

# Crime Analysis using Machine Learning Algorithm



G. Raj Kumar, Kothamasu G.V Saikumar, Dupuguntla Vaibhav Gopinath, Vemisetty Rochish

**Abstract:** *Counteraction is better than Cure. Forestalling a wrongdoing from happening is superior to examining what or how the wrongdoing had happened. When I pick out do expand this venture the fundamental hassle is growing the centralized server. Awful conduct scene want has relies mostly on the certain awful conduct record and various geospatial and part data. In existing machine they're proposed only getting the crime from the consumer most effective until now they didn't have system for prediction the crime. Wrongdoing that happens nowadays are have following key qualities, for example, violations rehashing in an occasional style, wrongdoings happening because of some other action and event of violations pre shown by some other data. In our proposed system we overcome that answer and we enforce the Prediction System. We need to accumulate raw facts and method in addition. We use Random forest Algorithm*

**Keywords :** Machine learning , Random Forest Algorithm.

## I. INTRODUCTION

A wrongdoing or offense (or criminal offense) is a demonstration hurtful not exclusively to some individual or people yet additionally to a network, society or the express. Numerous significant inquiries in open wellbeing and assurance identify with wrongdoing, and a superior comprehension of wrongdoing is gainful in different manners: it can prompt focused on and touchy practices by law implementation specialists to moderate wrongdoing, and progressively purposeful endeavours by residents and specialists to make sound neighbourhood situations. With the coming of the Big Data period and the accessibility of quick, proficient calculations for information investigation, understanding examples in wrongdoing from information is a functioning and developing field of research. Machine Learning. Is additionally broadly utilized in logical applications, for example, bioinformatics, medication, and

stargazing. Machine learning is the investigation of having PCs choose decisions without human mediation. It has moreover made foreseeing bad behaviour reliant on referenced data conceivable. Portrayal is a directed desire framework which thinks about apparent class names. Game plan has been used in various spaces including atmosphere foreseeing, clinical thought, assets and banking, nation security, and business information. The essential target of this work is to make a forecast model that can precisely anticipate crime In our exploration we utilized Random forest Algorithm.

## II. CONTEMPLATED WORK

In our journal we propose approach for crime detection and analysis using Machine learning algorithm. We introduce improvements for crime analysis in comparison with other work on given literature Random Forest algorithm is used to identify the crime

### Advantages

- . Predicting hotspots of crime.
- . Understanding crime pattern.

## III. SPECIFICATION REQUIREMENT

### Software Requirement

Python 2.7  
Anaconda Navigator

### Hardware Requirement

.Windows 7,8,10 64 bit  
.RAM 4GB

## IV. MODULE DESCRIPTION

### Module-1:

#### Data Set:

.In this module we need to gather the crime subtleties, for example, which kind of crime and guilty parties and so forth. Every one of these subtleties will be put away into Data Base. .In server side DB has the forecast Data set.

### Module-2:

#### Pre Processing:

It consist of two fields:

- .Dates
- .Address

From the 'Dates' field, we extracted the Day, the Month, the Year, the Hour, the Minute, the Weekday, and the number of days since the first day in the data.

Revised Manuscript Received on April 25, 2020.

\* Correspondence Author

G. Raj kumar M.TECH\*, Assitant professor(Sr.G) , Department of CSE SRM Institute of science and Technology ,Chennai, Tamilnadu., India.

annaunivraj@gmail.com

Kothamasu G.V Saikumar Department of CSE, SRM Institute of Science and Technology chennai, Tamilnadu, India kothamasusaikumarcredits@gmail.com

Dupuguntla Vaibhav Gopinath Department of CSE,SRM Institute of Science and Technology ,Chennai, Tamilnadu, India. gopinathv8@gmail.com

Vemisetty Rochish Department of CSE,SRM Institute of Science and Technology ,Chennai, Tamilnadu, India. Rochish2372@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

From the 'Address' field we extracted if the incident has taken place in a crossroad or on a building block.

### Module-3:

#### Random Selection:

- . .Pick N irregular records from the dataset
- . Fabricate a choice tree dependent on these N records.
- . Pick the quantity of trees you need in your calculation and rehash stages 1 and 2.
- . For grouping issue, each tree in the woodland predicts the classification to which the new record has a place. At long last, the new record is relegated to the class that successes the larger part vote.

### Module-4:

#### Trained & Testing:

Training /Testing data creation to evaluate the performance of our models and fine-tune their hyperparameters.

## V. METHODOLOGY

### Data Pre-processing:

Data preprocessing is an information mining system which is utilized to change the crude information in a valuable and proficient arrangement.

It consists of three sub parts:

### Data Cleaning:

The information can have numerous unessential and missing parts. To deal with this part, information cleaning is finished. It includes treatment of missing information, loud information and so on.

