Amrita School of Arts and Sciences

Department of Computer Science

MINOR PROJECT – UG – 2021 -2022

CRIME AGAINST WOMEN ANALYSIS AND PREDICTION USING MACHINE LEARNING LANGUAGE

CERTIFICATE:

Crime against women Analysis and Prediction using Machine Learning algorithm

A project report submitted by

Rakshith S Nadiger [MY.SC.I5MCA19021]

Sumukha M [MY.SC.I5MCA19037]

Sumukh MR [MY.SC.I5MCA19035]

in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF COMPUTER APPLICATIONS

under the guidance of

Mr. Akhil K M

Asst. Professor, Department of Computer Science,

Amrita Vishwa Vidyapeetham,

Mysuru Campus

AMRITA VISHWA VIDYAPEETHAM

MYSURU CAMPUS

January 2022

AMRITA VISHWA VIDYAPEETHAM

MYSURU CAMPUS

BONAFIDE CERTIFICATE

This is to certify that the project entitled **"Crime against women Analysis and Prediction using machine learning algorithm"** submitted by,

Rakshith S Nadiger [MY.SC.I5MCA19021]

Sumukha M [MY.SC.I5MCA19037]

Sumukh M R [MY.SC.I5MCA19035]

Supreeth N M [MY.SC.I5MCA19003]

for the award of the degree of **Bachelor of Computer Applications** at **Amrita School of Arts and Sciences** is a bonafide record of the work carried out by them under my guidance and supervision at Amrita University, Mysuru.

SUPERVISOR

Mr. Akhil K M

Asst. Professor
Department of Computer Science
Amrita School of Arts and Sciences

PROJECT CO-ORDINATOR (UG)

Dr. Suresh K

Asst. Professor
Department of Computer Science
Amrita School of Arts and Sciences

CHAIRPERSON

Mr. ADWITIYA MUKHOPADHYAY

AMRITA VISHWA VIDYAPEETHAM

MYSURU CAMPUS

DECLARATION

We,

Rakshith S Nadiger [MY.SC.I5MCA19021]

Sumukha M [MY.SC.I5MCA19037]

Sumukh M R [MY.SC.I5MCA19035]

Supreeth N M [MY.SC.I5MCA19003]

hereby declare that this project report, entitled "**Crime against women Analysis and Prediction using machine learning algorithm"** is a record of the original work done by us under the guidance of **Mr. Akhil K M**, Asst. Professor, Department of Computer Science, Amrita School of Arts and Sciences, Mysuru and that to the best of our knowledge, this work has not formed the basis for the award of any degree/diploma/associate-ship/fellowship or a similar award, to any candidate in any University.

Signature of the Student

1. Rakshith S Nadiger
2. Sumukha M
3. Sumukh M R
4. Supreeth N M

Place: Mysuru

Date: 21/01/2022

ACKNOWLEDGEMENT

We would like to express our sincere thanks to **Amma**, our beloved chancellor "**Mata Amritanandamayi Devi**".

We would like to express our sincere thanks to **Br. Anantaananda Chaitanya**, Director, Mysuru Campus, Mysuru who supported us with continuous encouragement, motivation and lending us with right direction throughout our tenure of studies.

We would like to express our sincere thanks to **Br. Muktidamrita Chaitanya**, Correspondent, Mysuru Campus, Mysuru for providing us with well supported peaceful studying environment with excellent infrastructure and support.

We would like to express our sincere thanks to our beloved principal **Dr. G Ravindranath** for giving us moral support and continuous encouragement which has been the key for the successful completion of the project.

We are pleased to acknowledge **Mr. Adwitiya Mukhopadhyay**, Chairperson, Department of Computer Science, for his encouragement and support throughout the project.

We would like to express our heart-felt gratitude to our Project Co-ordinator **Dr. Suresh K,** Asst. Professor, Department of Computer Science for his valuable suggestions and excellent guidance rendered throughout this project.

We would like to express our heart-felt gratitude to our guide **Mr. Akhil K M**, Asst. Professor, Department of Computer Science for his valuable suggestions and excellent guidance rendered throughout this project.

