Human Resource Data Analysis & predictionusing Decision Tree Algorithm and Random Forest

Durga S,Prathusha Perugu,Nidhi Sree D, Venkateswararao Podile ,Mahesh Manohar Bhanushali, Revathi R

[1]KL Business School, KLEF, KL University e-mail: sdurga1234@gmail.com

[2]Department of AI,ML Data Science, Gates Institute of Technology, Andrapradesh, India. e-mail: Prathusha2009@gmail.com

[3]Department of Commerce, Bhavans Vivekananda College of Sciene, Humanitites and Commerce,KL University. e-mail: nidhishree.devarakonda@gmail.com

[4]KL Business School, KL University e-mail: vraopodile@gmail.com

[5]Management Studies Operation Institute, VPM, Dr.V.N. Bedekar Institute of Management Studies, Thane, maheshbhanu87@gmail.com

[6]Department of Computer Science, Karpagam Academy of Highern Education, Coimbatore, Indiae-mail: revathilakshay@gmail.com

Abstract:

Due to the complexity and unpredictability of human behaviour in contrastto equipment or other physical assets, allocating human capital efficiently has always been a challenging task. Because of this issue, human resource analytics has lately gained prominence. Employee attrition is a fascinating data set to investigate. Apartfrom the reasons mentioned in the dataset, other factors, such as distance from work, further education, establishing one's own business, politics, or relocation, may occur. However, let us do some analysis and prediction using the dataset that has been provided to us. Before we begin, let us clarify why we are doing this analysis: discover why a small percentage of workers leave the company and their satisfaction. The prediction is used to determine which workers will depart from the business. Future research may concentrate on developing validated models of the suggested theoretical frameworks and assessing the impact of Evidence-Based Management methods on human resources and organizational performance.

DOI Number: 10.14704/nq.2022.20.9.NQ44558 Neuro Quantology 2022; 20(9):4803-4809

1 Introduction

Besides talent management and development, human resource (HR) professionals must identify ways to improve human capital management[1]. It has always been challenging to allocate human capital effectively because of the complexity and un- predictability of human behaviour compared to equipment or other physical assets. Human resource analytics have been more popular during the last several years[2]. When new management policies are developed, assessed, and implemented based on data and analytics, the "tried and true" method of directing human resource strat- egy based on experience, intuition, and guesswork is slipping away. An increasingnumber of organizations are aware of the benefits of human resource analytics, yetfew decisions are based on data or well-researched aim deliberations[3]. Most of the material on analytics is promotional, with little guidance on how to turn ideas into reality.[11-15]

This research article is organized as follows, Section I described with the complete Introduction, Motivation and Background of the research work, Section II gives thedeep analysis of the existing works of our domain of research has been carried out,and presented the research gaps and previous work performance, Section IV , novel idea is presented for the research gaps

identified and Findings, Section V, results are observed with the experimentation with the performance analysis, Section VI, gives the conclusion and Future work.

2 Literature Survey

In reality, the field of human resource data analytics has just lately begun to acquire momentum. The importance of "big data" has grown in recent years, which has contributed to this trend. It is common to hear the term "big data" in reference to data that is too enormous to be handled and processed by conventional database technologies[21-25]. Certain factors may also be considered, such as "intelligence," or the extent to which data is used to explain and anticipate behavior and results. An in-depth knowledge of the data being utilized and the context in which it was gathered is required before any analytical insights can be generated [26-29]. All ofthe information provided here will be utilized to establish the resources that will be required, as well as the final format of the analysis to be conducted. [7] Before using HR analytics, human resource professionals and management must first understand the significance of human capital in the success of their organizations [8-12].

The author's method was used to calculate employee attrition. Logistic regression is used in statistical analysis. This system takes use of a dataset that includes data on employee counts. Members of the organization leave every year[13-20].

The existing work gaps are handled in our proposed work by the novel methodol- ogy by using Decision Tree Algorithm and Random Forest is described in Section 3

3 Proposed Work

The Basic Method we used is how can a company find out about what was happening n their organization for this we have taken some predefined data from kaggle and we performed analysis in different ways. Firstly we need to know what are the different data attributes we have and knowing about it would help us in many ways. Then we need to refine the data and know if there are any empty cells and we need totry to eliminate them. Then we need to look at the extremely important attributesin the data-set and try to find out an interlink between them. Firstly, We can taken all the attributes present in the data and form a correlation between them and see how it forms. Secondly we can see how many employees left for the number of projects increased. Thirdly we can also look at the satisfaction levels of each and every employee by performing analysis. We can also predict using decision tree and their satisfactory levels of who are all the employees satisfied [21-29].

