JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CRIME DATA RATE PREDICTION USING ENSEMBLEMACHINE LEARNING ALGORITHMS

[1]G Mahesh, [2]K Vishnu Vardhan, [3]Md Abdul Imran, [4]A Ravi Kumar, [5]Dr. Subhani Shaik
[1,2,3]IV Year Students, [4]Assistant Professor,[5]Associate Professor
Dept. of Information Technology, Sreenidhi Institute of Science and Technology(Autonomous), Hyderabad, India

Abstract : Crime Is A Pervasive Problem In Modern Civilization, With A Multitude Of Crimes Being Committed On A Day-To-Day, Causing Great Unrest Among The General Populace. Preventing Crime Is Therefore An Essential Task. In Recent Times, The Significance Of Artificial Intelligence Has Been Observed In Almost Every Field, Including Crime Prediction. To Achieve This, It Is Essential To Maintain Comprehensive Data On Past Crimes, As This Info Can Be Utilized For Future Orientation. Envisaging The Likelihood Of Upcoming Crimes Can Help Law Administration Agencies Prevent Them Before They Occur, Providing Valuable Strategic Evidence Based On Factors Such As Time And Location. However, Accurately Predicting Crime Is A Challenging Task, Given The Increasing Rate At Which Crimes Are Being Committed. Therefore, The Development Of Effective Crime Prediction And Analysis Methods Is Crucial In Detecting And Reducing Future Crimes. To This End, Many Researchers Have Led Trials To Predict Crimes Using Support Vectors, Logistics Regression, Decision Trees, And Xgboost Machine-Learning Techniques.
Keywords: Crime Rate Prediction, Accuracy, Machine Learning Algorithms

1. INTRODUCTION

Crime rate prediction is a field of study that aims to use machine learning methods to predict and analyze the incidence of criminal activity in a particular area ata specific time. The goal is to recognizeoutlines and tendencies in crime data and use this data to develop stratagems and interferences that can help avoid crime. Crime rate prediction typically involves the collection and analysis of large amounts of data, including crimeand other relevant data [1]. Machine learning procedures are castoff to recognizeoutlines in the data, and to progress models that can expect the likelihood of upcoming criminal movement. A few key applications of crime rate calculation and scrutinyconsist of the progress of prognostic policing approaches. the recognition of high-risk areas for lawbreaking, and the assessment of the efficiency of crime inhibition programs [2]. By providing law implementation agencies and legislators with correct and timely dataon near crime tendencies and patterns, crime rate prediction and investigation can assistin progressing public security and decreasing crime rates. utilized a crime dataset from Kaggle to train the model[3]. Opted to use the Random Forest algorithm to analyze the patterns of crime and predict impending crime rates based on the crime category. Additionally, conducted an investigation of crime rates across states over time, examining the time and kind of crime. Through these methods, our study provides a comprehensive and structured overview of trends in crime rates and evaluates existing methodologies for predicting crime rates [4].

Asummary of numerouscrime estimateapproaches, countingarithmetic models, and machine learning models, for spatial analysis, is provided in this survey. The preprocessing methods and data sources used in crime prediction studies are also covered in the article [5]. Predicting Crime with Machine Learning,this review essencesmainly on using machine learning methods to predict crime. An overview of the different machine learning algorithms that are used to analyze the different kinds of data used in crime prediction studies [6]. A Survey of Crime Prediction and Analysis Using Data Mining Techniques,this survey offers an overview of data mining procedures used for crime analysis and prediction. The paper covers various data mining methods, including clustering, classification, and association rule mining, that are applied in crime prediction studies [7].

A Systematic Review and Analysis of Benefits and Drawbacks of Predictive Policing: A Systematic Review and Future Directions,this systematic review focuses particularly on predictive policing, a contentious method of predicting crime that uses machine learning algorithms to pinpoint neighborhoods and people who are most likely to commit crimes [8]. In addition to outlining the present state of predictive policing study, the paper also discusses some of its potential advantages and disadvantages.Exploring Spatial Analysis Techniques in Crime Prediction [9]. Spatial Analysis and Crime Prediction: A Review of the Literature,the methods of spatial analysis employed in studies of crime prediction are the particular focus of

this review of the literature. The article gives asummary of several spatial analysis procedures in addition to the kinds of data that were used in these studies [10].

