each of these things is important in determining a student's trajectory in terms of their chosen profession, they are all taken into account. There are a lot of methods to gather data. Some of the data comes from college alum databases, some from LinkedIn's API, some from randomly generated samples, and some from employees at various companies. We gathered data from around 20,000 records, each with 36 columns [130-134].

Gathering data is one thing but turning it into something valuable is an entirely different kettle of fish. Data gathered from different sources will be in a disorganised format with a high likelihood of missing or incorrect values as well as undesirable information [135-141]. The fundamental processes in pre-processing data include cleaning all of this data, replacing it with suitable or approximative data, eliminating missing or null values, and replacing it with some predetermined alternate values. The values contained in the obtained data can be entirely useless. Neither the format nor the intended execution may be perfect. In order to make the data relevant and usable for additional processing, it is necessary to verify and replace all such situations with other values. It is essential to maintain data in a structured manner [142-146].

Testing is the last step after data processing and training. Here you may see the desired output, data quality, and algorithm performance. We use 80% of the data for training and set aside 20% for testing from the massive dataset that we acquire. As previously said, training is the act of imparting knowledge to the machine so that it can learn and generate additional predictions using the training data. Tests, on the other hand, involve using a preset data set with labelled output to see if the model is accurate in its predictions. We can proceed with confidence in the model's accuracy if all of the maximum forecasts are correct; otherwise, it's best to try a different model. In order to find bugs related to interfaces, integration testing is a methodical approach to building the program's structure while running tests [147-149]. In other words, testing the product as a whole entail testing each of its individual parts. Using unproven modules, construct a programme structure is the goal. The tester needs to find the modules that are crucial. The earliest opportunity to test critical modules should be taken. One strategy is to wait for each unit to pass testing individually before combining and testing them. Starting with the unstructured testing of smaller applications, this method has grown. Making smaller batches of tested units and adding them to the final product is another option. Integrating and testing a small collection of modules is the first step in adding another module and testing it in tandem with the others. Continuation follows. The approach's benefits include the fact that interface dispenses are easily discoverable.

OPEN ACCESS JOURNALS
International Journal of Innovative Analyses and Emerging Technology
| e-ISSN: 2792-4025 | http://openaccessjournals.eu | Volume: 4 Issue: 01

A linkage fault was the most significant problem that arose throughout the process. The link is not set correctly with all support files when all the modules are integrated. The next thing we did was verify the connections and interconnections. The new module and all of its interactions are the exact source of the errors. Development of the product can be done in stages, with modules being integrated when unit tests are finished. Once testing of the last module is finished, testing is considered complete. The goal of testing is to identify and fix bugs in a programme. It is highly likely that an error that has not been found yet will be found by a good test case. If the test passes, it finds an error that hasn't been found yet. As a part of the implementation process, system testing verifies that the system performs as intended before going live. This ensures that the entire suite of programmes functions as expected. Adopting a new system successfully necessitates system testing, which entails a number of critical actions and stages. If you want to make sure everything is perfect before installing the system for user acceptability testing, this is your final chance. Program development, documentation, and data structure design are the first steps in the software testing process. Fixing bugs in software requires testing. In that case, we wouldn't call it a finished programme or project. At its core, software testing is an inspection of the code and design specifications that constitutes the pinnacle of software quality assurance.

Determine what needs doing and then go about doing it in a methodical fashion. Software testing templates, or a plan for including several test case creation methodologies, should, thus, exhibit the following qualities. A methodical approach to building the software structure and simultaneously testing it to find flaws is known as integration testing. Because of the high likelihood of interface problems, we cannot assume that the combined modules will immediately function. Naturally, the hurdle lies in "putting them together" or integrating them. Problems can arise with global data structures, there is a danger that data will be lost throughout sub-functions, and that combined functions won't generate the desired major function. Even acceptable impressions can be exaggerated to intolerable levels.

Product testing has identified the syntactic and logical mistakes. An error in programme code that does not adhere to the rules of the language it was written in is called a syntax error. Common syntax issues include missing keywords or field dimensions that are not defined correctly. The computer displays these mistakes with error messages. Conversely, erroneous data fields, out-of-range items, and invalid combinations are the domain of logic errors. The programmer must check the output for logical errors since compilers cannot do it. During condition testing, a module's logical conditions are exercised. Elements that can make up a condition can be anything from a relational operator to an arithmetic expression to a Boolean operator, variable, or pair of parentheses. Each programme condition is tested using the condition testing approach. One of the main goals of a condition test is to identify any faults within the programme, not just those related to its state. The purpose of security testing is to ensure that a system's built-in safeguards are effective in preventing unauthorised access. There needs to be testing to ensure the system is impenetrable from both frontal and rear attacks. The security exam involves the tester assuming the persona of an intruder.

Conclusion

The software is put together as a bundle at the process of software package assembly begins at We have found and fixed the interface problems, and now we can start the last round of software test validation. Many definitions exist for validation testing; one common one is that it is successful when the programme performs as the client reasonably anticipates. Black box tests show that the software is compliant with requirements, which is how software is validated. There are two possible outcomes following the validation test. Acceptance and conformity with specifications are indicated by the function or performance qualities. A validation that was not part of the specification is found, and a gap is filled. In order to fix any mistakes or deviations found at this stage of the project, the user and project manager work together to negotiate a solution. As a result, validation testing has confirmed that the under-consideration system is functioning as expected. There were some minor but noticeable flaws in the system. The success of every system hinges on the level of user acceptability. By maintaining continual communication with prospective systems and users during development and implementing adjustments as needed, the system being evaluated is thoroughly assessed for user acceptance.

