CRIME ANALYSIS AND PREDICTION USING MACHINE LEARNING

Syed Saood Hasan[1], Khaja Moinuddin Khan[2], Mohammed Moiduddin[3], Dr. Syed Asadullah Hussaini[4]

[1,2]B. E Student, Department of CSE, ISL College of Engineering, India.

[3]Associate Professor, Department of CSE, ISL College of Engineering, Hyderabad, India.

Abstract:

To be better prepared to respond to criminal activity, it is important to understand patterns in crime. In our project, we analyze crime data from the city.

At the outset, the task is to predict which category of crime is most likely to occur given a time and place in city. The use of AI and machine learning to detect crime via sound or cameras currently exists, is proven to work, and expected to continue to expand.

The use of AI/ML in predicting crimes or an individual's likelihood for committing a crime has promise but is still more of an unknown. The biggest challenge will probably be "proving" to politicians that it works. When a system is designed to stop something from happening, it is difficult to prove the negative. Companies that are directly involved in providing governments with AI tools to monitor areas or predict crime will likely benefit from a positive feedback loop. Improvements in crime prevention technology will likely spur increased total spending on this technology. We also attempt to make our classification task more meaningful by merging multiple classes into larger classes. Finally, we report and reflect on our results with different classifiers, and dwell on avenues for future work.

Introduction

Crime is a significant aspect of public safety and protection, and gaining a comprehensive understanding of crime has several advantages. It enables law enforcement authorities to implement focused and tactful strategies to reduce crime. Additionally, it encourages both citizens and authorities to collaborate more effectively in order to establish safe and thriving neighborhood environments.

The emergence of the Big Data age and the accessibility of rapid and effective algorithms for data analysis have led to a thriving and expanding area of study focused on comprehending crime trends via data. The parameters for our algorithms consist of time (hour, day, month, and year), location (latitude and longitude), and crime category.

- Act 379 - Robbery
- Act 13 - Gambling
- Act 279 - Accident
- Act 323 - Violence
- Act 302 - Murder
- Act 363 - Kidnapping

The output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests.

We also perform multiple classification tasks – we first try to predict which of 6 classes of crimes are likely to have occurred, and later try to differentiate between violent and non-violent crimes.

Methodology

Machine learning

Machine learning is the process of automatically identifying significant patterns in data. Over the last several decades, it has become a widely used technique in practically every work that involves extracting information from massive data sets.We are encompassed by a technology that utilizes machine learning: search engines acquire the ability to provide us with optimal results (while displaying profitable advertisements), anti-spam software acquires the ability to sift through our email messages, and credit card transactions are safeguarded by software that acquires the ability to identify fraudulent activities. Digital cameras are becoming capable of facial recognition, while intelligent personal assistant programs on smartphones are developing the ability to understand and respond to spoken requests. Automobiles are outfitted with collision avoidance systems that are constructed using machine learning techniques.

Machine learning is extensively used in scientific domains such as bioinformatics, medicine, and astronomy. A distinguishing characteristic of all these applications is that, unlike conventional computer uses, a human programmer is unable to offer a clear and explicit specification for how these activities should be completed owing to the intricate nature of the patterns that need to be identified. Like intelligent creatures, we acquire or improve many of our talents via experiential learning rather than relying only on explicit instructions. Machine learning tools focus on equipping programs with the capacity to acquire knowledge and adjust accordingly.