2. PROPOSED SYSTEM AND ARCHITECTURE

Crime hotspot estimate is a method used to recognize areas whereupcoming crime actions are likely to happen. In this study, the RF algorithm is applied to predict crime trials. The RF algorithm has two aspects of randomness. The primary aspect is the use of a bagging algorithm to arbitrarily select training samples, the second partincludes the assortment of a randomlydividedfeature set [11]. The finestdivided mode of attributes is used to split the nodes. The multi-decision tree formed by the random forest algorithm ultimately determines the endingcatalogingoutcome based on the tree classifier. The target of this research is to grow a model for foreseeing crime. The training dataset is cast-off to produce the model, and the test dataset is used to verify it. Based on its accuracy, the algorithm cast-off to explain the model is selected. The Random Forest algorithm is applied in this situation to predict crime. The dataset is pictured to explain the crimes, that have happened in the country, which can support in discoveryof tendencies and possible hotspots for crime actions in the future [12].

3.1 Benefits Of The Proposed System

The persistence of this research is to advance the earlier proposed estimate framework by usingalternate crime plotting and feature engineering methods and to deliver an open-source execution that can be applied by policy analysts for additional effective predictive policing. This study aids law enforcement agencies in India by providing more accurate crime prediction and detection, ultimately leading to a reduction in the overall crime rate [13].

3.2 Proposed System Architecture

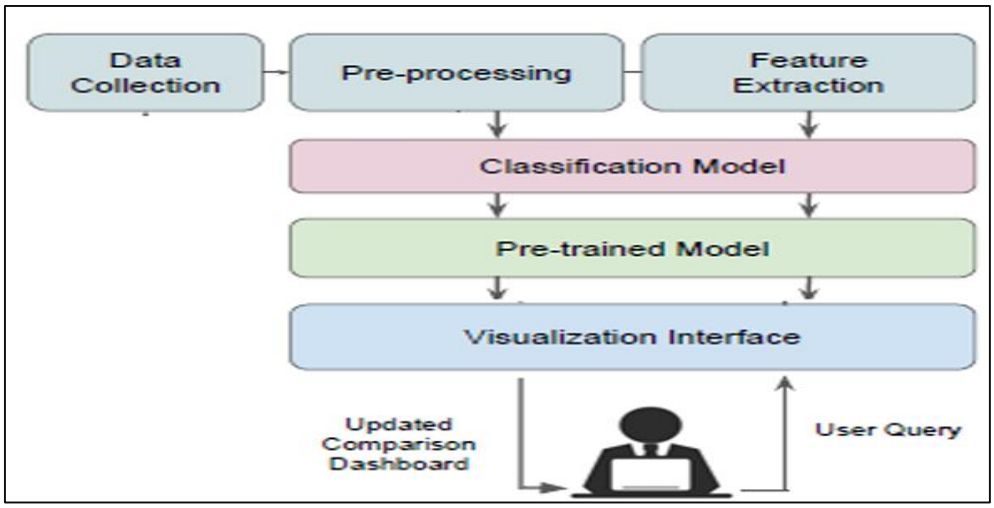

Figure 1: system architecture

4. MODULE DESCRIPTION

The following phases are implemented to find the results. Every phase is crucial for getting accurate results [14].

- Collection of data
- Description of dataset
- Pre-processing of data(loading the data)
- Selection of model
- Analysis of data
- Check Accuracy on test set
- Saving the Trained Model for further testing

4.1 Collection Of Data

In the machine learning model, data collection is the primarysubstantial step togenuineformation. The Tata collection step is a vital phase influenced by more and better data; we can collect. Different approaches were used for collecting the data counting physical interferences, and web scraping. Machine Learning Algorithms for predictionof Crime Hotspots from Kaggle and other sources.

4.2 Dataset Description

The dataset has different data points with clear attribute types. The dataset contains 15 columns, everyone is detailed below.

State: One of the Indian state
District:A particular district
Year: 2001-2023
Murder: Murder rate
Rape: Rape rates
Theft: Theft rates
Total number of crimes: Whole crime rate