References

1. Mouthami, K., Devi, K. N., & Bhaskaran, V. M. (2013). Sentiment analysis and classification based on textual reviews. 2013 International Conference on Information Communication and Embedded Systems (ICICES). IEEE.
2. Kyriakopoulou, A., & Kalamboukis, T. (2007). Using clustering to enhance text classification. Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, USA: ACM.
3. Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. ACM Transactions on Information Systems, 26(3), 1–34.
4. Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82–89.
5. Zhu, H., & Cao, G. (2006). RDCF: A relay-enabled medium access control protocol for wireless ad hoc networks. IEEE Transactions on Mobile Computing, 5(9), 1201–1214.
6. Awasthi, K.K., Kumar, M. (2022). Survey of Supervised Machine Learning Techniques in Wireless Sensor Network. In: Dhawan, A., Mishra, R.A., Arya, K.V., Zamarreño, C.R. (eds) Advances in VLSI, Communication, and Signal Processing. Lecture Notes in Electrical Engineering, vol 911. Springer, Singapore.
7. Linear Regression in Machine Learning. (2018). Retrieved January 10, 2024, from New Technology website: https://www.ntirawen.com/2018/06/linear-regression.html
8. Clustering in machine learning. (2018, January 15). Retrieved January 10, 2024, from GeeksforGeeks website: https://www.geeksforgeeks.org/clustering-in-machine-learning/
9. Rasul, H. O. (2023). Synthesis, evaluation, in silico ADMET screening, HYDE scoring, and

molecular docking studies of synthesized 1-trityl-substituted 1 H-imidazoles. Journal of the Iranian Chemical Society, 20(12), 2905-2916.

10. Rasul, H. O., Thomas, N. V., Ghafour, D. D., Aziz, B. K., Salgado M, G., Mendoza-Huizar, L. H., & Candia, L. G. (2023). Searching possible SARS-CoV-2 main protease inhibitors in constituents from herbal medicines using in silico studies. Journal of Biomolecular Structure and Dynamics, 1-15.
11. Rasul, H. O., Sabir, D. K., Aziz, B. K., Guillermo Salgado, M., Mendoza-Huizar, L. H., Belhassan, A., & Ghafour, D. D. (2023). Identification of natural diterpenes isolated from Azorella species targeting dispersin B using in silico approaches. Journal of Molecular Modeling, 29(6), 182.
12. Rasul, H. O., Aziz, B. K., Morán, G. S., Mendoza-Huizar, L. H., Belhassan, A., Candia, L. G., ... & Sadasivam, K. (2023). A Computational Study of The Antioxidant Power Of Eugenol Compared To Vitamin C. Química Nova, 46, 873-880.
13. Hjazi, A., Nasir, F., Noor, R., Alsalamy, A., Zabibah, R. S., Romero-Parra, R. M., ... & Akram, S. V. (2023). The pathological role of CXC chemokine receptor type 4 (CXCR4) in colorectal cancer (CRC) progression; special focus on molecular mechanisms and possible therapeutics. Pathology-Research and Practice, 154616.
14. Althomali, R. H., Al-Hawary, S. I. S., Gehlot, A., Qasim, M. T., Abdullaeva, B., Sapaev, I. B., ... & Alsalamy, A. (2023). A novel Pt-free counter electrode based on MoSe2 for cost effective dye-sensitized solar cells (DSSCs): Effect of Ni doping. Journal of Physics and Chemistry of Solids, 182, 111597.
15. Hjazi, A., Ahsan, M., Alghamdi, M. I., Kareem, A. K., Al-Saidi, D. N., Qasim, M. T., ... & Mirzaei, R. (2023). Unraveling the Impact of 27-Hydroxycholesterol in Autoimmune Diseases: Exploring Promising Therapeutic Approaches. Pathology-Research and Practice, 154737.
16. Gupta, J., Suliman, M., Ali, R., Margiana, R., Hjazi, A., Alsaab, H. O., ... & Ahmed, M. (2023). Double-edged sword role of miRNA-633 and miRNA-181 in human cancers. Pathology-Research and Practice, 154701.
17. Al-Hawary, S. I. S., Kadhum, W. R., Saleh, E. A. M., Yacin, Y., Abdullah, E. A., Qasim, M. T., ... & Alsalamy, A. (2023). Tunneling induced swapping of orbital angular momentum in a quantum dot molecule. Laser Physics, 33(9), 096001.
18. Gaffar Sarwar Zaman, Ibrahem Waleed, Ruaa Ali Obeid, Shaymaa Abdulhameed Khudair, Saafa Abaas Abd Al-Kahdum, Kadhum Al-Majdi, Ahmed S. Abed, Ali Alsalamy, Maytham T. Qasim, Ahmed Hussien Radie Alawadi. (2023). Electrochemical determination of zearalenone in agricultural food samples using a flower like nanocomposite-modified electrode, Materials Chemistry and Physics,Volume 305,127986.
19. Al-dolaimy, F., Kzar, M.H., Hussein, S.A. et al. (2023). Incorporating of Cobalt into UiO-67 Metal–Organic Framework for Catalysis CO2 Transformations: An Efficient Bi-functional Approach for CO2 Insertion and Photocatalytic Reduction. J Inorg Organomet Polym.