The model mainly contains six parts:

1. Data-set input
2. Analyzing the Data
3. Decision Tree
4. Random Forest

The model requires a specific type of the data-set which will contain the list ofthe Employees left. It'll also contain the data like the total number of projects done

, the satisfaction level, Promotion last 5 years and also their salary.

3.1 Decision Tree

To find out the decision from the data we used a concept called decision trees. Core algorithm for building decision trees called ID3 which employs a top-down, greedy search [30-40] through the space of possible branches with no backtracking. Using R programming

For performing analysis we will be going to use R tool. R is a programming tool which is majorly used to analyze data and also find out the statistical results such asmean, median etc from the data. We can plot the graphs using R programming.

4 Analyzing the data

In this section, we perform analysis and see weather we can use those graphs and infer something from it. We have taken this data from kaggle and performed analysis on it.The graphs and plots shown are only from it.

4.1 Correlation

Load all the Libraries and the dataset into R

using this command.data = read.csv("../input/HRcommasep.csv")

sapply(data,class)

We have to keep a check on the datasets which do not contain numerical values. We have to try and convert them into numerical form so that we can form the desired correlation.

corrplot(cor(as.matrix(tempData), method = " ", use = " ") ,is.corr = FALSE, type = "lower", order = "hclust", tl.col = "black", tl.srt = 45)

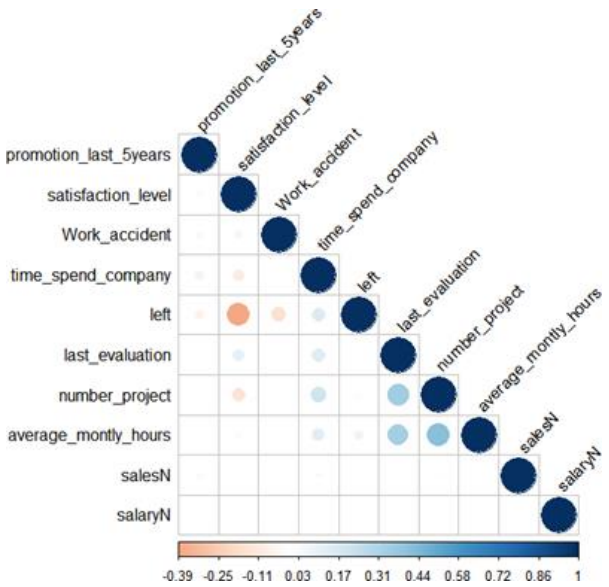

Fig. 1 Correlation

Figure 1 shows the correlation plot form the attributes we have.In the correlation plot, more darker and bigger the circle correlation is positive. In other case if lighter and bigger, it indicates those 2 variables are inversely correlated i.e if one increases other variable decreases. Coming to the employees plot, we can see lighter and bigger circle for satisfaction level and left.This implies more the satisfaction level, 'left' tends to zero i.e more satisfied employees does not quit.And also we can observe that employes' who are assigned with more number of project tend to spend more hours in office.

4.2 Barplots

We saw employee satisfaction is the major item which impact attrition in companies. In this section lets get more insight of this variable through Barplot. To achieve this we will seggregate satisfaction-level to high,good,average,low, lowest. And assign this to a new variable 'satisfaction'

If the Employees with good, high and highest satisfaction tend to stay in the company and the rest will tend to leave the organization. Lets see how projects haveeffected the employees left

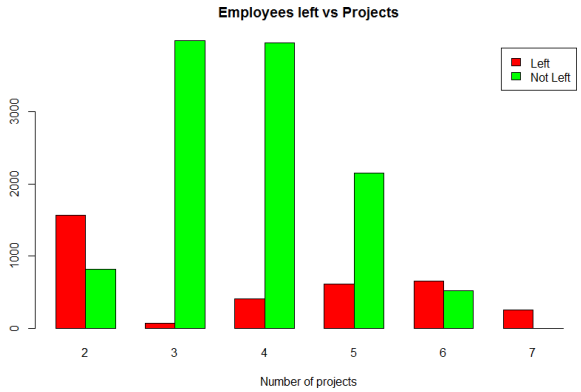

Fig. 2 Employees Left VSprojects

We can say that employees with very less projects and employees with too much projects tend to leave the organization. Employees with 3 to 4 projects in hand will often stay in the organization. So, it is better that the number of projects be 3 to 4.