Table 1: Crime Dataset

Id	DISTRICT	STATE-todel	ASSAULT ON WOMEN	DACOITY	DOWRY DEATHS	KIDNAPPING	MAJOR_CRIMES_T	MURDER	RAPE	RIOTS	ROBBERY	THEFT	total_crim	YEAR
1	Ariyalur	Tamil nadu	64	0	1	26	233	27	20	19	9	76	1977	2010
2	Chennai	Tamil nadu	45	2	16	47	1858	103	47	60	74	1540	10869	2010
3	Chennai rh	Tamil nadu	2	1	2	0	250	0	0	14	58	232	647	2010
4	Chennai ci	Tamil nadu	61	1	18	30	1059	86	29	85	102	750	8670	2010
5	Coimbator	Tamil nadu	39	11	1	62	664	41	13	83	78	425	4806	2010
6	Coimbator	Tamil nadu	35	1	1	22	1202	22	15	23	25	1084	4180	2010
7	Cuddalore	Tamil nadu	8	1	8	39	530	58	27	149	16	241	6179	2010
8	Dharmapu	Tamil nadu	42	2	6	47	437	59	15	107	25	161	3761	2010
9	Dindigul	Tamil nadu	43	4	1	43	644	57	9	81	29	410	5299	2010
10	Erode	Tamil nadu	43	3	2	64	681	57	15	39	73	461	6702	2010
11	Kanchipura	Tamil nadu	23	2	1	9	494	57	17	102	21	285	5757	2010
12	Kanyakum	Tamil nadu	10	3	2	57	441	42	12	0	90	318	4752	2010
13	Karur	Tamil nadu	28	1	4	54	379	19	10	23	34	241	2974	2010
14	Krishnagiri	Tamil nadu	27	0	5	64	515	58	20	120	14	221	4580	2010
15	Madurai	Tamil nadu	62	2	14	58	726	59	17	168	95	348	5348	2010
16	Madurai C	Tamil nadu	12	0	11	15	592	34	3	17	77	500	2672	2010
17	Nagapattir	Tamil nadu	57	0	1	38	367	50	9	39	18	173	4027	2010
18	Namakkal	Tamil nadu	52	6	2	48	487	50	10	70	14	255	4690	2010
19	Nilgiris	Tamil nadu	10	0	0	8	115	11	7	18	9	61	1320	2010
20	Perambalu	Tamil nadu	16	2	0	10	144	13	5	7	25	93	1536	2010
21	Pudukotta	Tamil nadu	42	3	5	54	448	45	7	45	55	250	4808	2010
22	Ramnatha	Tamil nadu	14	1	1	61	402	33	16	96	26	181	3462	2010
23	Salem	Tamil nadu	165	9	4	140	945	81	53	281	42	221	8634	2010
24	Salem City	Tamil nadu	25	1	3	44	473	27	21	32	70	321	3218	2010
25	Sivagangai	Tamil nadu	16	3	4	32	310	38	1	21	21	198	2611	2010
26	Thanjavur	Tamil nadu	75	4	3	42	1183	82	7	104	141	870	6946	2010

5. RESULTS AND ANALYSIS

5.1 Random Forest And Decision Tree Regressor

In our experiment, we used random forest and decision tree regressor combined to give 95%+ accuracy of classification. The following figures 1 and 2 show the login page and theft crime rate.