Further, we extend our thanks to all the faculty members and technical staff of our department and ICTS for their suggestions, support and providing resources at needed times.

Finally, we express our sincere gratitude to our parents and friends who have been the embodiment of love and affection which helped us to carry out the project in a smooth and successful way.

CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

ABBREVIATIONS	FULL FORM
NCRB	National Crime Records Bureau
RF	Random Forest
GBR	Gradient Boosting Regressor

Table 1.1

ABSTRACT

Crime Investigation is the use of facts to analyse a crime. It involves systematically studying the crime scene and the evidence gathered. There are many social, temporal, spatial and demographic factors which assist the police in the evaluation of crime. As we all know crime is an offense that is prohibited by law.

Crimes are social irritation and causes deep damage to our society in several ways. Any research that helps solving the crime quickly will pay for itself. Using data mining we can extract some important/critical information which can help the local authorities like police, to detect crime. The main purpose of this paper is to analyse the crime pattern of India and take measures to curb crime rates. In addition to all the existing project we have planned to check how Covid-19 has affected India's crime rate and analyse state wise crime pattern and also crime against on women.

1.1 INTRODUCTION TO BROAD AREA OF RESEARCH:

Data Analysis is our Broad area of research. Data Analysis is the process of collecting, pre-processiong, analyzing and visualizing data in order to draw conclusions. As previously mentioned, Data Analysis consists of five major steps.

1. Collection of data:
   This is the first step of data analysis which includes accumulation/ collecting data from various sources. In our project, we have collected data from NCRB (National Crime Records Bureau).
2. Pre-processing of data:
   This is one of the important step in Data analysis. In this step, we will clean the data, fill missing value, deal with outliers and validate the data.
3. Analyzing data:
   Here, we will analyze the processed data, make some observation which will be helpful to obtain results/conclusions.
4. Visualization:
   In this step, we will visualize data by using graphs and charts. It is very effective in quick understanding of the data. These graphs and charts are even helpful to present the results that are obtained by our analysis.
5. Results/Conclusions:
   This is the final phase where we will draw conclusions based on our observations and these conclusions will be the result of our analysis which will be helpful in further decision making.

1.2 INTRODUCTION TO SPECIFIC AREA OF RESEARCH:

Predictive analysis is specific area of research. Predictive analytics is a branch of advanced analytics that makes predictions about future outcomes using historical data. This is very much helpful in decision making. To perform predictive analysis, we have to build model using algorithm or any other statistical methods.

Selection of algorithm plays a vital role. After selection of algorithm, we will be able to build a model which will be able to make prediction. For instance, we have used GBR to build our model. After modelling we must evaluate and validate the model. We have used metrics such as r2 score ,mean squared error and mean absolute error for evaluation of model. For Validation, we have used K-fold Cross validation(10 folds). After evaluation and validation we will be able to predict future data using historical data

1.3 INTRODUCTION TO BACKGROUND OF PROBLEM RESEARCH:

- Analyze state wise crime rates and crime against women from 2011-2020.
- Performing predictive analysis using Supervised learning algorithm.

1.4 OBJECTIVES OF RESEARCH:

- The main objective of the study is to analyse crime on women across different states.
- The second objective is to analyse the impact of covid-19 on crime.
- Third objective of the study is to predict the future crime rates by building predictive model.

1.5 Applications and contributions:

- Can be applied in police station for crime investigation.
- Can be used to know future crime rate
- Useful to control crime

LITERATURE REVIEW:

Sl.No	Author and Year of publication	Title of the paper	Methodology used	Observation	Limitations
1	Bhargava, D., P. Singh, and R. S. Sangwa-2018	Analysis of Crime data using Data mining	Decision tree algorithm and Apriori algorithm.	Data source of various crime is provided and it is very detailed.	Limited visualization
2	Deepika, K., and Smitha Vinod-2018	Crime analysis in India using	Data pre-processing, K means	Used BPN classifier for predictions.	Some visualizations are difficult to understand(doesn’t