20. Muzammil Khursheed, Kzar Mazin Hadi, Mohammed Faraj, et al. (2023). Methanol extract of Iraqi Kurdistan Region Daphne mucronata as a potent source of antioxidant, antimicrobial, and anticancer agents for the synthesis of novel and bioactive polyvinylpyrrolidone nanofibers. Frontiers in Chemistry. Vol.1, 2296-2646.
21. Rasul, H. O., Aziz, B. K., Ghafour, D. D., & Kivrak, A. (2022). In silico molecular docking and dynamic simulation of eugenol compounds against breast cancer. Journal of molecular modeling, 28(1), 17.
22. Rasul, H. O., Aziz, B. K., Ghafour, D. D., & Kivrak, A. (2023). Discovery of potential mTOR inhibitors from Cichorium intybus to find new candidate drugs targeting the pathological protein related to the breast cancer: an integrated computational approach. Molecular Diversity, 27(3), 1141-1162.
23. Rasul, H. O., Aziz, B. K., Ghafour, D. D., & Kivrak, A. (2023). Screening the possible anti-cancer constituents of Hibiscus rosa-sinensis flower to address mammalian target of rapamycin: An in silico molecular docking, HYDE scoring, dynamic studies, and pharmacokinetic prediction. Molecular Diversity, 27(5), 2273-2296.
24. Batool, Kiran; Zhao, Zhen-Yu; Irfan, Muhammad; Żywiołek, Justyna (2023): Assessing the role of sustainable strategies in alleviating energy poverty: an environmental sustainability paradigm. w: Environmental science and pollution research international 30 (25), s. 67109–67130.
25. Nayyar, Anand; Żywiołek, Justyna; Rosak Szyrocka, Joanna; Naved, Mohd (2023): Advances in distance learning in times of pandemic. First edition. Boca Raton, FL: Chapman & Hall/CRC Press.
26. Zywiolek, Justyna; Matulewski, Marek; Santos, Gilberto (2023): The Kano Model As A Tool For Assessing The Quality Of Hunting Tourism - A Case From Poland. w: IJQR 17 (3), s. 1097–1112.
27. Żywiołek, Justyna (2018): Monitoring of information security system elements in the metallurgical enterprises. w: MATEC Web Conf. 183, s. 1007.
28. Żywiołek, Justyna (2019): Personal data protection as an element of management security of information. w: Multidisciplinary Aspects of Production Engineering 2 (1), s. 515–522.
29. Żywiołek, Justyna; Schiavone, Francesco: The Value of data sets in Information and Knowledge Management as a Threat to Information Security, Garcia-Perez, Alexeis; Simkin, Lyndon (red.), w: European Conference on Knowledge Management, s. 882–891, dostępne na stronie internetowej.
30. Żywiołek, Justyna; Schiavone, Francesco (2021): Perception of the Quality of Smart City Solutions as a Sense of Residents' Safety. w: Energies 14 (17), s. 5511..
31. Tak, A. (2021). Multi-Modal Fusion for Enhanced Image and Speech Recognition in AI Systems. International Journal of Science and Research, 10(6), 1780-1788.
32. Tak, A. (2021). The Data Mining Techniques for Analyzing Employee Performance and

Productivity. International Journal of Science and Research, 10(10), 1575-1578.

33. Tak, A. (2022). The Impact of Electronic Health Records on Patient Care in the US Healthcare System. Journal of Health Statistics Reports, 1(2), 1–7.
34. Tak, A. (2022). Big Data Analytics in Healthcare: Transforming Information into Actionable Insights. Journal of Health Statistics Reports, 1(3), 1-6.
35. Ananda Shankar Hati, and T. K. Chatterjee, "Symmetrical component filter based online condition monitoring instrumentation system for mine winder motor" Measurement (Elsevier), vol. 82, pp. 284-300, 2016.
36. Prashant Kumar and Ananda Shankar Hati "Review on Machine Learning Algorithm Based Fault Detection in Induction Motors," Archives of Computational Methods in Engineering, vol: 28, pp: 1929-1940, 2021.
37. Kumar Prashant and Hati, Ananda Shankar "Convolutional Neural Network with batch normalization for fault detection in SCIM," IET Electric Power Application, vol: 15, issue: 1, pp. 39-50, 2021.
38. Kumar Prashant and Hati, Ananda Shankar "Deep Convolutional Neural Network based on adaptive gradient optimizer for fault detection in SCIM," ISA Transactions, vol: 111, pp: 350-359, 2021.
39. Prince, Hati Ananda Shankar, Chakrabarti Prasun, Abawajy Jemal Hussein and Ng Wee Keong "Development of Energy Efficient Drive for Ventilation System using Recurrent Neural Network," Neural Computing and Applications, Vol. 33, no. 14, pp. 8659-8668, 2021.
40. Sinha Ashish Kumar, Hati Ananda Shankar, Benbouzid Mohamed and Chakrabarti Prasun "ANN-based Pattern Recognition for Induction Motor Broken Rotor Bar Monitoring under Supply Frequency Regulation" Machines (2021), vol: 9(5).
41. Prince and Hati Ananda Shankar "A Comprehensive Review of Energy-Efficiency of Ventilation System using Artificial Intelligence" Renewable and Sustainable Energy Reviews (2021), vol: 146, 2021.
42. Kumar Prashant and Hati, Ananda Shankar "Transfer Learning Based Deep CNN Model for Multiple Faults Detection in SCIM" Neural Computing and Applications (2021).
43. Prince and Hati Ananda Shankar "Temperature and Humidity Dependent MRAS Based Speed Estimation Technique for Induction Motor used in Mine Ventilation Drive" Journal of Mining Science, 2021, Vol. 57, No. 5, pp. 842–851.
44. Kumar Prashant and Hati, Ananda Shankar "Dilated Convolutional Neural Network Based Model For Bearing Faults and Broken Rotor Bar Detection in Squirrel Cage Induction Motors" Expert Systems With Applications (2022).
45. Prince and Hati Ananda Shankar "Convolutional Neural Network-Long Short Term Memory Optimization for Accurate Prediction of Airflow in a Ventilation System" Expert Systems with Applications (2022).

