[1] **Sandramol M Kamal**

[2]**Dr. A. S. Aneeshkumar**

A literature study on Cybercrime in Social Media using Genetic Algorithm based Datamining Techniques

Abstract: - An organization or an individual would often experience the effects of cybercrime. Clustering techniques like K-means, hierarchical clustering and combination of Multiple Classifiers (CMC) are extensively used in cybercrime detection. Automated identification and analysis of criminal networks has received relatively little attention in last twenty years, despite the fact that social network analysis and social community mining are major areas of study. The methods presented in this study are primarily concerned with cybercrime forensics tools, specifically with methods for visualizing cybercriminal networks and methods for analyzing online social media posts to reveal different cybercriminal links. Crime data analysis has made excellent use of clustering methods in recent years. Therefore, it is important to investigate the causes, variables, and correlations of crime in order to identify the best means of crime control and prevention. In this study, we aim to categories clustered crimes according to their frequency throughout years. Data mining has found widespread use is in criminal activity analysis, investigation and pattern identification.

Keywords: Crime; Clustering; K-Means Algorithm; Genetic Algorithm

Introduction

Due to development of more sophisticated information technology. The number of cybercrimes has increased at a far faster rate than in prior years. Current cyber defences, such intrusion prevention systems and anti-malware software, aren't very good at predicting or forensic criminality. In order to detect cybercrimes, current cyber security systems use software coding signatures and low-level aspects of network traffic. The difficulty in detecting cybercriminals from the fact that they are capable of continually changing their assault strategies. Therefore, in order to tackle the ever-increasing trend of cybercrimes, this article applies cutting-edge computational approaches to uncover the subterranean networks used by cybercriminals. Dark marketplaces have emerged in online social media, and there is mounting evidence that cybercriminals use them to trade information about cyber attacks and even buy and sell cyber attack gear. This development presents researchers and analysts in the field of cyber defence with an opportunity to get a deeper understanding of cybercrime and cybercriminal networks by mining online social media for conversational communications. Cybercrime forensics may be made more efficient and successful by using an automated way of mining cybercriminal networks [1].

Along with evolutionary programming, evolutionary methods, and genetic programming, the genetic algorithm is a member of the evolutionary algorithm family. As a general category of stochastic optimisation methods, evolutionary algorithms cover a lot of ground. An evolutionary algorithm keeps track of a pool of potential answers to the current challenge. A collection of stochastic operators is applied iteratively to the population, causing it to develop. Mutation, recombination, and selection, or a mixture of these, are the common operators in this collection. Let's talk about a few of the publications that have been published in the past that deal with Genetic Algorithms [2].

Crime Analysis

Today, security services cannot function without collecting and analysing data pertaining to crimes. Most importantly, we need to figure out how to organise this data coherently according to frequency and location, find the underlying pattern in the crimes that happened at various periods, and forecast how they will relate to each other in the future. Using real-world datasets and presenting an appropriate framework that is robust against outliers are important steps to go in this direction [12,13]. Outliers have a major impact on the outcomes during

[1] Research Scholar, Research Department of Computer Science, AJK College of Arts and Science, Coimbatore, Tamilnadu-641105, India

[2]Research Supervisor and Head, Research Department of Computer Science and Applications, AJK College of Arts and Science, Coimbatore, Tamilnadu-641105, India

ORCID: 0000-0001-7196-1894

preprocessing, a crucial step in data mining. Thus, it is necessary to use a suitable method for locating and removing the data points that constitute outliers. Optimization of outlier detection operator settings using GA and formulation of a fitness function are both grounded in accuracy and classification error. Utilizing the weighting strategy allowed for the elimination of low-value features, which had a domino effect on data clustering and classification quality, resulting in decreased prediction accuracy and increased classification error.[3].

The following are some of the primary goals of criminal analysis:

- Using crime data and research to identify trends in criminal activity,
- Predicting crime rates using a range of data mining techniques and crime rates depending on the distribution of data across different regions
- Identifying criminals.