### Data Transformation:

This progression is taken so as to change the information in proper structures reasonable for mining process. This includes following ways: Normalization, Attribute Selection, Discretization, Concept Hierarchy Generation.

### Data Reduction:

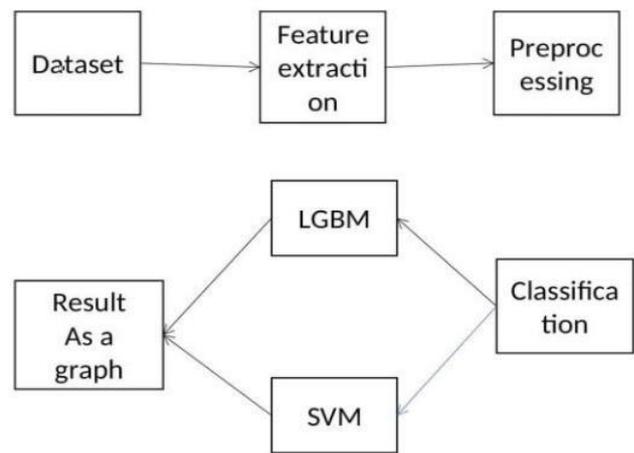
Since information mining is a strategy that is utilized to deal with gigantic measure of information. While working with immense volume of information, investigation got more diligently in such cases. So as to dispose of this, we utilizes information decrease procedure. It means to expand the capacity effectiveness and decrease information stockpiling and investigation costs.

### Feature Scaling:

Deciding to continue with a tree-based algorithm there was no need for scaling on the final dataset. After the element building portrayed above, we wound up with 11 highlights. To distinguish if any of them expanded the multifaceted nature of the model without adding Note worthy addition to the model, we utilized the strategy for Permutation Importance. The thought is that the significance of a component can be estimated by taking a gander at how much the misfortune diminishes when an element isn't accessible. To do that we can expel each component from the dataset, re-train the estimator and check the effect. Doing this would require re-preparing an estimator for each component, which can be computationally escalated. Rather, we can supplant it with claby mix esteems for a component. The execution of the above method demonstrated that there is no requirement for any element

expulsion since every one of them have a positive effect in the dataset.

### Architecture:



**Fig 1: Block Diagram**

The most appropriate evaluation metric for such problems is the multi-class logarithmic loss. Logarithmic misfortune quantifies the presentation of an order model where the forecast yield is a likelihood esteem somewhere in the range of 0 and 1. For each incident, we will predict a set of predicted probabilities (one for every class), and we will calculate the average deviation from the real values. To get a little more intuition on the metric, for a specific incident: We get 0 loss from the categories of crimes that did not happen (since  $y_{ij} = 0$   $y_{ij} \log(p_{ij})$ ) no matter our predicted probability. We get  $\log(p_{ij})$  loss from the category that happened, where  $p_{ij}$  is our predicted probability for the specific category. We have to note here that since all the probabilities for a specific incident sum to 1, each probability we predict for a category that did not happen creates an "indirect" loss since it decreases our predicted probability for the category of crime that happened. In other words, the metric evaluates the certainty of our model for each category of crime/incident.

## VI. CLASSIFICATION

### Random Forest

Random forest is a kind of administered AI calculation dependent on group learning. Group learning is a kind of realizing where you join various sorts of calculations or same calculation on different occasions to shape an all the more impressive expectation model. The arbitrary timberland calculation joins numerous calculation of a similar sort for example various choice trees, bringing about a timberland of trees, consequently the name "Arbitrary Forest". The irregular woods calculation can be utilized for both relapse and grouping [1]. Using Machine Learning to Assist Crime Prevention, 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI). [2]. Prediction Analysis of Crime in India Using a Hybrid Clustering Approach, Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2018).

- [3]. Crime against Women: Analysis and Prediction Using Data Mining Techniques, 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon).
- [4]. Crime Data Analysis and Prediction of Perpetrator Identity using Machine Learning Approach Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018).
- [5] Hybrid Approach to Crime Prediction using Deep Learning ,2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI). errands.

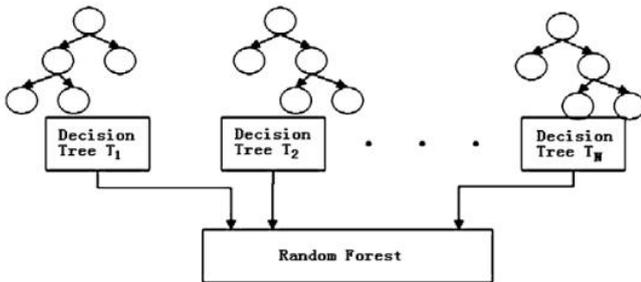


Fig 2: Random Forest

**Decision Tree:**

Decision Tree calculation has a place with the group of administered learning calculations. In contrast to other managed learning calculations, the choice tree calculation can be utilized for taking care of relapse and order issues as well. The objective of utilizing a Decision Tree is to make a preparation model that can use to anticipate the class or estimation of the objective variable by taking in straightforward choice principles construed from earlier data (training information). In Decision Trees, for foreseeing a class name for a record we start from the base of the tree. We look at the estimations of the root quality with the record's trait. Based on correlation, we follow the branch comparing to that worth and hop to the following hub. Decision Trees follow Sum of Product (SOP) representation. The Sum of product (SOP) is also known as Disjunctive Normal Form. For a class, every branch from the root of the tree to a leaf node having the same class is conjunction (product) of values, different branches ending in that class form a disjunction