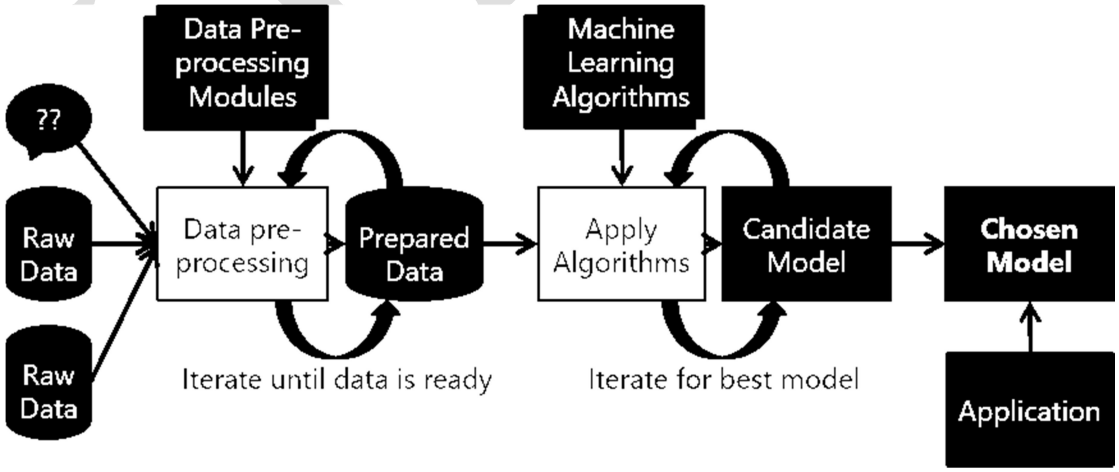

Fig 1-Machine learning process

The inputs to our algorithms are time (hour, day, month, year), place (latitude and longitude), class of crime

- Act 379-Robbery
- Act 13-Gambling
- Act 279-Accident
- Act 323-Violence
- Act 302-Murder
- Act 363-Kidnapping

The output is the class of crime that is likely to have occurred. We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests.

After the preprocessing described in the previous sections, we had three different classifications problems to solve, which we proceeded to attack with an assortment of classification algorithms. The following are the algorithms which we are using:

- KNN(K- Nearest neighbors)
- Decision Tree
- Random Forests

Analysis and Design

Activity diagram

The activity diagram is a graphical representation for representing the flow of interaction within specific scenatios. It is similar to a flowchart in which various activities that can be performed in the system are represented.

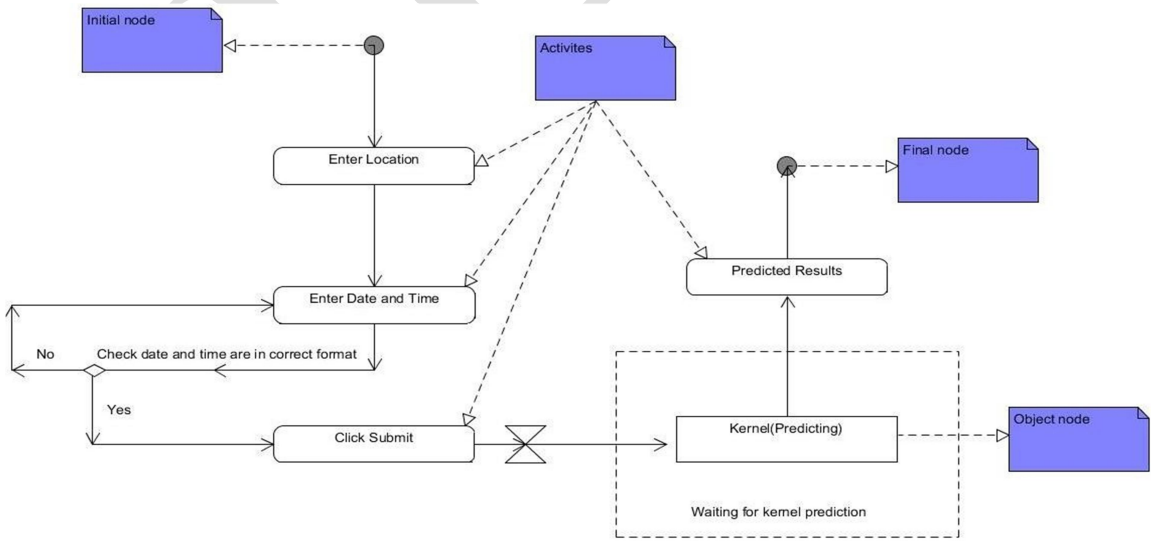

Fig 2-Activity diagram

System architecture

The system architectural design is the design process for identifying the subsystems making up the system and framework for subsystem control and communication. The goal of the architectural design is to establish the overall structure of software system.