Clustering

Clustering is the process of dividing a dataset or collection of items into many groups. Clusters are therefore collections of data that exhibit consistent group behaviour due to their shared characteristics. One may argue that clustering is synonymous with classification; the main distinction is that clustering does not need human oversight and the classes are not predefined [5,6].

Clustering by K-means Algorithm

There are a number of clustering algorithms used in scientific and industrial applications, but K-means is by far the most popular and easiest to implement [7, 8]. A major reason for K-means's popularity is how easy it is to implement. There is a linear relationship between the number of data points and the algorithm's computing cost, making it ideal for clustering huge datasets [5]. While this method is easy to understand and implement, it does have a few drawbacks that can reduce its efficiency. These include: being susceptible to outlier data, dealing with high-dimensional data, and being too sensitive to cluster centres when creating initial clusters, which increases the risk of getting stuck in a local minimum [10,11].

Literature Review

The crime rate has been declining from 1990 to 2011, according to research by J. Agarwal, R. Nagpal, and R. Sehgal in [14]. The authors took the relevant year into consideration while analysing the crime rate. Because RapidMiner is a robust and comprehensive package with configurable support options, they have employed the k-means clustering algorithm to extract meaningful information from the crime dataset.

In order to forecast the frequency of criminal incidents in Delhi (India), J. Han used a linear regression in [15]. In order to forecast the frequency of certain crimes, such as robbery, burglary, murder, and so on, they examine a dataset spanning the last 59 years. The local police stations will find their work useful for criminal monitoring and decision making.

In this study, Bryant examined the outcomes of several mutation and crossover operators developed for the travelling salesman problem. He found that operators using heuristic knowledge or a matrix representation of the graph produced the most favourable outcomes. To find optimal solutions to optimisation issues, genetic algorithms use an evolutionary strategy known as "survival of the fittest" by making use of mutation and crossover operators. Their effective application spans several difficulties, one of them being the travelling salesman dilemma. The objective of the traveling salesman problem is to find a route across a weighted graph such that the total weight is minimized. There are a lot of practical uses for finding a good solution to the NP-hard traveling salesman problem. [19].

The term "Evolutionary Algorithms" was first used by Carlos Coello in his article "Evolutionary Algorithms," which describes a class of stochastic optimisation methods inspired by evolution. In addition to explaining how EAs generally work, this article also provides a high-level overview of these methods, outlining the primary groups into which they fall. Then it delves into the various parts of an EA and shows you how to put them into action with several examples. Last but not least, it provided a sneak peek at the many uses for these methods. Developmental calculations, hereditary programming, transformative techniques, and transformative writing

computer programs are a portion of the strategies accessible. The representation schemes, reproduction operators, and selection procedures are the crux of the disagreements between these models [20].

Hanne offered a coordination mechanism and GA as a solution to real-world scheduling challenges. Effective creation the board and convenient conveyance are two of the most challenging issues to handle in today's dynamic circumstances. Finding an optimal allocation plan that maximises particular performance measures is the objective of scheduling, which entails assigning a group of machines to complete a series of tasks within a certain time frame. To address the implementation concerns, we use the order crossover operator and encode the solutions using natural representation. As a mutation operator, they used the inversion method [21].

At last, the job-shop scheduling issue was resolved by Madureia et al., who demonstrated the viability of GA by solving it using a series of static scheduling methods. Sandstrom proposed GA, which is used to define priority for tasks and offset them to ensure that real-time timing restrictions are met. Assigning tasks with specific time constraints is a complex issue in real-time systems. They demonstrated how properties of periodic tasks executed on conventional preemptive RTOSes might be transferred to time restrictions. They turned to GA because, when all time constraints cannot be satisfied, it may provide a solution that fulfils some of the criteria. As a population grows, GA, a natural selection process, becomes better at assigning time limitations to individuals. The results from the various test scenarios that it was tested on were positive [22].