		Data mining techniques	Clustering, Random forest algorithm, Neural networks	KNN classification used for criminal identification, WEKA tool used for analysis	have labels/headdings)
3	Yadav, Sunil, et al-2017	Crime Pattern Detection, Analysis and Prediction	Association Mining, K-Means clustering, Naïve Bayes classification technique, Correlation & Regression.	This paper helps in finding out the crime patterns, trends, mapping criminal networks	No Limitations
4	Gupta Neha-2019	Crime in India: An Inter-State analysis	Data pre-processing, K Means Clustering	Detailed study of every states. Perfect data collection.	No future Predictions
5	Maity, Shrabanti, and Sucharita Roy-2021	Analysis of Growth and Identifications of the Determinants of Crime against Women: Insight from India	Augmented Dicky fuller test. Phillips perron test.	Detailed study of every states. Perfect data collection.	Limited visualization
6	Joshi, Anant, A. Sai Sabitha, and Tanupriya Choudhury-2017	Crime analysis using K-means clustering	K-Means Clustering	Made crime analysis using both qualitative and quantitative methods. Good visualization	Insufficient textual information. No predictions
7	Sri, L. A., Manvitha, K., Amulya, G., Sanjuna, I. S., & Pavani, V – 2020	FBI CRIME ANALYSIS AND PREDICTION USING MACHINE LEARNING	RF Regression algorithm	In this paper, they have used Random forest algorithm for Crime rate prediction. Implementation is done using Python language.	they have not explained the analysis part in detail and they have not used any validation for their model.

2.2 MOTIVATION:

The crime rates accelerate continuously and the crime patterns are constantly changing. As a result, the behaviours in crime pattern are difficult to explain. The aim is to provide a comprehensive review of theory and research with respect to the prevention of the crime in the society and to implement different data analysis algorithms which address the connections between crime and its pattern.

2.3 Implementation of base paper 1:

Title: FBI CRIME ANALYSIS AND PREDICTION USING MACHINE LEARNING
Journal: JES – Journal of Engineers and Sciences
Year: 2020(April)
In this paper, they have used Random forest algorithm for Crime rate prediction. Implementation is done using Pyhton language.

2.4.1 Results and discussion

After using Random forest, they were able to achieve accuracy of 86.5%. They have given detailed description of the algorithm and few details of existing systems.

2.4.2 Conclusion

Some of the limitations are, they have not explained the analysis part in detail and they have not used any validation for their model.

2.4 COMPARISON BETWEEN TWO APPROACHES:

EXISTING SYSTEM:

- Haven’t done analysis related to covid period.
- Used random forest regression algorithm for predictive analysis.
- Efficiency of 86.7% is achieved.

PROPOSED SYSTEM:

- Analyzing crime rates in covid period
- We have used gradient-booster regression algorithm.
- Efficiency of 98.92% is achieved