46. Vatsa Aniket and Hati Ananda Shankar "Depolarization Current Prediction of Transformers OPI System Affected From Detrapped Charge Using LSTM," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-11, 2022, Art no. 2511711.
47. Gorai Rahul, Hati Ananda Shankar, and Maity Tanmoy, "A new cascaded multilevel converter topology with a reduced number of components" 3rd IEEE 2017 Conference on International conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017), 21-22 September 2017 | IEEE, Chennai, India., pp. 539-543.
48. Kumar Prashant, Hati, Ananda Shankar, Sanjeevikumar Padmanaban, Leonowicz Zbigniew and Prasun Chakrabarti "Amalgamation of Transfer Learning and Deep Convolutional Neural Network for Multiple Fault Detection in SCIM" 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), 9th-12th June 2020, Madrid, Spain.
49. Sinha Ashish Kumar, Kumar Prashant, Prince and Hati, Ananda Shankar, "ANN Based Fault Detection Scheme for Bearing Condition Monitoring in SRIMs using FFT, DWT and Band-pass Filters" 2020 International Conference on Power, Instrumentation, Control, and Computing (PICC) 2020 IEEE.
50. Prince Kumar and Hati, Ananda Shankar, "Sensor-less Speed Control of Ventilation System Using Extended Kalman Filter For High Performance," 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2021, pp. 1-6.
51. Kumar Prashant and Hati, Ananda Shankar "Support Vector Classifiers based broken rotor bar detection in Squirrel cage induction motor" Machines, Mechanisms and Robotics, Springer, Singapore, 429-438.
52. Hati, Ananda Shankar, and Chatterjee, T. K., "Some studies on condition monitoring techniques for online condition monitoring and fault diagnosis of mine winder motor", International Journal of Engineering Science and Technology (IJEST), vol. 4, no. 08, pp. 3785-3793, August 2012.
53. Hati, Ananda Shankar, and Chatterjee, T. K., "Axial leakage flux-based online condition monitoring instrumentation system for mine winder motor" Journal of Mines, Metals & Fuels, vol. 63, no. 5&6, pp. 132-140, May-June 2015.
54. Hati, Ananda Shankar, and Chatterjee, T. K., "Current monitoring Instrumentation system for detecting airgap eccentricity in mine winder motor", International Journal of Applied Engineering Research, vol. 10, no. 22, pp. 43000-43007, 2015.
55. Hati, Ananda Shankar, "Vibration monitoring instrumentation system for detecting airgap eccentricity in mine winder motor" Journal of Mine Metals and Fuels, vol. 64, no. 5&6, pp. 240-248, May-June 2016.
56. Tak, A. (2023). The Role of Cloud Computing in Modernizing Healthcare IT Infrastructure. Journal of Artificial Intelligence & Cloud Computing, 2(2), 1–7.

57. Tak, A., & Sundararajan, V. (2023, December 2). Pervasive Technologies and Social Inclusion in Modern Healthcare: Bridging the Digital Divide. FMDB Transactions on Sustainable Health Science Letters, 1(3), 118-129.
58. H. Lakhani, D. Undaviya, H. Dave, S. Degadwala, and D. Vyas, “PET-MRI Sequence Fusion using Convolution Neural Network,” in 2023 International Conference on Inventive Computation Technologies (ICICT), 2023, pp. 317–321.
59. F. Ahamad, D. K. Lobiyal, S. Degadwala, and D. Vyas, “Inspecting and Finding Faults in Railway Tracks Using Wireless Sensor Networks,” in 2023 International Conference on Inventive Computation Technologies (ICICT), 2023, pp. 1241–1245.
60. D. Rathod, K. Patel, A. J. Goswami, S. Degadwala, and D. Vyas, “Exploring Drug Sentiment Analysis with Machine Learning Techniques,” in 2023 International Conference on Inventive Computation Technologies (ICICT), 2023, pp. 9–12.
61. C. H. Patel, D. Undaviya, H. Dave, S. Degadwala, and D. Vyas, “EfficientNetB0 for Brain Stroke Classification on Computed Tomography Scan,” in 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2023, pp. 713–718.
62. V. Desai, S. Degadwala, and D. Vyas, “Multi-Categories Vehicle Detection For Urban Traffic Management,” in 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), 2023, pp. 1486–1490.
63. D. Vyas and V. V Kapadia, “Evaluation of Adversarial Attacks and Detection on Transfer Learning Model,” in 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), 2023, pp. 1116–1124.
64. D. D. Pandya, S. K. Patel, A. H. Qureshi, A. J. Goswami, S. Degadwala, and D. Vyas, “Multi-Class Classification of Vector Borne Diseases using Convolution Neural Network,” in 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2023, pp. 1–8.
65. D. D. Pandya, A. K. Patel, J. M. Purohit, M. N. Bhuptani, S. Degadwala, and D. Vyas, “Forecasting Number of Indian Startups using Supervised Learning Regression Models,” in 2023 International Conference on Inventive Computation Technologies (ICICT), 2023, pp. 948–952.
66. S. Degadwala, D. Vyas, D. D. Pandya, and H. Dave, “Multi-Class Pneumonia Classification Using Transfer Deep Learning Methods,” in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2023, pp. 559–563.
67. D. D. Pandya, A. Jadeja, S. Degadwala, and D. Vyas, “Diagnostic Criteria for Depression based on Both Static and Dynamic Visual Features,” in 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), 2023, pp. 635–639.
68. K. Butchi Raju, P. Kumar Lakineni, K. S. Indrani, G. Mary Swarna Latha and K. Saikumar, "Optimized building of machine learning technique for thyroid monitoring and analysis," 2021