In his article titled "Test functions for optimisation needs," Molga proposes a literature review that will serve as a standard for optimisation processes that are specifically designed to tackle multidimensional, continuous optimisation tasks. The quality test for competing optimisation techniques (e.g., GA, SA, TS, etc.) has been multiple-extreme functions, which have received particular focus. Common standard literature standards are often used to assess the quality of optimisation processes, both existing and proposed. All of these test functions are continuous and fall into one of several categories: There are four types of multimodal distributions: one is convex and has one mode; another is two-dimensional and has few local extremes; a third is multimodal and has many local extremes; and finally, a fourth is multimodal and complicated.. When the first class includes both benign and harmful functions, the convergence to a single global extreme is either very slow or nonexistent. In an unpleasant environment, defined as one with few local extremes and one global one, the quality of conventional optimisation processes is tested in Class 2, which is intermediate between Classes 1 and 3. When evaluating intelligent resistant optimisation algorithms, it is best to use classes three and four [23].

Omar put out a GA-based solution for jobshop scheduling. They created a starting population at random, which included the outcome of applying certain popular priority criteria, such the quickest and longest processing times. After then, the population will continue reproducing via mutation, crossover, and reproduction until the set stopping requirements are met. As a cutoff point, the study uses the number of generations. The key block neighbourhood and measured distance are used to assess the schedules in mutation and crossover. The results demonstrate that various approaches may achieve the same outcome when crucial block neighbourhood and measured distance are used [24].

The "Genetic Algorithm approach to Operating system process scheduling problem" was proposed by Inazawa in his article. Operating system scheduling greatly affects system throughput and performance as a whole. The performance of the system depends on an efficient scheduling. Schedules are thought of as NP-hard problems. In order to provide effective process scheduling, the genetic algorithm is used. Obtaining a scheduler that efficiently allocates and schedules the task to the CPU is the goal [25].

In his research titled "A Genetic Algorithm on Single Machine Scheduling Problem to Minimise Total Weighted Completion Time," Dehuri proposed several ideas. In this article, he discusses a scheduling issue involving a single machine family and several workloads. There is a setup time between each family of jobs, and each family has its own processing time and positive weight. To solve this issue, he suggests a genetic algorithm that takes use of an undirected bipartite graph's optimised crossover operator to discover the best possible schedule that minimises the overall weighted completion time of the tasks, all while taking sequence independent family setup durations into account [26].

According to Snehal Kamalapur's "Efficient CPU Scheduling: A Genetic Algorithm based Approach" research, CPU scheduling has a significant impact on the performance and throughput of operating systems. An NP issue

is what scheduling is thought of as. Optimal scheduling enhances the efficiency of the system. She describes and assesses a process scheduling strategy in her article. She wrote about genetic algorithms and how they might improve process scheduling in this work. And see how well the suggested method works in contrast to existing deterministic algorithms, all while optimising for certain performance metrics via simulation [27,28,29].

S. Ramya proposed a novel method for window constrained scheduling in his article titled "Window Constrained Scheduling of Processes in Real Time CPU Using Multi Objective Genetic Algorithm." This method is well-suited for real-time systems that are weakly hard. Virtual Deadline Scheduling (VDS) is the original technique that aims to ensure that real-time operations, such periodic CPU workloads, are serviced by meeting m out of k deadlines. VDS can use all of its resources to generate a realistic window-constrained schedule. Virtual Desktop Service (VDS) should refresh the related virtual cutoff time at whatever point it serves an undertaking or changes the solicitation time frame. The method becomes more time-consuming due to this update, which is a bottleneck. Additionally, the number of context flips rises as VDS attempts to resolve the latency issue. Delay and context switching are two criteria that are at odds with one another. It is possible to strike a balance between the two parameters of context switching and latency by using a multi-objective genetic algorithm. Since updating is an extra step in the original VDS algorithm, we made sure to include it into our design so that it can also handle updates [30,31].