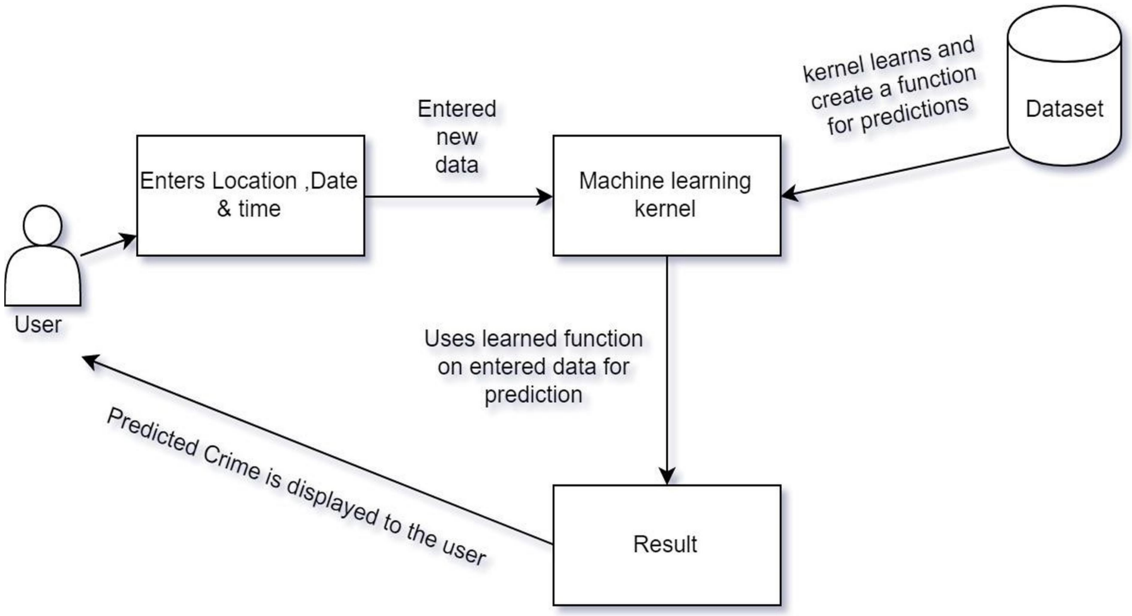

Fig 3-System architecture

Implementation

The implementation of the project is done with the help of python language. To be particular, for the purpose of machine learning Anaconda is being used.

Anaconda is one of several Python distributions. Anaconda is a new distribution of the Python. It was formerly known as Continuum Analytics. Anaconda has more than 100 new packages. Anaconda is used for scientific computing, data science, statistical analysis, and machine learning.

On Python technology, we found out Anaconda to be easier. Since it helps with the following problems:

- Installing Python on multiple platforms.
- Separating out different environments.
- Dealing with not having correct privileges.
- Getting up and running with specific packages and libraries.

KNN (K-Nearest neighbors)

A powerful classification algorithm used in pattern recognition K nearest neighbors stores all available cases and classifies new cases based on a similarity measure (e.g. distance function).One of the top data mining algorithms used today. A non-parametric lazy learning algorithm (An Instance based Learning method).

KNN: Classification Approach

- An object (a new instance) is classified by a majority votes for its neighbor classes.
- The object is assigned to the most common class amongst its K nearest neighbors.(measured by distance function)