2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 1-6.

69. P. K. Lakineni, D. J. Reddy, M. Chitra, R. Umapriya, L. V. Kannan and S. R. Barkunan, "Optimal Feature Selection and Classification Using Convolutional Neural Network-Based Plant Disease Prediction," 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 2023, pp. 1-6.
70. P. K. Lakineni, S. Kumar, S. Modi, K. Joshi, V. Mareeskannan and J. Lande, "Deepflow: A Software-Defined Measurement System for Deep Learning," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1217-1221.
71. P. K. Lakineni, R. Singh, B. Mandaloju, S. Singhal, M. D. Bajpai and M. Tiwari, "A Cloud-Based Healthcare Diagnosis Support Network for Smart IoT for Predicting Chronic Kidney Failure," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1858-1863.
72. P. K. Lakineni, K. M. Nayak, H. Pallathadka, K. Gulati, K. Pandey and P. J. Patel, "Fraud Detection in Credit Card Data using Unsupervised & Supervised Machine Learning-Based Algorithms," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-4.
73. V. Dankan Gowda, K. Prasad, R. Shekhar, R. Srinivas, K. N. V. Srinivas and P. K. Lakineni, "Development of a Real-time Location Monitoring App with Emergency Alert Features for Android Devices," 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2023, pp. 1-8.
74. Karimisetty, S., Kumar Lakineni, P., Kamalanjali, M.L., Parwekar, P. (2021). Automated Water Flow Control System in Overhead Tanks Using Internet of Things and Mobile Application. In: Mahapatra, R.P., Panigrahi, B.K., Kaushik, B.K., Roy, S. (eds) Proceedings of 6th International Conference on Recent Trends in Computing. Lecture Notes in Networks and Systems, vol 177. Springer, Singapore.
75. C. S. Rao, A. Balakrishna, K. B. Raju, L. P. Kumar, S. V. Raju and G. V. P. Raju, "Implementation of Interactive Data Knowledge Management and Semantic Web for Step-Data from Express Entities," 2010 3rd International Conference on Emerging Trends in Engineering and Technology, Goa, India, 2010, pp. 537-542.
76. K. B. Raju, C. S. Rao, L. PrasannaKumar and S. V. Raju, "PDM data classification from step - An object oriented string matching approach," 2011 5th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, 2011, pp. 1-9.
77. D. R. Giri, S. P. Kumar, L. Prasannakumar and R. N. V. V. Murthy, "Object oriented approach to SQL injection preventer," 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), Coimbatore, India, 2012, pp. 1-7.

78. K. Butchi Raju, P. Kumar Lakineni, K. S. Indrani, G. Mary Swarna Latha and K. Saikumar, "Optimized building of machine learning technique for thyroid monitoring and analysis," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 1-6.
79. R. Rai, V. Y. Nguyen, and J. H. Kim, "Variability analysis and evaluation for major cut flower traits of F1 hybrids in Lilium brownii var. colchesteri," Journal of multidisciplinary sciences (Online), vol. 4, no. 2, pp. 35–41, Dec. 2022.
80. V. Y. Nguyen, R. Rai, J.-H. Kim, J. Kim, and J-K Na, "Ecogeographical variations of the vegetative and floral traits of Lilium amabile Palibian," Journal of Plant Biotechnology, vol. 48, no. 4, pp. 236–245, Dec. 2021.
81. R. Rai and J.H. Kim, "Performance Evaluation And Variability Analysis For Major Growth And Flowering Traits Of Lilium longiflorum Thunb. Genotypes," Journal of Experimental Biology and Agricultural Sciences, vol. 9, no. 4, pp. 439–444, Aug. 2021.
82. R. Rai, V. Y. Nguyen, and J. H. Kim, "Estimation Of Variability Analysis Parameters For Major Growth And Flowering Traits Of Lilium leichtlinii var. maximowiczii GERMPLASM," Journal of Experimental Biology and Agricultural Sciences, vol. 9, no. 4, pp. 457–463, Aug. 2021.
83. R. Rai and J. H. Kim, "Effect Of Storage Temperature And Cultivars On Seed Germination OF Lilium×formolongi HORT.," Journal of Experimental Biology and Agricultural Sciences, vol. 8, no. 5, pp. 621–627, Oct. 2020.
84. R Rai, &J.H. Kim ,Estimation of combining ability and gene action for growth and flowering traits in Lilium longiflorum.International Journal of Advanced Science and Technology,Vol.29 No.8S pp 1356-1363,2020
85. R. Rai, A. Badarch, and J.-H. Kim, "Identification Of Superior Three Way-Cross F1s, Its Line×Tester Hybrids And Donors For Major Quantitative Traits In Lilium×formolongi," Journal of Experimental Biology and Agricultural Sciences, vol. 8, no. 2, pp. 157–165, Apr. 2020.
86. R. Rai, J. Shrestha, and J. Kim, "Line×tester analysis in lilium×formolongi: identification of superior parents for growth and flowering traits," SAARC Journal of Agriculture, vol. 17, no. 1, pp. 175–187, Aug. 2019.
87. R. Rai,J.Shrestha and J.H.Kim "Combining Ability and Gene Action Analysis of Quantitative Traits in Lilium × formolongi," vol. 30, no. 3, pp. 131–143, Dec. 2018.
88. T. X. Nguyen, S.-I. Lee, R. Rai, N. Kim, and Jong Hwa Kim, "Ribosomal DNA locus variation and REMAP analysis of the diploid and triploid complexes ofLilium lancifolium," Genome, vol. 59, no. 8, pp. 551–564, Aug. 2016.
89. N. X. Truong, J. Y. Kim, R. Rai, J. H. Kim, N. S. Kim, and A. Wakana, "Karyotype Analysis of Korean Lilium maximowiczii Regal Populations," Journal of The Faculty of Agriculture Kyushu University, vol. 60, no. 2, pp. 315–322, Sep. 2015.

