

An Exploration of Crime Type and Prediction Using RALASD Feature Selection Algorithm with Deep Learning Technique

R. Ananda DhanaLakshmi¹, Dr. Grasha Jacob²

¹Research scholar, Register Number:18121172162002

¹Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli 627 012, Tamil Nadu, India

²Associate Professor & Head

^{1,2}Dept of Computer Science

Rani Anna Govt College for women ,Tirunelveli 627 008,Tamil nadu, India

grasha.ananthi@gmail.com

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract - In daily life, there is an enormous number of crimes that are frequently committed. The crime tracking and maintaining crime dataset is a challenging. Prediction of crime is an administration of regulation in the society by analyzing the statistical by using the data employed from a source. The source of data is used for the analysis of the crime patterns and crime rates in the particular region using data mining and deep learning techniques. The key objective of this work is to analyze the crime activities based on the information set using data mining techniques and predicting the crime rates using deep learning techniques. The work employs the crime dataset with various crime types occurs in the various region for the analysis. The analysis of the crime dataset includes pre-processing steps to construct a crime profile for the prediction. The pre-processing steps include removal of missing values and duplicate information from the dataset and finally convert the dataset into an encoded format for further identification. Then the encoded format of the dataset is employed in the selection phase for the attribute selection using feature selection strategy to reduce and select significant crime variables for the prediction. Finally, the crime prediction process is performed with the deep learning strategy. The prediction of the crime types is performed with the selected subset of the crime variables to increase the prediction accuracy.

Keywords —Crime Detection, Deep Learning, Crime Variables, Feature Selection, RALASD, ANN.

I. INTRODUCTION

Crime is an injurious action; it affects the whole community not only single person. It is one of the momentous subjects that grow in complexity. For example, violent crime includes homicide, aggravated assault, sexual assault, and robbery, while non-violent crime includes burglary, larceny, motor vehicle theft, and arson. Crime prediction is a law enforcement procedure that utilizes data and statistical analysis for the identification of crimes most likely to be happening. The primary target of the framework is to recognize the pattern and type of crime analyze it and predict the crime. Crime pattern recognition analysis and prediction systems are principally founded on Data Mining concepts in which it analyzes different AI and Deep Learning algorithms. The focus of the framework is to dissect the dataset related to criminal records in different areas and to predict the conceivable sort of crime that may occur in different areas.

This research has expanded to predict criminal activities at a faster rate, and it is the responsibility of the police division to control and lessen crimes. Crime prediction is a serious issue for the police department as there is an enormous measure of crime data that exists. There is a requirement for technology through which case-solving could be quicker. The above issue caused this research about how could to settle a criminal case made simpler.

II. REVIEW OF LITERATURE

Chen et al. [1], in which the authors present an overall system for crime data mining, where a large number of these standard devices are accessible as a feature of the COPLINK software package. Much modern work has absorbed on discovering and examining hotspots, which are limited high crime-density areas.

Adeyayo M. Balogun [2] work in a bid to address this issue, looks, to begin with recognizing the specific advantages of criminal profiling have brought to effective modern crime investigations and the advantages it can mean cybercrime examinations, distinguish the difficulties presented by the cyber-scene to its execution in cybercrime investigations, and proffer a practicable answer.

Tonkin et al. [3] took a gander at how much they could effectively characterize violations as connected or unlinked across various sorts of crimes. They looked at violent, sexual, and property offenses. Their outcomes demonstrated that both inter-crime distance and temporal proximity were acceptable pointers for if violations were

connected or unlinked.

This research [4] focuses the utilization of Foursquare data for identify the crime. This system uses feature selection methods to research the force of several highlights got from foursquare to recognize crime comprises in the United States.

This work [5] is to face the issue of crime in the social environment. Predicting and analyzing crimes that occurred in the world would present us a Broadview in understanding the crime area to be utilized to take necessary precautions to predict the crime ratios.This work concentrates on the growth of a crime prediction prototype model with the decision tree (J48) algorithm since it has been measured as the most effective ML algorithm for the recognition of crime as defined in the related literature [6].

III. PROPOSED METHODOOGY

This research intends to make crime predictions using the features present in the dataset. The dataset is extracted from the official sites. Implementing this research with python, here predict the crime type that is violent or non-violent. The objective would be to train a prediction model. The training would be done using the training data set, which will be validated using the test dataset. Building the model will be done using a better algorithm depending upon the accuracy.

The main objectives of the work are:

- To construct the clustering based on the type and hotspot of the crime profile with the pre-processed crime dataset.
- To employ a feature selection strategy for the crime profile identification using the best feature selection for enhancing the prediction.
- To predict the crime with the selected subset based on the crime types along with Dee- learning model.



Fig1.Crime Analysis and Prediction Framework

Fig1.Shows the crime prediction framework that depicts the framework consists of four main layers, which are explained as follows: Pre-Processing Layer, Determination, and Prediction Layer.