3. DESIGN OF PROPOSED APPROACH:

3.1 ARCHITECTURE DIAGRAM:

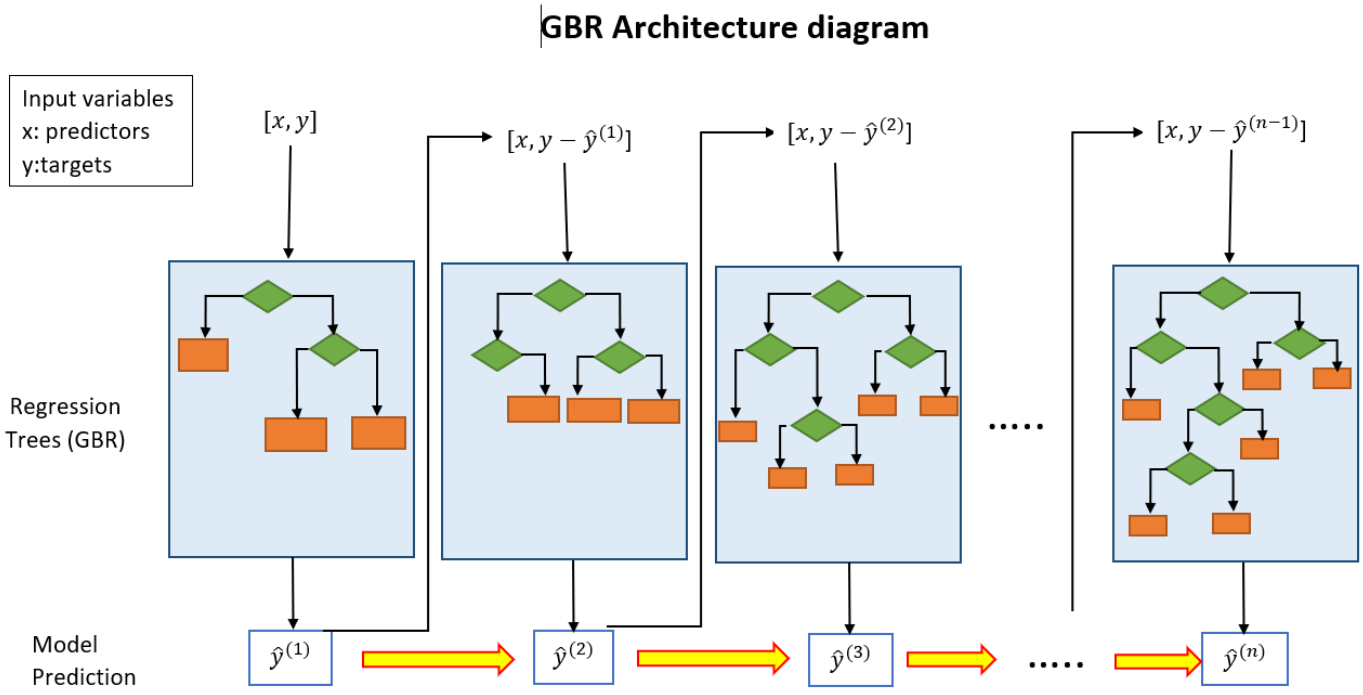

Figure 1.1

CONCLUSION

Crime Data information mining and examination is a dynamic territory of research. The consequences of this investigation may help new potential clients in understanding the scope of accessible wrongdoing information mining methods and advances. Crime information mining can be utilized to give entire crime measurements of a specific locale or territory that gives advantage to the general public by striking the administration and law implementation organizations to comprehend the different causes that expansion the crime rates. The administration and law requirement offices can take better choices for better living of the nationals that normally mean parcel of lives.

METRICS

Name od Metrics	GBR
Mean Squared Error	1001.91
Mean Absolute Error	586.25
R2 Score	0.9892(98.92%)
Cross Validation Score (10 folds)	0.9841(98.41%)

BIBILIOGRAPHY

[1] Sri, Linga Akhila, et al. "FBI CRIME ANALYSIS AND PREDICTION USING MACHINE LEARNING."

[2] Deepika, K., and Smitha Vinod. "Crime analysis in India using data mining techniques." International Journal of Engineering & Technology 7.2.6 (2018): 253-258.

[3] Yadav, Sunil, et al. "Crime pattern detection, analysis & prediction." 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA). Vol. 1. IEEE, 2017.

[4] Gupta, Neha. "Crime in India: An Inter-State Analysis." (2019).

[5] Maity, Shrabanti, and Sucharita Roy. "Analysis of Growth and Identifications of the Determinants of Crime against Women: Insight from India." Journal of International Women's Studies 22.1 (2021): 293-311.

[6] Joshi, Anant, A. Sai Sabitha, and Tanupriya Choudhury. "Crime analysis using K-means clustering." 2017 3rd International conference on computational intelligence and networks (CINE). IEEE, 2017

[7] Sri, L. A., Manvitha, K., Amulya, G., Sanjuna, I. S., & Pavani, V. FBI CRIME ANALYSIS AND PREDICTION USING MACHINE LEARNING.

URLs:

https://data.gov.in/catalog/crime-india

https://en.wikipedia.org/wiki/Random_forest

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

https://en.wikipedia.org/wiki/Gradient_boosting

https://ncrb.gov.in

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html

https://machinelearningmastery.com/gradient-boosting-with-scikit-learn-xgboost-lightgbm-and-catboost/

https://en.m.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_crime