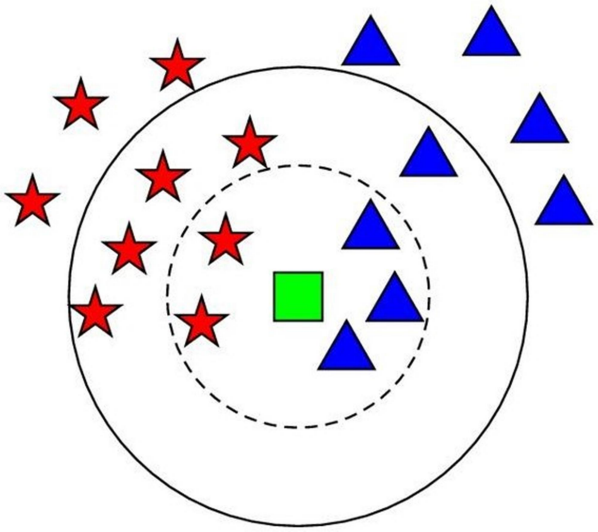

Fig 4 Principle diagram of KNN

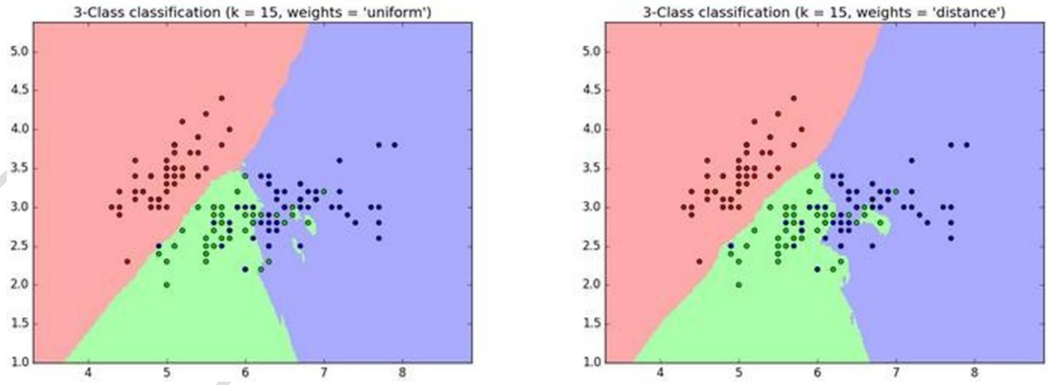

Fig 4.1 Shows graphical representation of KNN

As the name says all about it, it is a tree which helps us by assisting us in decision-making. Used for both classification and regression, it is a very basic and important predictive learning algorithm.

- It is different from others because it works intuitively i.e., taking decisions one-by-one.
- Non-parametric: Fast and efficient.

It consists of nodes which have parent-child relationships:

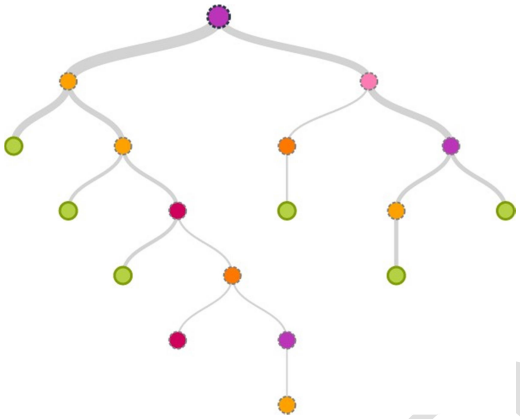

Fig 4.2 Decision tree

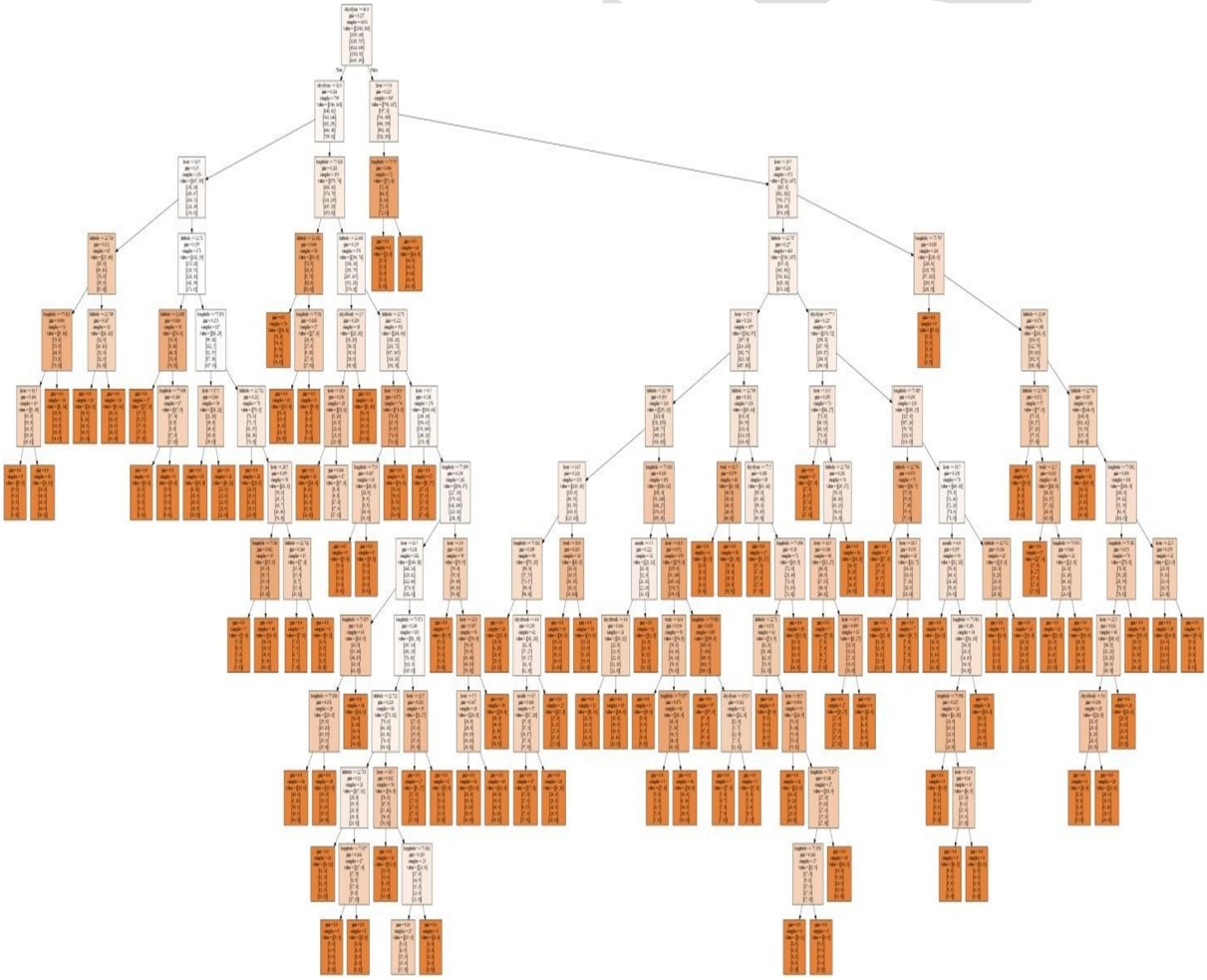

Fig 4.3 Decision Tree

Testing

Test Case ID	Test Name	Test Description	Steps	Executed result	Actual result	Test case statement
01	Check for correct entered numeric values and date and time.	The entered values are in correct format.	1. Enter details in fields. 2. Click submit.	If format is correct details are sent to kernel successfully.	As expected.	Pass
02	Check for correct entered time.	The entered values are in correct format.	1. Enter details in fields. 2. Click submit.	If format is correct details are sent to kernel successfully	As expected.	Pass
03	Check for correct entered location	The entered values are correct.	1. Enter details in fields. 2. Click submit.	If format is correct details are sent to kernel successfully	As expected	Pass
04	Predicted Result	Output is displayed		If kernel predicts successfully output is then showed to the screen	As expected	Pass
05	Analysis Button	Data visualization is displayed.	1.Click Analysis	Shows the overall analysis on screen 3	As expected	Pass

Conclusion and future scope

The initial problem of classifying 6 different crime categories was a challenging multi-class classification problem, and there was not enough predictability in our initial data-set to obtain very high accuracy on it. We found that a more meaningful approach was to collapse the crime categories into fewer, larger groups, in order

to find structure in the data. We got high accuracy and precision on Prediction. However, the Violent/Non-violent crime classification did not yield remarkable results with the same classifiers – this was a significantly harder classification problem. Thus, collapsing crime categories is not an obvious task and requires careful choice and consideration.

Possible avenues through which to extend this work include time-series modeling of the data to understand temporal correlations in it, which can then be used to predict surges in different categories of crime. It would also be interesting to explore relationships between surges in different categories of crimes – for example, it could be the case that two or more classes of crimes surge and sink together, which would be an interesting relationship to uncover. Other areas to work on include implementing a more accurate multi-class classifier, and exploring better ways to visualize our results.

Future Scope

The goal of any society shouldn't be to just catch criminals but to prevent crimes from happening in the first place

Predicting Future Crime Spots: By using historical data and observing where recent crimes took place we can predict where future crimes will likely happen. For example a rash of burglaries in one area could correlate with more burglaries in surrounding areas in the near future. System highlights possible hotspots on a map the police should consider patrolling more heavily