A. Pre-Processing Layer

Our Proposed Model contains dataset is a crime-based Dataset that occurred in Chicago City for the COVID-19 period. It got from the Chicago Cop Section system named CLEAR (Citizen Law Enforcement Analysis and Reporting). To secure the protection of crime ID, Crime Types that have appeared at the specific locations have appeared. The dataset contains more than One Lakh records of data and it is complicated to be seen in full in Microsoft Excel. This section of the work includes the crime data pre-processing phases. Initially, the system needs to preprocess data by removing all null values and removing all columns are unnecessary. This procedure comprises a function to delete any null or infinite values that may disturb the accuracy of the system.

- 1) **Crime Variables:** The field in the crime dataset is entitled Crime Variable (Feature) Such as Name, Date, and Block, etc. This Crime Dataset has 22 Types of unique Variables. These variables are used to identify the values in the crime records. Table 1 describes the crime variable, which includes in the crime data set for the crime prediction. The variables in the table, which assist for, further selection and prediction, are included.

Table 1: Crime Variable

| ID | Name | Description |
|----|----------------------------------|--|
| 1 | Crime Identification Number (ID) | Distinctive identifier for the Criminal Record |
| 2 | Case Number | This Number is a reference used for the Criminal Record Division for Cop Stations. |
| 3 | Date | It indicates the Date of the Crime |
| 4 | Block | It means the partial address where the crime incident ensued |
| 5 | IUCR Code | Illinois Uniform Crime Reporting code. |
| 6 | Primary Type | It is the main detail about the IUCR Code |
| 7 | Description | The subordinate detail of the IUCR code refers to the Sub Category of the Crime. |
| 8 | Description of the Location | Report of the place where the Crime ensued. |
| 9 | Arrest | It Specifies whether an arrest occurred or not. |
| 10 | Domestic Data | It describes the incident was domestic-related or not |
| 11 | Beat Data | It refers to the smallest cop geographic area where the crime occurred. |
| 12 | District Data | This field describes the District of the cop where the crime happened. |
| 13 | Ward Information | It refers to the ward information about where the crime befallen |
| 14 | Community Area Details | It noted the community area where the crime has arisen. This filed Seventy-Seven areas. |
| 15 | FBI Code | It designates the crime category as drawn in the FBI. |
| 16 | X Coordinate | This refers to the location where the crime arose in State Plane X-Axis. |
| 17 | Y Coordinate | This refers to the location where the crime arose in State Plane Y-Axis. This location is shifted from the actual location for partial redaction but falls on the same block |
| 18 | Year | The Year of the crime occurred. |
| 19 | Updated On | Recently updated date and time of the crime. |
| 20 | Latitude | The latitude of the location |

| | | |
|----|-----------|--|
| | | where the incident occurred. |
| 21 | Longitude | The longitude of the location where the incident occurred. |
| 22 | Location | Location Refers to the combination of the Latitude and Longitude Details. Both are the angles computed with the center of the earth. |

Table 1 shows the sample dataset for the above-mentioned crime variables. The following table 2 depicts that separate values for the sample dataset based on the variables included.

| ID | Case Number | Date of Crime | Block | Type | Description | Location Description |
|--------------|--------------|---------------------------|-------------------------|------------------------|----------------------------------|------------------------------------|
| 12119 112 | jd312 488 | 07/26/2020 05:00:00 pm | 048xx n central ave | motor vehicle theft | automobile | other (specify) |
| 12116 789 | jd310 153 | 07/26/2020 12:39:00 am | 047xxs rockwell st | weapons violation | reckless firearm discharge | residence - yard (front / back) |
| 12117 549 | jd311 038 | 07/26/2020 08:50:00 pm | 032xx w maypole ave | battery | simple | Street |
| 12117 662 | jd311 173 | 07/26/2020 06:00:00 pm | 029xx e 77th st | battery | simple | park property |
| 12123 105 | jd317 646 | 06/29/2020 10:00:00 am | 012xx n hoyne ave | motor vehicle theft | cycle, scooter, bike with vin | Street |
| 12117 603 | jd311 158 | 07/26/2020 11:35:00 pm | 049xx w crystal st | battery | domestic battery simple | Sidewalk |
| 12116 719 | jd310 162 | 07/26/2020 01:51:00 am | 060xx s carpenter st | weapons violation | unlawful possession handgun | Sidewalk |

Table 2: Depiction of Sample Dataset Based on Crime Variable

2) **Crime Variable Encoding:** After the recognition of the data types of the variable that exists in the crime set, the further process is to make the data in a way that is suitable to put into the Deep Learning Algorithm to predict the crime. This pre-processing layer involves numerous phases, which include Missing and Null Value Treatment, Data Type Detection, Column Removal, Feature Encoding, and so on. Out of these strategies, the one that identifies with the above-examined data types is Feature Encoding. Feature Encoding is the transformation of Categorical features to numeric values, as Deep learning models cannot deal with the categorical information straightforwardly. The vast majority of the Deep Learning Algorithm's performance varies based on how the Categorical data is encoded. In this encoding technique, the categorical data is assigned values from '1' to 'N' (Where 'N' ∈ No. of different categories dataset). This type of encoding function is applied to the ordinal data. The assigning value in ascending or descending orders only cannot be changed this is randomized order. For Example, the crime location should be a Street, Sidewalk, Residence, or Park. It makes no sense in ordering them randomly like Street, Sidewalk, and Residence.