Incorporating several local directional searches into the GA process, Birch offers four variants of computationally efficient GAs in this study. The foundation of these local searches is a technique that integrates the methods of derivative-free directional search (DFDS), the Newton-Rampson method (NR), and the method of steepest descent (SD). The improvement of these suggested approaches is shown by a Monte Carlo reenactment research utilizing a split-plot configuration, utilizing benchmark works, for example, a low-layered capability versus a high-layered capability and a to some degree lopsided capability versus an exceptionally lopsided capability. To additionally exhibit how these recommended approaches beat both the traditional GA and the strategy remembered for the Plan Master measurable program, a certifiable circumstance relating to the multi-reaction improvement issue is likewise utilized. Integrating a neighborhood coordinated search into the GA cycle may often increase the accuracy and computing efficiency of the GA, as shown by our findings [32,33].

Molecular biologists have found success in using clustering algorithms to analyse gene expression data in the last few years. Now that microarray technology has advanced, scientists can see how hundreds of genes are expressed in response to different stimuli or cellular processes all at once. Genes are grouped according to how similar their expression patterns are using clustering techniques. This is how genes that have a common function are located. New productive and powerful bunching calculations are expected to manage the ever-increasing amounts of biological data generated by molecular biology laboratories, which is fueled by the exponential growth of data produced by cutting-edge technologies like microarray.

Due to its great computing efficiency, K-means [34,35] is among the most prominent approaches employed in gene expression data analysis among the numerous clustering algorithms. It is known, however, that K-means may converge to a local optimum, and that the initialization procedure, which produces the first clustering at random, affects the output. Put simply, it is possible for several K-means runs on the same input data to get distinct results. Genetic algorithms [36,37] have been suggested by several academics as a method for grouping. The main concept is to model the process of natural selection and pass on improved solutions over successive generations. While K-means may converge to a local optimum, these genetic algorithms don't care about the initialization and will always find the global optimum. However, computational expense is a common issue with these techniques, limiting their practical use in areas like gene expression data processing [38,39].

Data Mining Examples with Genetic Algorithm Method

Their research demonstrated that a few of fuzzy if-then rules may be used to pick several continuous pattern categorization issues. The first step of this method is to define the population of potential rules using data mining rule assessment metrics; the second step is to pick the rules using multi-objective genetic algorithms. We used a multi-objective evolutionary algorithm to narrow down the pool of candidate rules to a manageable number of fuzzy if-then rules based on numerical data. This technique integrates data mining support and confidence metrics into the rule population. Evolutionary algorithms are designed to discover the number of

non-dominant sets of rules. Previous research has shown that using computer simulations may enhance the efficiency of procedures when using fuzzy rule selection approaches [40,41].

A technique was described for categorising students with the purpose of predicting their final grade using attributes taken from recorded data in an education web-based system. A far better classification performance is achieved when many classifiers are used together. An optimisation of the prediction accuracy was achieved by utilising a genetic algorithm to weight the features and vectors; this achieved a significant improvement over raw classification. Feature weighting outperforms feature selection alone when the number of features is small, as was shown further [42,43].

The significance of using mining algorithms to sift through massive amounts of data in search of relevant outcomes. Observations indicate that when the search space contains many items, a complete search becomes prohibitively difficult. When this occurs, the significance of a thorough search becomes clear. The usefulness of genetic algorithms in data mining was examined in this work, along with their applicability to other domains, and the function of algorithms within data mining methods was defined and described explicitly [44,45].

Knowledge discovery and data mining using various genetic algorithms. One of the key arguments in favour of using genetic algorithm techniques in data mining applications is that they have certain desirable traits while removing some of the negative ones. For example, using genetic algorithms ensures reliability even when dealing with noisy data, and they have introduced new solutions like the ability to interpret data without any prior knowledge. Interactive algorithms are a potential hotspot for data mining-related research and new information discovery [45]. It was shown how genetic algorithms work and describes the features of these algorithms, which are widely employed in data mining. The purpose of this research was to use genetic algorithm approach to identify the most reliable company personnel database [46].