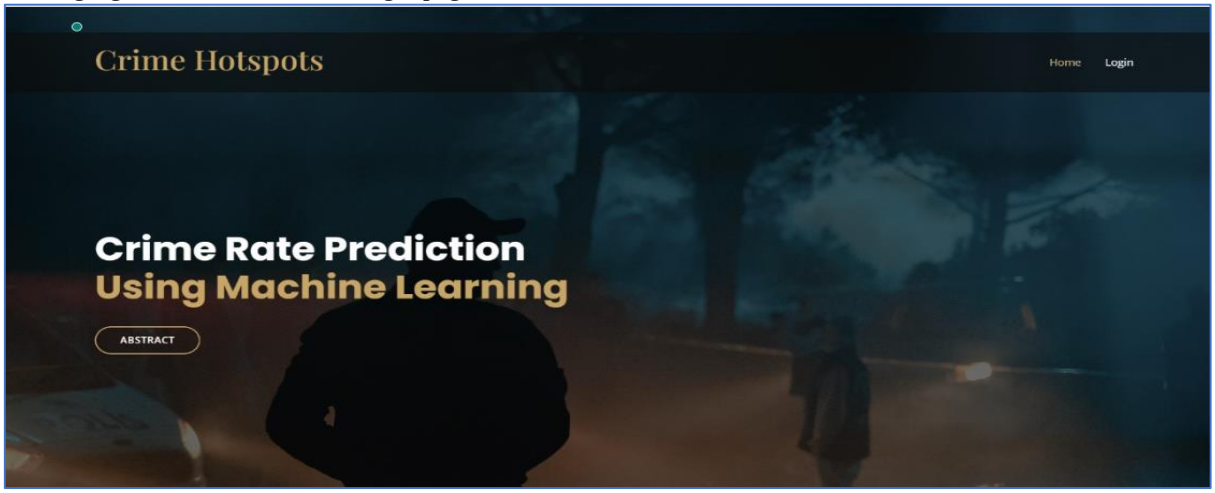

Figure 2: Login page

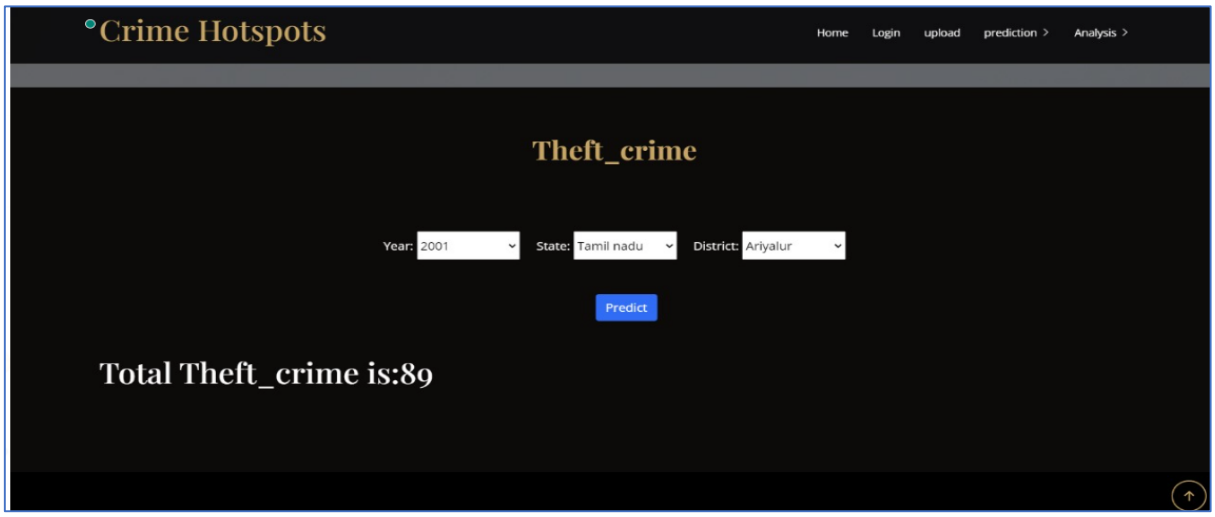

Figure 3: Total Theft Crime Page

Figure 3 shows the murder crime rate of a particular state.

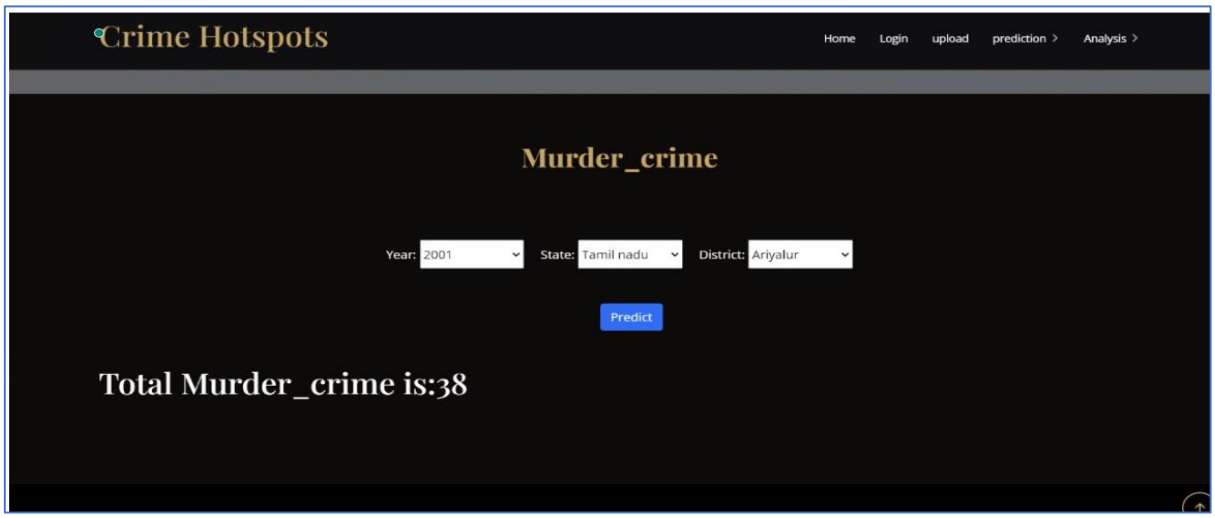

Figure 4: Total Murder Crime Page

Figure 4 shows the crime rate of Tamilnadu, Delhi, Kerala, and Puducherry states.