**VII. EXPERIMENTAL RESULTS**

Following is a corresponding intermedi-ate dump table of correlational values.

X\Y	Piracy	cheating	Sexual_of	personati	Offence_	Forgery	Offence_	Genocide	Violent_offence
Piracy	0.023645	0.023457	0.023924	0.024613	0.025608	0.024796	0.024613	0.02439024	0.023842377
cheating	0.023457	0.023645	0.02435	0.023842	0.02508	0.025019	0.024877	0.02439024	0.025080151
Sexual_of	0.023924	0.02435	0.024077	0.025466	0.024599	0.025628	0.024898	0.02577006	0.024877237
personati	0.024613	0.023842	0.025466	0.024309	0.024816	0.02437	0.025306	0.02505996	0.024309978
Offence_	0.025608	0.02508	0.024599	0.024816	0.026322	0.024857	0.023376	0.024025	0.025790349
Forgery	0.024796	0.025019	0.025628	0.02437	0.024857	0.024471	0.024431	0.0252019	0.023720628
Offence_	0.024613	0.024877	0.024898	0.025306	0.023376	0.024431	0.024025	0.02489759	0.024999805
Genocide	0.02439	0.02439	0.02577	0.02506	0.024025	0.025202	0.024898	0.02223936	0.025668601
Violent_o	0.023842	0.02508	0.024077	0.024309	0.02579	0.023721	0.024999	0.0256686	0.024063532

FIG 3: Probability Matrix

A sample visualisation of internal events graph along with prbability of there relations is show in following image. this simply is the representation of reated events the and their respective probability graph.

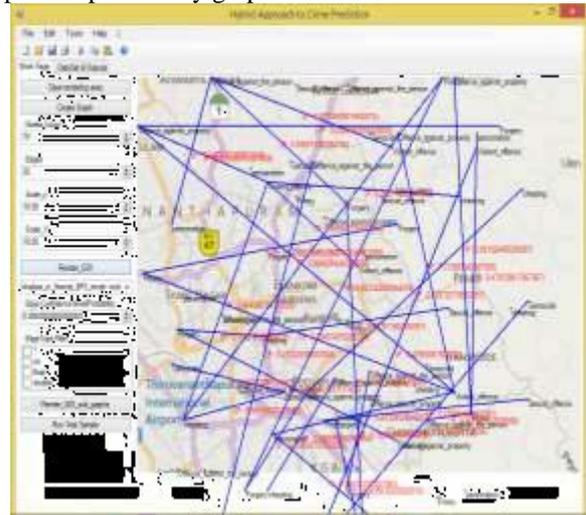


Fig 4: Crime and its relations across various events

**VIII. CONCLUSION**

With the assistance of Machine learning innovation, it has gotten simple to discover connection and examples among different data's. The work in this venture mostly rotates around foreseeing the sort of crimes which may occur on the off chance that we know the area of where it has happened. Utilizing the idea of AI we have manufactured a model utilizing preparing informational collection that have experienced information cleaning and information change. A Random forest Algorithm of various decision trees is used to predict the crime

**REFERENCES**

1. Using Machine Learning to Assist Crime Prevention, 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI).
2. Prediction Analysis of Crime in India Using a Hybrid Clustering Approach, Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2018).
3. Crime against Women: Analysis and Prediction Using Data Mining Techniques, 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon).
4. Crime Data Analysis and Prediction of Perpetrator Identity using Machine Learning Approach Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018).
5. Hybrid Approach to Crime Prediction using Deep Learning ,2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI).

**AUTHORS PROFILE**



**G. Raj Kumar** is an Assistant Professor of the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Chennai. He completed his M.Tech .He participated in a state level workshop on Prediction Techniques & Big Data Machine Learning In Data Workshop.

## Crime Analysis using Machine Learning Algorithm



**Kothamasu G.V Saikumar** is a Final Year B.tech Student in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Chennai. He has better knowledge on IoT ,Machine Learning



**Dupuguntla Vaibhav Gopinath** is a Final Year B.tech Student in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Chennai. He has better knowledge on IoT, Machine Learning.



**Vemisetty Rochish** is a Final Year B.tech Student in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Chennai. He has better knowledge on IoT, Machine Learning.