Feature selection of SNP association working data was conducted by mining utilising a genetic algorithm approach. Thousands upon thousands of single nucleotide polymorphisms (SNPs) are produced by genomic investigations, constituting massive amounts of data. Analysis of SNPs allows for the determination of correlations between phenotypic and genotypic data. Data mining and genetic algorithm technology have been used to provide a method for forecasting the efficacy of drugs. To choose useful genes, we use a worldwide hunt instrument, a weighted choice tree, a choice tree-based covering, a connection based heuristic, and we look for overlapping feature sets [47].

The goal of proposing this hybrid decision tree/genetic algorithm technique was to find rules for categorization. A key component of this combined approach is based on the data mining principle of minor disjuncts. In contrast to the traditional decision-tree method, which is used to generate rules for big disjunct cases, the suggested model incorporates two genetic algorithms (GA) that are tailored to find rules for tiny disjunct examples. Results from tests of the proposed algorithms on 22 real-world datasets are given in [48].

We surveyed data miners to find out how genetic algorithms fit into the process. This paper reviewed the usage of genetic algorithm approach in the domains of feature extraction, feature selection, classification, and clustering. They looked at evolutionary algorithms on their own and how they worked with other algorithms, such as decision trees and artificial neural networks [49].

When it comes to criminology data mining, there are two basic areas: crime control and crime suppression. An algorithm for crime trend analysis that clusters people according to their profiles and uses a novel distance metric to compare all individuals. It suggests a query-based interface that police may use as a crime analysis tool and builds on top of the e-governance projects already in place in India. This is used for the purpose of mining the massive crime database kept by the National Crime Record Bureau (NCRB) for valuable information and identifying problem areas using clustering and other related crime data mining methods. Crime statistics in India show that the suggested interface works well [50].

A Visual Interactive Malaysia Crime News Retrieval System that was developed and described together with the methodology, user studies, intended features, system design, and future plans. They aimed to build crime-based events, study how they can improve classification and clustering, create an interactive system to retrieve crime news, find effective ways to visualise crime news, combine all of that into a robust system, and finally, evaluate

how well the system works and how usable it is. This study will help us understand how people in Malaysia use crime data, and the system they built with visualisation features will help us fight crime [51].

Cluster analysis's use in accounting, and more specifically, audit discrepancy identification. The researcher is doing this investigation to learn more about audit fraud filtering automation via clustering technology. To save time and energy while reviewing group life insurance claims, he used cluster analysis. Missing value and clustering method data mining for crime pattern prediction and accelerated investigation [52].

In order to examine the city's crime statistics, the police department used data mining protocols. Someday soon, this data mining may help reduce or perhaps eliminate crime altogether. Using various data mining methods, we will create a crime analysis tool tailored to the Indian environment. This tool will aid law enforcement in their fight against crime. The suggested technology makes it simple and affordable for agencies to clean, characterise, and analyse crime data in order to find trends and patterns that may be used for the better [53].

An analysis and detection system for wrongdoing and criminal information using choice tree calculations for order and straightforward k-implies calculation for bunching. As a whole, the article is useful for experts in the fields of pattern recognition, trend analysis, forecasting, criminal network mapping, suspect identification, and creating connections and explanations.

Perhaps auditing might benefit from clustering technology. Automating the process of fraud screening may greatly benefit ongoing audits. Through this research, they want to shed light on cluster analysis' potential as a fresh approach to anomaly identification in the wire transmission system. Data mining and the development of a simulation model were the means by which K. Zakir Hussain et al. [11] sought to incorporate human experience into computational models [54].

Proposed System

Building Design Once the literature research is complete, a user-friendly, open-source data mining application should be used for easier analysis. A criminal dataset is analysed in this case by using the k-means clustering method via the Matlab programme. Following is the process:

1. First we take cybercrime dataset
2. data collection - Create a new dataset with attributes according to the analysis that has to be done after filtering the existing dataset according to requirements.
3. Data preprocessing
4. Feature selection using CMC
5. Classification-Perform k means clustering on resultant dataset formed
6. Prediction
7. visualizing and analyzing the result

Conclusion and Future Scope

A review of literature on data mining algorithms for cybercrime data analysis has been discussed. The evidence shows that there is an increase of cybercrime from 2011 to 2024. The results make it easy to see how crime has changed over the years, which can be used to create prevention strategies. Based on these encouraging findings, we think crime data mining will be an important for improving intelligence and criminal analysis in the future. Patterns of criminal activity may be an important factor for intelligent field investigation. Additionally, we can able to do analyses on a wide variety of datasets, including those pertaining to crime surveys, poverty crime, the efficacy of help, and many more.