5 Prediction

We have done a lot of analysis now let us look at the prediction models.

5.1 Decision Tree

This dataset is an ideal example of classification problem and also data is not continuous so decision tree suits best for this. First create a training and testing data set. Lets do a 75-25 partition of the data.Then create a fancy R plot with the partition decisionTreeModel <- train(left .,method=" ",data=)

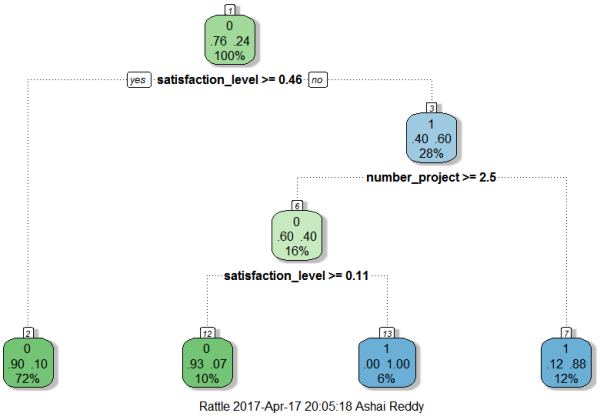

Fig. 3 Decision Tree Modelused for Prediction

A clear indication that satisfaction-level and number of project have a affect on employee staying or leaving the company.we have occured with 90 percent accuracyin decision tree model. This is comparatively good, the future work may include some other technique to improvise the accuracy.

5.2 Random Forest

Random forest is nothing but collection of decision trees. We has to take multiple decision trees and check the accuracy of the model. With this data we would get an accuracy of about 99percent which is a very high accuracy and also the results withrandom forest would be better. It is also evident from this that satisfaction level andnumber of projects have a clear effect on the employees leaving and staying in the organization.

6 Conclusion and Future Work

Satisfaction is a critical factor in determining whether an employee will remain with the business. We still need to analyze the parameter upon which the satisfaction level is dependent. Perhaps in a future edition, I'll change this script. Additionally, the quantity of initiatives has an effect. Employees allocated to three to four projects aremore likely to remain. Perhaps this has a direct correlation with work-life balance.

HR analytics solutions should not be used only for the sake of being trendy or competitive. Since human resource analytics have the ability to substantially improve both current and future performance of a company, it is important to make use of this technology.

Despite the recent and rapid increase in popularity of human resource analytics, it is prudent to maintain a healthy dose of skepticism about their efficacy. It is possible to assess the maturity of a business in a number of ways. Until recently, about 75% of human resource departments lacked even the most basic of safeguards to protect their employees. As a result, many companies will be forced to undertake a significant transition from their present form to one that successfully uses analytics.

References

1. Zhang, Yucheng, Shan Xu, Long Zhang, and Mengxi Yang. "Big data and human resource management research: An integrative review and new directions for future research." Journal of Business Research 133 (2021): 34-50.
2. Kshetri, Nir. "Evolving uses of artificial intelligence in human resource management in emerg- ing economies in the global South: some preliminary evidence." Management Research Review (2021).
3. Rahul, Kumar, Rohitash Kumar Banyal, Puneet Goswami, and Vijay Kumar. "Machine learning algorithms for big data analytics." In Computational Methods and Data Engineering, pp. 359-367. Springer, Singapore, 2021.
4. Sato, Shintaro, Tae-Ahn Kang, Ebe Daigo, Hirotaka Matsuoka, and Munehiko Harada. "Grad- uate employability and higher education's contributions to human resource development in sport business before and after COVID-19." Journal of Hospitality, Leisure, Sport Tourism Education 28 (2021): 100306.
5. Wei, Guohua, and Yi Jin. "Human resource management model based on three-layer BP neural network and machine learning." Journal of Intelligent Fuzzy Systems Preprint (2021): 1-12.
6. Vrontis, Demetris, Michael Christofi, Vijay Pereira, Shlomo Tarba, Anna Makrides, and Eleni Trichina. "Artificial intelligence, robotics, advanced technologies and human resource man- agement: a systematic review." The International Journal of

Human Resource Management (2021): 1-30.