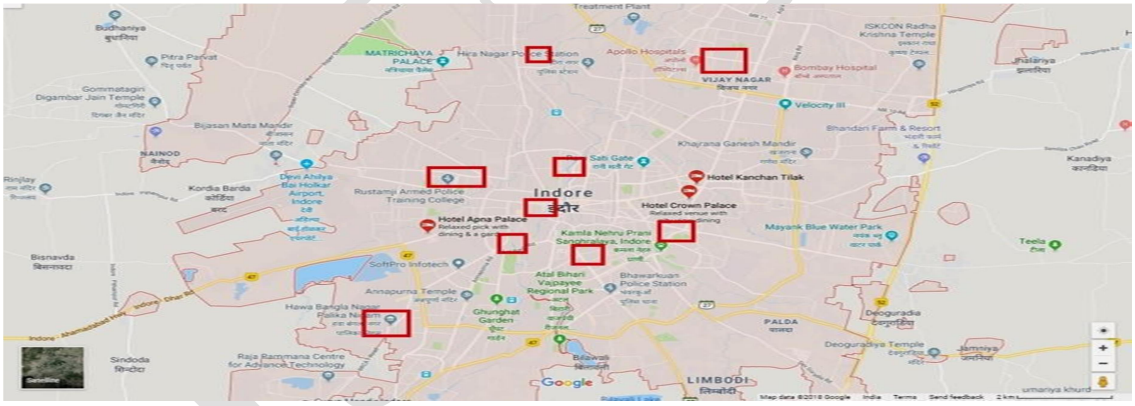

Fig 5 Predicting Surges

References

[1] Bogomolov, Andrey and Lepri, Bruno and Staiano, Jacopo and Oliver, Nuria and Pianesi, Fabio and Pentland, Alex.2014. Once upon a crime: Towards crime prediction from demographics and mobile data, Proceedings of the 16th International Conference on Multimodal Interaction.

[2] Yu, Chung-Hsien and Ward, Max W and Morabito, Melissa and Ding, Wei.2011. Crime forecasting using data mining techniques, pages 779-786, IEEE 11th International Conference on Data Mining Workshops (ICDMW)

[3] Kianmehr, Keivan and Alhajj, Reda. 2008. Effectiveness of support vector machine for crime hot-spots

prediction, pages 433-458, Applied Artificial Intelligence, volume 22, number 5.

[4] Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay," *Routing Performance Analysis of Infrastructure-less Wireless Networks with Intermediate Bottleneck Nodes*", International Journal of Intelligent Systems and Applications in Engineering, ISSN no: 2147-6799 IJISAE,Vol 12 issue 3, 2024, Nov 2023

[5] Md. Zainlabuddin, "*Wearable sensor-based edge computing framework for cardiac arrhythmia detection and acute stroke prediction*", Journal of Sensor, Volume2023.

[6] Md. Zainlabuddin, "*Security Enhancement in Data Propagation for Wireless Network*", Journal of Sensor, ISSN: 2237-0722 Vol. 11 No. 4 (2021).

[7] Dr MD Zainlabuddin, "*CLUSTER BASED MOBILITY MANAGEMENT ALGORITHMS FOR WIRELESS MESH NETWORKS*", Journal of Research Administration, ISSN:1539-1590 | E-ISSN:2573-7104 , Vol. 5 No. 2, (2023)

[8] Vaishnavi Lakadaram, " Content Management of Website Using Full Stack Technologies", Industrial Engineering Journal, ISSN: 0970-2555 Volume 15 Issue 11 October 2022

[9] Dr. Mohammed Abdul Bari,Arul Raj Natraj Rajgopal, Dr.P. Swetha ," *Analysing AWSDevOps CI/CD Serverless Pipeline Lambda Function's Throughput in Relation to Other Solution*", International Journal of Intelligent Systems and Applications in Engineering , JISAE, ISSN:2147-6799, Nov 2023, 12(4s), 519–526

[10] Ijteba Sultana, Mohd Abdul Bari and Sanjay," *Impact of Intermediate per Nodes on the QoS Provision in Wireless Infrastructure less Networks*", Journal of Physics: Conference Series, Conf. Ser. 1998 012029 , CONSILIO Aug 2021

[11] M.A.Bari, Sunjay Kalkal, Shahanawaj Ahamad," *A Comparative Study and Performance Analysis of Routing Algorithms*", in 3rd International Conference ICCIDM, Springer - 978- 981-10-3874-7_3 Dec (2016)

[12] Mohammed Rahmat Ali,: BIOMETRIC: AN e-AUTHENTICATION SYSTEM TRENDS AND FUTURE APLLICATION", International Journal of Scientific Research in Engineering (IJSRE), Volume1, Issue 7, July 2017

[13] Mohammed Rahmat Ali,: BYOD.... A systematic approach for analyzing and visualizing the type of data and information breaches with cyber security", NEUROQUANTOLOGY, Volume20, Issue 15, November 2022

[14] Mohammed Rahmat Ali, Computer Forensics -An Introduction of New Face to the Digital World, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169-453 – 456, Volume: 5 Issue: 7

[15] Mohammed Rahmat Ali, Digital Forensics and Artificial Intelligence ...A Study, International Journal of Innovative Science and Research Technology, ISSN:2456-2165, Volume: 5 Issue:12.