90. Gaurav Kumawat, Santosh Kumar Viswakarma, Prasun Chakrabarti, Pankaj Chittora, Tulika Chakrabarti , Jerry Chun-Wei Lin, "Prognosis of Cervical Cancer Disease by Applying Machine Learning Techniques", Journal of Circuits, Systems, and Computers, 2022.
91. Akhilesh Kumar Sharma, Gaurav Aggarwal, Sachit Bhardwaj, Prasun Chakrabarti, Tulika Chakrabarti, Jemal Hussain, Siddhartha Bhattarcharyya, Richa Mishra, Anirban Das, Hairulnizam Mahdin, "Classification of Indian Classical Music with Time-Series Matching using Deep Learning", IEEE Access, 9 : 102041-102052 , 2021.
92. Akhilesh Kumar Sharma, Shamik Tiwari, Gaurav Aggarwal, Nitika Goenka, Anil Kumar, Prasun Chakrabarti, Tulika Chakrabarti, Radomir Gono, Zbigniew Leonowicz, Michal Jasiński , "Dermatologist-Level Classification of Skin Cancer Using Cascaded Ensembling of Convolutional Neural Network and Handcrafted Features Based Deep Neural Network", IEEE Access , 10 : 17920-17932, 2022.
93. Abrar Ahmed Chhipa , Vinod Kumar, R. R. Joshi, Prasun Chakrabarti, Michal Jaisinski, Alessandro Burgio, Zbigniew Leonowicz, Elzbieta Jasinska, Rajkumar Soni, Tulika Chakrabarti, "Adaptive Neuro-fuzzy Inference System Based Maximum Power Tracking Controller for Variable Speed WECS", Energies ,14(19) :6275, 2021.
94. Chakrabarti P. , Goswami P.S., "Approach towards realizing resource mining and secured information transfer", International Journal of Computer Science and Network Security, 8(7), pp.345-350, 2008.
95. Chakrabarti P., Choudhury A., Naik N. , Bhunia C.T., "Key generation in the light of mining and fuzzy rule", International Journal of Computer Science and Network Security, 8(9), pp.332-337, 2008.
96. Chakrabarti P., De S.K., Sikdar S.C., "Statistical Quantification of Gain Analysis in Strategic Management" , International Journal of Computer Science and Network Security,9(11), pp.315-318, 2009.
97. Chakrabarti P. , Basu J.K. , Kim T.H., "Business Planning in the light of Neuro-fuzzy and Predictive Forecasting", Communications in Computer and Information Science , 123, pp.283-290, 2010.
98. Prasad A. , Chakrabarti P., "Extending Access Management to maintain audit logs in cloud computing", International Journal of Advanced Computer Science and Applications ,5(3),pp.144-147, 2014.
99. Sharma A.K., Panwar A., Chakrabarti P. ,Viswakarma S., "Categorization of ICMR Using Feature Extraction Strategy and MIR with Ensemble Learning", Procedia Computer Science, 57,pp.686-694,2015.
100. S. Mandvikar, "Factors to Consider When Selecting a Large Language Model: A Comparative Analysis," International Journal of Intelligent Automation and Computing, vol. 6, no. 3, pp. 37–40, 2023.
101. S. Mandvikar, "Augmenting intelligent document processing (IDP) workflows with

contemporary large language models (LLMs)," International Journal of Computer Trends and Technology, vol. 71, no. 10, pp. 80–91, 2023.