7. Liu, Shui. "Human Resource Management of Internet Enterprises Based on Big Data Mobile Information System." Mobile Information Systems 2021 (2021).
8. Liu, Jintao. "Impact of enterprise human capital on technological innovation based on machine learning and SVM algorithm." Journal of Ambient Intelligence and Humanized Computing (2021): 1-13.
9. Hickman, Louis, Rachel Saef, Vincent Ng, Sang Eun Woo, Louis Tay, and Nigel Bosch. "De- veloping and evaluating language-based machine learning algorithms for inferring applicant personality in video interviews." Human Resource Management Journal (2021).
10. Liu, Yishu, Wenjie Zhang, Qi Zhang, and Monire Norouzi. "An optimized human resource management model for cloud-edge computing in the internet of things." Cluster Computing (2021): 1-13.
11. Rathore, M. S., Poongodi, M., Saurabh, P., Lilhore, U. K., Bourouis, S., Alhakami, W., ... & Hamdi, M. (2022). A novel trust-based security and privacy model for Internet of Vehicles using encryption and steganography. Computers and Electrical Engineering, 102, 108205.
12. Gupta, S., Iyer, S., Agarwal, G., Manoharan, P., Algarni, A. D., Aldehim, G., & Raahemifar, K. (2022). Efficient Prioritization and Processor Selection Schemes for HEFT Algorithm: A Makespan Optimizer for Task Scheduling in Cloud Environment. Electronics, 11(16), 2557.
13. Balyan, A. K., Ahuja, S., Lilhore, U. K., Sharma, S. K., Manoharan, P., Algarni, A. D., ... & Raahemifar, K. (2022). A Hybrid Intrusion Detection Model Using EGA-PSO and Improved Random Forest Method. Sensors, 22(16), 5986.
14. Poongodi, M., Bourouis, S., Ahmed, A. N., Vijayaragavan, M., Venkatesan, K. G. S., Alhakami, W., & Hamdi, M. (2022). A Novel Secured Multi-Access Edge Computing based VANET with Neuro fuzzy systems based Blockchain Framework. Computer Communications.
15. Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES. Malaysian Journal of Computer Science, 132-148.
16. Poongodi, M., Malviya, M., Hamdi, M., Vijayakumar, V., Mohammed, M. A., Rauf, H. T., & Al-Dhlan, K. A. (2022). 5G based Blockchain network for authentic and ethical keyword search engine. IET Commun., 16(5), 442-448.
17. Poongodi, M., Malviya, M., Kumar, C., Hamdi, M., Vijayakumar, V., Nebhen, J., & Alyamani, H. (2022). New York City taxi trip duration prediction using MLP and XGBoost. International Journal of System Assurance Engineering and Management, 13(1), 16-27.
18. Poongodi, M., Hamdi, M., & Wang, H. (2022). Image and audio caps: automated captioning of background sounds and images using deep learning. Multimedia Systems, 1-9.
19. Poongodi, M., Hamdi, M., Gao, J., & Rauf, H. T. (2021, December). A Novel Security Mechanism of 6G for IMD using Authentication and Key Agreement Scheme. In 2021 IEEE Globecom Workshops (GC Wkshps) (pp. 1-6). IEEE.
20. Ramesh, T. R., Vijayaragavan, M., Poongodi, M., Hamdi, M., Wang, H., & Bourouis, S. (2022). Peer-to-peer trust management in intelligent transportation system: An Aumann's agreement theorem based approach. ICT Express.
21. Hamdi, M., Bourouis, S., Rastislav, K., & Mohmed, F. (2022). Evaluation of Neuro Image for the Diagnosis of Alzhimer's Disease Using Deep Learning Neural Network. Frontiers in Public Health, 35.
22. Poongodi, M., Hamdi, M., Malviya, M., Sharma, A., Dhiman, G., & Vimal, S. (2022). Diagnosis and combating COVID-19 using wearable Oura smart ring with deep learning methods. Personal and ubiquitous computing, 26(1), 25-35.
23. Sahoo, S. K., Mudligiriyappa, N., Algethami, A. A., Manoharan, P., Hamdi, M., & Raahemifar, K. (2022). Intelligent Trust-Based Utility and Reusability Model: Enhanced Security Using Unmanned Aerial

Vehicles on Sensor Nodes. Applied Sciences, 12(3), 1317.