Fig. 1: Cybercrime statistic 2011 to 2024

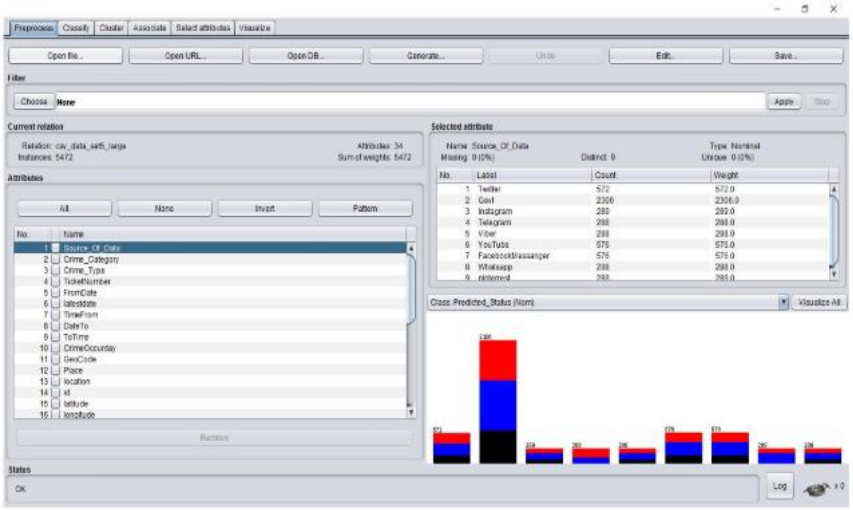

Fig. 2: Crime Data Sets Attributes

Based on the findings of the results, the accuracy of information obtained from the internet via social media platforms such as Facebook is 79.54% at the minimum accuracy and 99.5% at the maximum accuracy. When compared to other classification algorithms, our model with all key features shows the highest accuracy rate for the cybercrime dataset.

References

[1] P. Berkhin, Survey of clustering data mining techniques, In: Accrue Software, 2003.

[2] W. Li, Modified k-means clustering algorithm, IEEE Congress on Image and Signal Processing, pp. 616- 621, 2006.

[3] D.T Pham, S. Otri, A. Afifty, M. Mahmuddin, and H. Al-Jabbouli, Data clustering using the Bees algorithm, proceedings of 40th CRIP International Manufacturing Systems Seminar, 2006.

[4] J. Han, and M. Kamber, Data mining: concepts and techniques, 2nd Edition, Morgan Kaufmann Publisher, 2001.

[5] S. Joshi, and B. Nigam, Categorizing the document using multi class classification in data mining, International Conference on Computational Intelligence and Communication Systems, 2011.

[6] T. Phyu, Survey of classification techniques in data mining, Proceedings of the International Multi Conference of Engineers and Computer Scientists Vol. IIMECS 2009, March 18 - 20, 2009, Hong Kong.

[7] S.B. Kim, H.C. Rim, D.S. Yook, and H.S. Lim, Effective Methods for Improving Naïve Bayes Text Classifiers, In Proceeding of the 7th Pacific Rim International Conference on Artificial Intelligence, Vol.2417, 2002.

[8] S. Sindhiya, and S. Gunasundari, A survey on Genetic algorithm based feature selection for disease diagnosis system, IEEE International Conference on Computer Communication and Systems(ICCCS), Feb 20- 21, 2014.

[9] P. Gera, and R. Vohra, Predicting Future Trends in City Crime Using Linear Regression, IJCSMS (International Journal of Computer Science & Management Studies) Vol. 14, Issue 07Publishing Month: July 2014.