102. R. Boina, A. Achanta, and S. Mandvikar, "Integrating data engineering with intelligent process automation for business efficiency," International Journal of Science and Research, vol. 12, no. 11, pp. 1736–1740, 2023.
103. S. Mandvikar and A. Achanta, "Process automation 2.0 with generative AI framework," Int. J. Sci. Res. (Raipur), vol. 12, no. 10, pp. 1614–1619, 2023.
104. Mandvikar, S. (2023). Indexing robotic process automation products. International Journal of Computer Trends and Technology, 71(8), 52–56.
105. Mohd Akbar, Irshad Ahmad, Mohsina Mirza, Manavver Ali, Praveen Barmavatu "Enhanced authentication for de-duplication of big data on cloud storage system using machine learning approach", Cluster Computing, Springer Publisher , 2023.
106. S. Rangineni, A. Bhanushali, M. Suryadevara, S. Venkata, and K. Peddireddy, "A Review on Enhancing Data Quality for Optimal Data Analytics Performance," International Journal of Computer Sciences and Engineering, vol. 11, no. 10, pp. 51–58, 2023.
107. S. Rangineni, A. Bhanushali, D. Marupaka, S. Venkata, and M. Suryadevara, "Analysis of Data Engineering Techniques With Data Quality in Multilingual Information Recovery," International Journal of Computer Sciences and Engineering, vol. 11, no. 10, pp. 29–36, 1973.
108. A. K. Bhardwaj, S. Rangineni, and D. Marupaka, "Assessment of Technical Information Quality using Machine Learning," International Journal of Computer Trends and Technology, vol. 71, no. 9, pp. 33–40, 2023.
109. S. Sellamuthu et al., "AI-based recommendation model for effective decision to maximise ROI," Soft Computing, Jul. 2023
110. L. Thammareddi, M. Kuppam, K. Patel, D. Marupaka, and A. Bhanushali, "An extensive examination of the devops pipelines and insightful exploration," International Journal of Computer Engineering and Technology, vol. 14, no. 3, pp. 76–90, 2023.
111. S. Parate, L. ThammaReddi, S. Agarwal, and M. Suryadevara, "Analyzing the Impact of Open Data Ecosystems and Standardized Interfaces on Product Development and Innovation," International Journal of Advanced Research in Science, Communication and Technology, vol. 3, no. 1, pp. 476–485, 2023.
112. Kaushikkumar Patel "A Review on Enhancing Data Quality for Optimal Data Analytics Performance," International Journal of Computer Sciences and Engineering, vol. 11, no. 10, pp. 51–58, 2023
113. Kaushikkumar Patel, Amit Bhanushali, Satish Kumar, "Evaluating regression testing performance through machine learning for test case reduction," International Journal of Computer Engineering and Technology, vol. 14, no. 3, pp. 51–66, 2023.
114. K, Patel, "Big Data in finance: An architectural overview", International Journal of Computer

Trends and Technology, Vol. 71, no. 10, pp. 61-68, 2023.

115. Kaushikkumar Patel, "Credit Card Analytics: A Review of Fraud Detection and Risk Assessment Techniques," International Journal of Computer Trends and Technology, vol. 71, no. 10, pp. 69-79, 2023.
116. Latha, Thammareddi, Shashank, Agarwal, Amit, Bhanushali, Kaushikkumar, Patel, Srinivas, Reddy V, "Analysis On Cybersecurity Threats in Modern Banking and Machine Learning Techniques For Fraud Detection" The Review of Contemporary Scientific and Academic Studies, vol. 3, no. 11, pp.
117. Shashank, Agarwal "Graph Networks: Transforming Provider Affiliations for Enhanced Healthcare Management," International Journal of Computer Trends and Technology, vol. 71, no. 6, pp. 86-90, 2023.
118. Shashank, Agarwal, Siddharth Sharma, Sachin Parate "Exploring the Untapped Potential of Synthetic data: A Comprehensive Review," International Journal of Computer Trends and Technology, vol. 71, no. 6, pp. 86-90, 2023.
119. Kulbir, Singh, "Artificial Intelligence & Cloud in Healthcare: Analyzing Challenges and Solutions Within Regulatory Boundaries," SSRG International Journal of Computer Science and Engineering, vol. 10, no. 9, pp. 1-9, 2023.
120. Amit Bhanushali, "Challenges and Solutions in Implementing Continuous Integration and Continuous Testing for Agile Quality Assurance," International Journal of Science and Research (IJSR), Vol. 12, no.10, pp. 1626-1644, 2023.
121. Chadrasekar.P, Beaulah david, H.Shaheen , " Dissemination of Link State Information for Enhancing Security in Mobile Ad Hoc Networks", IOSR Journal of Computer Engineering (IOSR-JCE), Vol 16, Issue 1, Ver VIII, PP 24-31, February 2014.
122. H. Shaheen, "Characteristics-Centered Cryptography with Cloud Revocation Authority and its Application", "International Journal of Computer and Mathematical Science", Volume 7, Issue 5, May 2018.
123. H. Shaheen, "Modelling and Analytical Proofing of Low Energy Temperature Control using Earth/Ground Water Heat Exchanger", "International Journal of Pure and Applied Mathematics",Volume 119 No. 16 2018, 3575-3588.
124. H. Shaheen, "Secured Message Exchange in Mission Critical Infrastructure using Conditional Privacy Preserving Authentication", "International Journal of Computer and Mathematical Science", Volume 7, Issue 5, May 2018.
125. H. Shaheen, "Web Innovation with IOT in Social Environment for Sharing Efficient Information", "International Journal of Mechanical and Production Engineering Research and Development",Vol. 8, Special Issue 6, Aug 2018, 717-729.
126. H. Shaheen, "Wireless Sensing Element Network-Based Pet Location Observation System for Domestic Situations", "International Journal of Recent Advances in Science and Technology", Vol 5, Issue 3, 2018, 41- 49.