24. Muniyappan, A., Sundarappan, B., Manoharan, P., Hamdi, M., Raahemifar, K., Bourouis, S., & Varadarajan, V. (2022). Stability and numerical solutions of second wave mathematical modeling on covid-19 and omicron outbreak strategy of pandemic: Analytical and error analysis of approximate series solutions by using hpm. Mathematics, 10(3), 343.
25. Rawal, B. S., Manogaran, G., & Poongodi, M. (2022). Implementing and Leveraging Blockchain Programming.
26. Bourouis, S., Band, S. S., Mosavi, A., Agrawal, S., & Hamdi, M. (2022). Meta-Heuristic Algorithm-Tuned Neural Network for Breast Cancer Diagnosis Using Ultrasound Images. Frontiers in Oncology, 12, 834028.
27. Lilhore, U. K., Poongodi, M., Kaur, A., Simaiya, S., Algarni, A. D., Elmannai, H., ... & Hamdi, M. (2022). Hybrid Model for Detection of Cervical Cancer Using Causal Analysis and Machine Learning Techniques. Computational and Mathematical Methods in Medicine, 2022.
28. Dhiman, P., Kukreja, V., Manoharan, P., Kaur, A., Kamruzzaman, M. M., Dhaou, I. B., & Iwendi, C. (2022). A Novel Deep Learning Model for Detection of Severity Level of the Disease in Citrus Fruits. Electronics, 11(3), 495.
29. Dhanaraj, R. K., Ramakrishnan, V., Poongodi, M., Krishnasamy, L., Hamdi, M., Kotecha, K., & Vijayakumar, V. (2021). Random Forest Bagging and X-Means Clustered Antipattern Detection from SQL Query Log for Accessing Secure Mobile Data. Wireless Communications and Mobile Computing, 2021.
30. Maurya, S., Joseph, S., Asokan, A., Algethami, A. A., Hamdi, M., & Rauf, H. T. (2021). Federated transfer learning for authentication and privacy preservation using novel supportive twin delayed DDPG (S-TD3) algorithm for IIoT. Sensors, 21(23), 7793.
31. Poongodi, M., Nguyen, T. N., Hamdi, M., & Cengiz, K. (2021). Global cryptocurrency trend prediction using social media. Information Processing & Management, 58(6), 102708.
32. Poongodi, M., Sharma, A., Hamdi, M., Maode, M., & Chilamkurti, N. (2021). Smart healthcare in smart cities: wireless patient monitoring system using IoT. The Journal of Supercomputing, 77(11), 12230-12255.
33. Rawal, B. S., Manogaran, G., & Hamdi, M. (2021). Multi-Tier Stack of Block Chain with Proxy Re-Encryption Method Scheme on the Internet of Things Platform. ACM Transactions on Internet Technology (TOIT), 22(2), 1-20.
34. Poongodi, M., Malviya, M., Hamdi, M., Rauf, H. T., Kadry, S., & Thinnukool, O. (2021). The recent technologies to curb the second-wave of COVID-19 pandemic. Ieee Access, 9, 97906-97928.
35. Rawal, B. S., Manogaran, G., Singh, R., Poongodi, M., & Hamdi, M. (2021, June). Network augmentation by dynamically splitting the switching function in SDN. In 2021 IEEE International Conference on Communications Workshops (ICC Workshops) (pp. 1-6). IEEE.
36. Poongodi, M., Hamdi, M., Varadarajan, V., Rawal, B. S., & Maode, M. (2020, July). Building an authentic and ethical keyword search by applying decentralised (Blockchain) verification. In IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 746-753). IEEE.
37. Poongodi, M., Vijayakumar, V., Al-Turjman, F., Hamdi, M., & Ma, M. (2019). Intrusion prevention system for DDoS attack on VANET with reCAPTCHA controller using information based metrics. IEEE Access, 7, 158481-158491.
38. Poongodi, M., Hamdi, M., Sharma, A., Ma, M., & Singh, P. K. (2019). DDoS detection mechanism using trust-based evaluation system in VANET. IEEE Access, 7, 183532-183544.
39. Md Hossain, MM Kamruzzaman, Shuvo Sen, Mir Mohammad Azad, Mohammad Sarwar Hossain Mollah, Hexahedron Core with Sensor Based Photonic Crystal Fiber,2021
40. Md Nazirul Islam Sarker, Md Lamiur Raihan, Yang Peng, Tahmina Chumky, MM Kamruzzaman, Roger C Shouse, Huh Chang