[10] L. Ding PerpSearch: an integrated crime detection system,‖ 2009 IEEE 161-163 ISI 2009, June 8-11, 2009, Richardson, TX, USA.

[11] K. Bogahawatte, and S. Adikari, Intelligent criminal identification system, IEEE 2013 The 8th International Conference on Computer Science & Education (ICCSE 2013) April 26-28, 2013. Colombo, Sri Lanka.

[12] A. Babakura, N. Sulaiman, and M. Yusuf, Improved method of calssification algorithms for crime prediction, International Symposium on Biometrics and Security Technologies (ISBAST) IEEE 2014.
[13] S. Sathyadevan, and S. Gangadharan, Crime analysis and prediction using data mining, IEEE 2014.
[14] J. Agarwal, R. Nagpal, and R. Sehgal, Crime analysis using k-means clustering, International Journal of Computer Applications, Vol. 83 – No4, December 2013.
[15] J. Han, and M. Kamber, Data mining: concepts and techniques, Jim Gray, Series Editor Morgan Kaufmann Publishers, August 2000.
[16] Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Boston: AddisonWesley.
[17] Davis, L. (1991). Handbook of Genetic Algorithm. Von Nostrand Reinhold, Newyork.
[18] Navatha, S., et al. "Multitask Learning Architecture For Vehicle Over Speed As Traffic Violations Detection And Automated Safety Violation Fine Ticketing Using Convolution Neural Network And Yolo V4 Techniques." *Chinese Journal of Computational Mechanics* 5 (2023): 431-435.
[19] Bryant, Kylie (2000). Genetic Algorithms and the Traveling Salesman Problem, in Proceedings of 1st GNT Regional Conference on Mathematics, Statistics and Applications.
[20] Coello, Carlos (2000). An updated survey of GA-based multiobjective optimization techniques, ACM Computing Surveys, vol. 32, no. 2, pp. 109- 143.
[21] Hanne, Thomas (2000). Global multiobjective optimization using evolutionary algorithms. Journal of Heuristics, vol. 6, no. 3, pp. 347-360.
[22] Madureira, A., Ramos, C., & Silva, S. Carmo, (2002). A Coordination Mechanism for Real World Scheduling Problems using Genetic Algorithms. Evolutionary Computation, in Proceedings of the 2002 CEC, vol. 1, pp. 175 –180.
[23] Molga, M. & Smutnicki, C. (2005). Test functions for optimization needs, in Proceedings of 4th Conference on Genetic Algorithms.
[24] Omar, M., Baharum, A., & Hasan, Y. Abu (2006). A Job-Shop Scheduling Problem (JSSP) Using Genetic Algorithm ,in Proceedings of 2nd IMT-GT Regional Conference on Mathematics, Statistics
[25] Inazawa, H. & Kitakaze, K. (2006). Locus-Shift Operator for Function Optimization in Genetic Algorithms. Complex Systems Publications, Inc.
[26] Dehuri, S. et al. (2008). Application of elitist multi-objective genetic algorithm for classification rule generation, Applied Soft Computing, pp. 477–487.
[27] Snehal Kamalapur (2006).Efficient CPU Scheduling: A Genetic Algorithm based Approach. IEEE, pp.206- 207
[28] Sivanandam, S. N. & Deepa, S. N. (2008). Introduction to Genetic Algorithms. Springer.
[29] H. Nazif(2009). A Genetic Algorithm on Single Machine Scheduling Problem toMinimise Total Weighted Completion Time. European Journal of Scientific Research,Vol.35 No.3, pp.444-452
[30] Dr. Rakesh Chawla(2010). Genetic Algorithm approach to Operating system process scheduling problem. International Journal of ngineering Science and Technology, pp. 4247-4252
[31] S. Ramya[2010] "Window Constrained Scheduling of Processes in Real Time CPU Using Multi Objective Genetic Algorithm" International Journal of Computer Applications Volume 1, pp.86-90.
[32] Birch, J. B. & Wan, W. (2009). An Improved Genetic Algorithm Using a Directional Search. Tech report presented at Virginia Polytechnic Institute and State University, Blacksburg.
[33] Shamir R, Sharan R: approaches to clustering gene expression data. In Current Topics in Computational Biology Edited by: Jiang T, Smith T, Y. Xu and Zhang MQ. MIT press; 2001.
[34] Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: Systematic determination of genetic network architecture. Nat Genet 1999, 22:281-285.
[35] Bhuyan JN, Raghavan VV, Elayavalli VK: Genetic algorithm for clustering with an ordered representation: ; San Mateo, CA, USA. ; 1991.
[36] Hall LO, B. OI, Bezdek JC: Clustering with a genetically optimized approach. IEEE Trans on Evolutionary Computation 1999, 3:103-112.
[37] Maulik U, Bandyopadhyay S: Genetic algorithm based clustering technique. Pattern Recognition 2000:1455-1465.
[38] Jones D, Beltramo M: partitioning problems with genetic algorithms: ; San Mateo, CA, USA. ; 1991.
[39] Reddy, Ch Subba, T. L. Yookesh, and E. Boopathi Kumar. "A Study On Convergence Analysis Of Runge-Kutta Fehlberg Method To Solve Fuzzy Delay Differential Equations." *Journal of Algebraic Statistics* 13.2 (2022): 2832-2838.
[40] Lu Y, Lu S, Fotouhi F, Deng Y, Brown S: FGKA: A Fast Genetic Kmeans Algorithm: March 2004. 2004.
[41] Lu Y, Lu S, Fotouhi F, Deng Y, Brown S: Fast genetic K-means algorithm and its application in gene expression data analysis. Detroit, Wayne State University; 2003.
[42] Kumar, E. Boopathi, et al. "Predicting the Fake Review of Products Using Graph Recurrent Neural Network." The International Journal of Interdisciplinary Organizational Studies ISSN: 2324-7649 (Print), ISSN: 2324-7657 (Online) Volume 18, Issue 2, July-December, 2023.
[43] De Hoon MJ, Imoto S, Nolan J, Miyano S: Open source clustering software. Bioinformatics 2004, 20:1453-1454.