Figure 5: Total Crimes in the State

The above figure 5 shows the different types of crimes represented in the form of percentages.

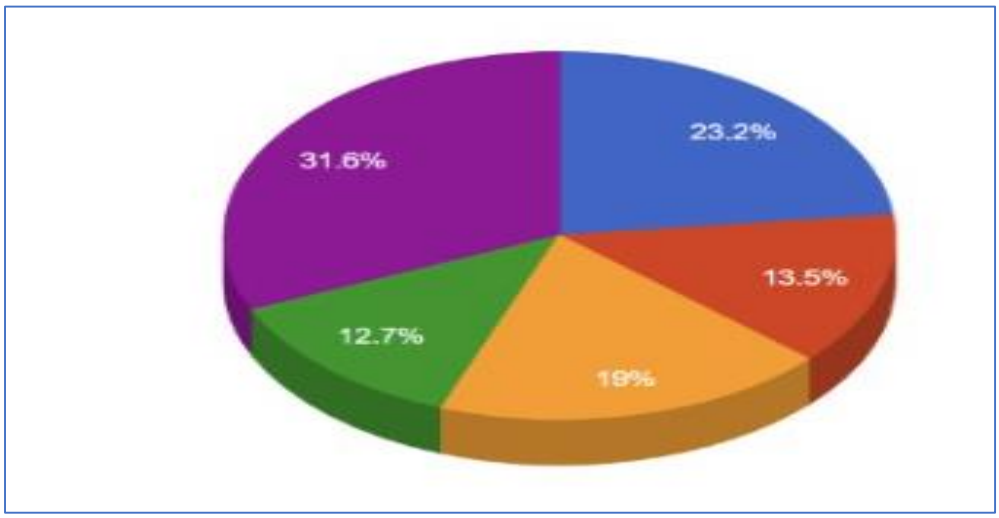

Figure 6: Total Crimes during the year 2017 to 2023

5.2 Support Vector and XGBoost regressor

In our experiment, we used a support vector regressor and XGBoost regressor combined to give different values of crime rates like theft, murder, and rape, etc. The following predictions of SVR and XGB Regressor model building. Each crime is predicted individually with the help of a Support Vector regressor and XCBoost regressor [15].

The following individual predictions for rapes, murders, theft, and overall crime rate. Here we check the mean square error rate of the support vector and XGBoost algorithms. In case 1 of rapes prediction,the XGBoost algorithm generatesminimum MSE. In case 2 murders SVM generates less MSE. In case 3 theft XGBoost generates less MSE. In case 4 overall crime prediction XGBoost generates less MSE.

Case 1: Handling rapes prediction

SVM Mean Squared Error: 85.26

XGBoost Mean Squared Error: 76.77

Case 2: Handling murder prediction

SVM Mean Squared Error: 53.23

XGBoost Mean Squared Error: 117.0

Case 3: Handling theft prediction

SVM Mean Squared Error: 190.4

XGBoost Mean Squared Error: 54.4

Case 4: Handling Total Crime Prediction

SVM Mean Squared Error: 93.6

XGBoost Mean Squared Error: 81.6

5.2 Comparative Study

If you observe Table 1, different machine learning algorithms are used for calculating the accuracy rate with the crime dataset.This table shows ensemble or hybrid learning and individual classification applied tothe crime dataset. The Random Forest tree algorithm predictsahigher accuracy of 97.7% and ensemble learning of RF and DT predicts an accuracy rate is 95.5%. Compared to all machine learning algorithms, RF generatesthe highest accuracy rate. In ensemble learning of our research RF and DT predict the highest accuracy compared to SVM and XGBoost algorithms.

Table 2: Accuracy of different Machine Learning Algorithms

Algorithm	**Accuracy Rate**
Random Forest	97.7%
Decision Tree	91.5%
Gradient Boosting	92%

Support Vector Machine & XGBoost	93%
Random Forest & Decision Tree	95.5%

6. CONCLUSION

Preventing crime is therefore an essential task. In recent times, the significance of artificial intelligence has been observed in almost every field, including crime prediction. To achieve this, it is essential to maintain comprehensive data on past crimes, as this info can be utilized for future orientation. Envisaging the likelihood of upcoming crimes can help law administration agencies prevent them before they occur, providing valuable strategic evidence based on factors such as time and location. However, accurately predicting crime is a challenging task, given the increasing rate at which crimes are being committed. Therefore, the development of effective crime prediction and analysis methods is crucial in detecting and reducing future crimes. To this end, many researchers have led trials to predict crimes using support vectors, Logistics regression, Decision trees, and XGboost machine-learning techniques.