Deog, "COVID-19: Access to Information, Health Service, Daily Life Facility and Risk Perception of Foreigners during Coronavirus pandemic in South Korea," Archives of Medical Science, 2021, https://doi.org/10.5114/aoms/141164

41. Y. Shi, S. Wang, S. Zhou and M. M. Kamruzzaman. (2020). Study on Modeling Method of Forest Tree Image Recognition Based on CCD and Theodolite. IEEE Access, vol. 8, pp. 159067-159076, 2020, doi: 10.1109/ACCESS.2020.3018180
42. Guobin Chen, Zhiyong Jiang, M.M. Kamruzzaman. (2020). Radar remote sensing image retrieval algorithm based on improved Sobel operator, Journal of Visual Communication and Image Representation, Volume 71, 2020, 102720, ISSN 1047-3203 https://doi.org/10.1016/j.jvcir.2019.102720.
43. M. M. Kamruzzaman, ""New Opportunities, Challenges, and Applications of Edge-AI for Connected Healthcare in Smart Cities,"" 2021 IEEE Globecom Workshops (GC Wkshps), 2021, pp. 1-6, doi: 10.1109/GCWkshps52748.2021.9682055."
44. Md Selim Hossain, MM Kamruzzaman, Shuvo Sen, Mir Mohammad Azad, Mohammad Sarwar Hossain Mollah, Hexahedron core with sensor based photonic crystal fiber: An approach of design and performance analysis," Sensing and Bio-Sensing Research, 32, 100426
45. Mingju Chen, Xiaofeng Han, Hua Zhang, Guojun Lin, M.M. Kamruzzaman, Quality-guided key frames selection from video stream based on object detection, Journal of Visual Communication and Image Representation, Volume 65, 2019, 102678, ISSN 1047-3203
46. M. M. Kamruzzaman: Performance of Decode and Forward MIMO Relaying using STBC for Wireless Uplink. JNW 9(12): 3200-3206 (2014)
47. M. M. Kamruzzaman, "Performance of Turbo Coded Vertical Bell Laboratories Layered Space Time Multiple Input Multiple Output system," Computer and Information Technology (ICCIT), 2013 16th International Conference on, Khulna, 2014, pp. 455-459.
48. S. V. N. Sreenivasu, S. Gomathi, M. Jogendra Kumar, Lavanya Prathap, Abhishek Madduri, Khalid M. A. Almutairi, Wadi B. Alonazi, D. Kali, S. Arockia Jayadhas, "Dense Convolutional Neural Network for Detection of Cancer from CT Images", BioMed Research International, vol. 2022, Article ID 1293548, 8 pages, 2022.
49. Mukesh Soni, Ihtiram Raza Khan, K. Suresh Babu, Syed Nasrullah, Abhishek Madduri, Saima Ahmed Rahin, "Light Weighted Healthcare CNN Model to Detect Prostate Cancer on Multiparametric MRI", Computational Intelligence and Neuroscience, vol. 2022, Article ID 5497120, 11 pages, 2022.
50. Mehraj, Haider, D. Jayadevappa, Sulaima Lebbe Abdul Haleem, Rehana Parveen, Abhishek Madduri, Maruthi Rohit Ayyagari, and Dharmesh Dhabliya. "Protection motivation theory using multi-factor authentication for providing security over social networking sites." Pattern Recognition Letters 152 (2021): 218-224.
51. Abhishek Madduri. "Human Gait Recognition using Discrete Wavelet and Discrete Cosine and Transformation Based Features." International Journal of Computer Trends and Technology, vol. 69, no. 6, June. 2021, pp.22-27.
52. Abhishek Madduri. Content based Image Retrieval System using Local Feature Extraction Techniques. International Journal of Computer Applications 183(20):16-20, August 2021.
53. Sharma, D. K., Chakravarthi, D. S., Boddu, R. S. K., Madduri, A., Ayyagari, M. R., & Khaja Mohiddin, M. (2023). Effectiveness of Machine Learning Technology in Detecting Patterns of Certain Diseases Within Patient Electronic Healthcare Records. In Proceedings of Second International Conference in Mechanical and Energy Technology (pp. 73-81). Springer, Singapore.