[44] Mewes HW, Frishman D, Gruber C, Geier B, Haase D, Kaps A, Lemcke K, Mannhaupt G, Pfeiffer F, Schuller C, Stocker S, Weil B: MIPS: a database for genomes and protein sequences. Nucleic Acids Res 2000, 28:37-40.
[45] J. Bobadilla.Generalization of recommender systems: Collaborative filtering extended to groups of users and restricted to groups of items, Expert Systems with Applications (2012)
[46] Y.-C. Chou l. Modeling behavior sequence for personalized fund recommendation with graphical deep collaborative filtering, Expert Systems with Applications (2022).
[47] P. Fang, Data analytics-enable production visibility for Cyber-Physical Production Systems, Journal of Manufacturing Systems (2020).
[48] A.M. Ikotun. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data, Information Sciences (2023)
[49] W. Li. Performance benchmark on semantic web repositories for spatially explicit knowledge graph applications, Computers, Environment and Urban Systems (2022).
[50] X. Li . Exploiting knowledge graphs in industrial products and services: A survey of key aspects, challenges, and future perspectives, Computers in Industry (2021)
[51] X. Li. Achieving cognitive mass personalization via the self-X cognitive manufacturing network: An industrial knowledge graph- and graph embedding-enabled pathway,Engineering (2023).
[52] S. Liu. A blockchain-based interactive approach between digital twin-based manufacturing systems, Computers & Industrial Engineering, (2023).
[53] Y. Liu. Hierarchical independence thresholding for learning bayesian network classifiers, Knowledge-Based Systems,(2021).
[54] S. Ma . Data-driven cleaner production strategy for energy-intensive manufacturing industries: Case studies from Southern and Northern China,Advanced Engineering Informatics,(2022).