127. H Shaheen, "A Pervasive Multi-Distribution Perceptron and Hidden Markov Model For Context Aware Systems", "The Journal of Research on the Lepidoptera ". June 2020, Volume 51 (2): 818-833.
128. H.Shaheen, "IoT Based System on Chip for Multiple Applications", "Journal of Advanced Research in Dynamical & Control Systems", Vol. 10, 13-Special Issue, 2018.
129. H.Shaheen, "An IoT Based System to Detect a Person/Wheelchair Fall", "International Journal of Engineering & Technology", 7 (3.20) (2018) 735-737.
130. H.Shaheen, "Secured Data Transmission in VANET using Vehicular Digital Hash Gen Model", "International Journal of Engineering and Advanced Technology", Vol. 8, Special Issue 6, Aug 2019, 1008-1011.
131. M. Farooq and M. Hassan, "IoT smart homes security challenges and solution," International Journal of Security and Networks, vol. 16, no. 4, p. 235, 2021.
132. M. Farooq and M. Khan, "Signature-Based Intrusion Detection System in Wireless 6G IoT Networks," Journal on Internet of Things, vol. 4, no. 3, pp. 155–168, 2023.
133. M. Farooq, "Artificial Intelligence-Based Approach on Cybersecurity Challenges and Opportunities in The Internet of Things & Edge Computing Devices," International Journal of Engineering and Computer Science, vol. 12, no. 07, pp. 25763–25768, Jul. 2023.
134. M. Farooq, "Supervised Learning Techniques for Intrusion Detection System Based on Multi-layer Classification Approach," International Journal of Advanced Computer Science and Applications, vol. 13, no. 3, 2022.
135. M. Farooq, R. Khan, and M. H. Khan, "Stout Implementation of Firewall and Network Segmentation for Securing IoT Devices," Indian Journal of Science and Technology, vol. 16, no. 33, pp. 2609–2621, Sep. 2023.
136. M. M. Kirmani and A. Wahid, "Impact of modification made in re-UCP on software effort estimation," Journal of Software Engineering and Applications, vol. 08, no. 06, pp. 276–289, 2015.
137. M. M. Kirmani and A. Wahid, "Revised use case point (re-UCP) model for software effort estimation," International Journal of Advanced Computer Science and Applications, vol. 6, no. 3, 2015.
138. R. Regin, S. Suman Rajest, Shynu T, & Steffi. R. (2023). Planning the Most Effective Itinerary for Tourists through the use of Data Analysis. International Journal of Human Computing Studies, 5(12), 77-92.
139. S. Suman Rajest, R. Regin, Shynu T, & Steffi. R. (2023). An Approach Based on Machine Learning for Conducting Sentiment Analysis on Twitter Data. International Journal of Human Computing Studies, 5(12), 57-76.
140. S. Suman Rajest, R. Regin, Shynu T, & Steffi. R. (2023). Using Voice Guidance, an Intelligent Walking Assistance Mechanism for the Blind. Central Asian Journal of Theoretical and

Applied Science, 4(11), 41-63.

141. S. Suman Rajest, S. Silvia Priscila, R. Regin, Shynu T, & Steffi. R. (2023). Application of Machine Learning to the Process of Crop Selection Based on Land Dataset. International Journal on Orange Technologies, 5(6), 91-112.
142. Shynu T, S. Suman Rajest, R. Regin, & Steffi. R. (2023). Android Application for Remote Control of Personal Computers. International Journal on Orange Technologies, 5(12), 44-58.
143. Shynu T, S. Suman Rajest, R. Regin, & Steffi. R. (2023). Region Segmentation and Support Vector Machine for Brain Tumour Stage Analysis, Detection, and Automatic Classification. Central Asian Journal of Medical and Natural Science, 25-43.
144. Steffi. R, Shynu T, S. Suman Rajest, & R. Regin. (2023). A Convolutional Neural Network with a U-Net for Brain Tumor Segmentation and Classification. Central Asian Journal of Medical and Natural Science, 4(6), 1326-1343.
145. Suman Rajest, S., Regin, R., Y, A., Paramasivan, P., Christabel, G. J. A., & T, Shynu. (2023). The Analysis of How Artificial Intelligence Has an Effect on Teachers and The Education System. EAI Endorsed Transactions on E-Learning, 9(4), 1-10.
146. Sundararajan, V., Steffi, R., & Shynu, T. (2023). Data Fusion Strategies for Collaborative Multi-Sensor Systems: Achieving Enhanced Observational Accuracy and Resilience. FMDB Transactions on Sustainable Computing Systems, 1(3), 112–123.
147. Syed Immamul Ansarullah, Syed Mohsin Saif, Pradeep Kumar, Mudasir Manzoor Kirmani, "Significance of Visible Non-Invasive Risk Attributes for the Initial Prediction of Heart Disease Using Different Machine Learning Techniques", Computational Intelligence and Neuroscience, vol. 2022, Article ID 9580896, 12 pages, 2022.
148. Syed Immamul Ansarullah, Syed Mohsin Saif, Syed Abdul Basit Andrabi, Sajadul Hassan Kumhar, Mudasir M. Kirmani, Dr. Pradeep Kumar, "An Intelligent and Reliable Hyperparameter Optimization Machine Learning Model for Early Heart Disease Assessment Using Imperative Risk Attributes", Journal of Healthcare Engineering, vol. 2022, 12 pages, 2022.
149. V. K. Nomula, R. Steffi, and T. Shynu, “Examining the Far-Reaching Consequences of Advancing Trends in Electrical, Electronics, and Communications Technologies in Diverse Sectors,” FMDB Transactions on Sustainable Energy Sequence, vol. 1, no. 1, pp. 27–37, 2023.