REFERENCES

[1] Aarthi S, Samyuktha M, Sahana M (2019) Crime hotspot detection with clustering algorithm using data mining. In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, pp 401–405.

[2] Aldossari BS, Alqahtani F M, Alshahrani NS, Alhammam MM, Alzamanan RM, Aslam N (2020) A comparative study of decision tree and naive Bayes machine learning model for crime category prediction in Chicago. In: Proceedings of 2020 the 6th International Conference on Computing and Data Engineering, pp 34–38.

[3] Baek MS, Park W, Park J, Jang KH, Lee YT (2021) Smart Policing Technique with Crime Type and Risk Score Prediction Based on Machine Learning for Early Awareness of Risk Situation. IEEE Access 9:131906–131915.

[4] Bandekar SR, Vijayalakshmi C (2020) Design and analysis of machine learning algorithms for the reduction of crime rates in India. Procedia Comput. Sci 172:122–127.

[5] Bappee FK, Soares A, Petry LM, Matwin S (2021) Examining the impact of cross-domain learning on crime prediction. J Big Data 8(1):1–27.

[6] Subhani Shaik and Dr. Uppu Ravibabu "Classification of EMG Signal Analysis based on Curvelet Transform and Random Forest tree Method",Journal of Theoretical and Applied Information Technology (JATIT), Vol. 95, December 2017.

[7] Ch. Shravya, Pravallika, and Subhani Shaik," Brest cancer prediction using machine learning Techniques", International Journal of Innovative Technology and Exploring Engineering, Vol. 8, Issue 6, 2019.

[8] KP Surya Teja, K Vigneswara Reddy, and Subhani Shaik," Flight Delay Prediction Using Machine Learning Algorithm XGBoost", Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No. 5, 2019.

[9] Dr. Sunil Bhutada and Subhani Shaik, "IPL Match Prediction using Machine Learning", IJAST, Vol.29, Issue 5, April-2020.

[10] Dr. R. Vijaya Kumar Reddy, Dr. Shaik Subhani, Dr. G. Rajesh Chandra, Dr. B. Srinivasa Rao," Breast Cancer Prediction using Classification Techniques", International Journal of Emerging Trends in Engineering Research, Vol. 8, No.9,2020.

[11] Mr. Sujan Reddy, Ms. Renu Sri and Subhani Shaik," Sentimental Analysis using Logistic Regression", International Journal of Engineering Research and Applications (IJERA), Vol.11, Series-2, July-2021.

[12] Ms. Mamatha, Srinivasa Datta and Subhani Shaik," Fake Profile Identification using Machine Learning Algorithms", International Journal of Engineering Research and Applications (IJERA), Vol.11, Series-2, July-2021.

[13] R. Vijaya Kumar Reddy, Subhani Shaik, B. Srinivasa Rao, "Machine learning based outlier detection for medical data" Indonesian Journal of Electrical Engineering and Computer Science, Vol. 24, No. 1, October 2021.

[14] Subhani Shaik and Dr. Uppu Ravibabu, "Detection and Classification of Power Quality Disturbances Using Curvelet Transform and Support Vector Machines", in the 5th IEEE International Conference on Information Communication and Embedded System (ICICES-2016) at S.A Engineering College, Chennai, India on 25th -26th, February 2016.

[15] Shiva Keertan J and Subhani Shaik," Machine Learning Algorithms for Oil Price Prediction", International Journal of Innovative Technology and Exploring Engineering, Volume-8 Issue-8, 2019.

[16] Dr. Subhani shaik, Vinay Kumar," Patient expenditure prediction in health care sector", International journal of science and Research, Vol. 11, Issue 6, June 2022.

[17] Dr. Subhani shaik, Ida Fann," performance indicator using machine learning techniques",Dickensian journal, volume 22, issue 6, June, 2022.

[18] Dr. Subhani shaik, Ganesh," Taming an autonomous surface vehicle for path following and collision avoidance using deep reinforcement learning",Dickensian journal, volume 22, issue 6, June, 2022.

[19] Dr. Subhani Shaik, Ch. Samyukth, "Geo tracking of waste and mapping areas with high waste index", Dickensian journal, volume 22, issue 6, June, 2022.