[16] Mohammed Rahmat Ali, Usage of Technology in Small and Medium Scale Business, International Journal of Advanced Research in Science & Technology (IJARST), ISSN:2581-9429, Volume: 7 Issue:1, July 2020.

[17] Mohammed Rahmat Ali, Internet of Things (IOT) Basics - An Introduction to the New Digital World, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169-32-36, Volume: 5 Issue: 10

[18] Mohammed Rahmat Ali, Internet of things (IOT) and information retrieval: an introduction, International Journal of Engineering and Innovative Technology (IJEIT), ISSN: 2277-3754, Volume: 7 Issue: 4, October 2017.

[19] Mohammed Rahmat Ali, How Internet of Things (IOT) Will Affect the Future - A Study, International Journal on Future Revolution in Computer Science & Communication Engineering, ISSN: 2454-424874 – 77, Volume: 3 Issue: 10, October 2017.

[20] Mohammed Rahmat Ali, ECO Friendly Advancements in computer Science Engineering and Technology, International Journal on Scientific Research in Engineering(IJSRE), Volume: 1 Issue: 1, January 2017

[21] Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay, "*Routing Quality of Service for Multipath Manets, International Journal of Intelligent Systems and Applications in Engineering"*, JISAE, ISSN:2147-6799, 2024, 12(5s), 08–16;

[22] Mr. Pathan Ahmed Khan, Dr. M.A Bari,: Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges", International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page 43-46

[23] Shahanawaj Ahamad, Mohammed Abdul Bari, Big Data Processing Model for Smart City Design: A Systematic Review ", VOL 2021: ISSUE 08 IS SN : 0011-9342 ;Design Engineering (Toronto) Elsevier SCI Oct : 021

[24] Syed Shehriyar Ali, Mohammed Sarfaraz Shaikh, Syed Safi Uddin, Dr. Mohammed Abdul Bari, "Saas Product Comparison and Reviews Using Nlp", Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022

[25] Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali," Smartphone Security and Protection Practices", International Journal of Engineering and Applied Computer Science (IJEACS) ; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021 (International Journal,U K) Pages 1-6

[26] .A.Bari& Shahanawaj Ahamad, "Managing Knowledge in Development of Agile Software", in International Journal of Advanced Computer Science & Applications (IJACSA), ISSN: 2156-5570, Vol: 2, No: 4, pp: 72-76, New York, U.S.A., April 2011

[27] Imreena Ali (Ph.D), Naila Fathima, Prof. P.V.Sudha ,"Deep Learning for Large-Scale Traffic-Sign Detection and Recognition", Journal of Chemical Health Risks, ISSN:2251-6727/ JCHR (2023) 13(3), 1238-1253

[28] Imreena, Mohammed Ahmed Hussain, Mohammed Waseem Akram" An Automatic Advisor for Refactoring Software Clones Based on Machine Learning", Mathematical Statistician and Engineering ApplicationsVol. 72 No. 1 (2023)

[29] Mrs Imreena Ali Rubeena,Qudsiya Fatima Fatimunisa "Pay as You Decrypt Using FEPOD Scheme and Blockchain", Mathematical Statistician and Engineering Applications: https://doi.org/10.17762/msea.v72i1.2369 Vol. 72 No. 1 (2023)

[30] Imreena Ali , Vishnuvardhan, B.Sudhakar," Proficient Caching Intended For Virtual Machines In Cloud Computing", International Journal Of Reviews On Recent Electronics And Computer Science , ISSN 2321-5461,IJRRECS/October 2013/Volume-1/Issue-6/1481-1486

[31] Heena Yasmin, A Systematic Approach for Authentic and Integrity of Dissemination Data in Networks by Using Secure DiDrip, INTERNATIONAL JOURNAL OF PROFESSIONAL ENGINEERING STUDIES, Volume VI /Issue 5 / SEP 2016

[32] Heena Yasmin, Cyber-Attack Detection in a Network, Mathematical Statistician and Engineering Applications, ISSN:2094-0343, Vol.72 No.1(2023)

[33] Heena Yasmin, Emerging Continuous Integration Continuous Delivery (CI/CD) For Small Teams, Mathematical Statistician and Engineering Applications, ISSN:2094-0343, Vol.72 No.1(2023)