The output to a Crime location of an individual is as follows:

| Crime Occurred Location | Encoding on Crime Occurred Location |
|-------------------------|-------------------------------------|
| Street | 1 |
| Sidewalk | 2 |
| Residence | 3 |

Table 3: Crime Variable Encoding

The encoding based on the crime occurred location of the crime variable is shown in Table 3

3) **Crime Profile Construction:** Crime Profile Construction includes such items as a report on the criminal dataset such as crime types, a description of the area in which the crime occurred, the detailed crime report, and so on. After the pre-processing phase, all pipelining processes are constructed into a single set, which is called Crime Profile. The crime profile constructed with the parameter details is described in Table 4. The table depicts the

An Exploration of Crime Type and Prediction Using RALASD Feature Selection Algorithm with Deep Learning Technique

parameter information and its description based on the crime dataset.

| ID | Parameter | Meaning |
|----|---------------------------------|---------------------------|
| 1 | CS | Raw Crime Dataset |
| 2 | N | Size of Dataset |
| 3 | CSP | Preprocessed Crime Set |
| 4 | CSC | Clustered Crime Set |
| 5 | CSFV | Crime Feature Vector |
| 6 | CSSF | Selected Feature Set |
| 7 | CType \in {Rape, Theft, etc.} | Type of Crimes in Dataset |
| 8 | CSTrain | Training Crime Dataset |
| 9 | CSTest | Testing Crime Dataset |
| 10 | CDate | Date of Crime |
| 11 | CTime | Time of Crime |
| 12 | CStreet | Crime occurs in Street |
| 13 | CDesc | Description of the Crime |

Table 4: Crime Profile Parameter Construction

The crime profile constructed with the parameter details is described in Table 4. The table depicts the parameter information and its description based on the crime dataset.

| CID | CBlock | CIUCR | CDesc | CArrest | CDOM | CBeat | CDistrict | CWard |
|----------|--------|-------|-------|---------|------|-------|-----------|-------|
| 12119112 | 13948 | 88 | 117 | 0 | 0 | 185 | 14 | 45 |
| 12116789 | 13838 | 152 | 306 | 1 | 0 | 107 | 8 | 15 |
| 12117549 | 10487 | 37 | 326 | 0 | 0 | 135 | 10 | 28 |
| 12117662 | 9442 | 37 | 326 | 0 | 0 | 42 | 3 | 7 |
| 12117569 | 1783 | 82 | 310 | 0 | 0 | 209 | 16 | 27 |
| 12116828 | 4994 | 48 | 153 | 1 | 1 | 60 | 5 | 17 |
| 12117603 | 14364 | 48 | 153 | 0 | 1 | 271 | 21 | 37 |
| 12116719 | 16844 | 147 | 372 | 1 | 0 | 72 | 6 | 16 |
| 12117294 | 361 | 37 | 326 | 1 | 0 | 214 | 16 | 2 |

Table5: Post Pre-Processing Crime Profile- Sample Data

The sample dataset based on the crime profile after the pre-processing phase is shown in Table 5.

4) **Removing Columns on Crime Dataset:** Columns in the Crime Dataset that have a similar value should probably be removed from the Crime dataset. To remove the missing value or null value in the dataset using a function IsNull() and not-null() which help in checking whether a value is 'NULL', 'NaN' or 'NOT Nan'. In this Crime Dataset 'Year' is the important column that should be removed because it has more duplication value than is '2020'. The reason for the repeated value of '2020' is the dataset is in the time range of Covid-19. The following table 6 indicates the column name and the reason for the removal of the other field of the crime dataset.

| ID | Columns | Reason for Removal |
|----|-------------|--|
| 1 | Year | The same Values occur in the year (2020) More duplication occurs. |
| 2 | Latitude | This value represents in Location so avoid redundancy removing this field. |
| 3 | Longitude | This Longitude represents in Location so avoid redundancy removing this field. |
| 4 | XCoordinate | It represents the integer value of |

| | | |
|---|-------------|---|
| | | Location (Latitude). It refers to the location where the incident occurred on State Plane. Using Location Field (Latitude) instead of X Coordinate. |
| 5 | YCoordinate | It represents the integer value of Location (Longitude). It refers to the location where the incident occurred. Using Location Field (Longitude) instead of X Coordinate. |

Table 6: Sample of Removed Columns in Crime Dataset

5) **Crime Duplication Checking and Removal:** occasionally the records as duplicates regardless of whether a key variable is repeated. Rather than utilizing the duplicated function on full information, use it on one variable. Duplication can mean two somewhat various things: More than one record that is the equivalent. More than one record is related with a similar perception, yet the values in the rows are not the equivalent, but rather eliminating these kinds of duplicated records is called "record association".

```

CDCR() Algorithm
{
  1. Read the Crime dataset CS
  2. For each i from CS
  3. Allocate the array size for i
  4. Get the chunks from the set
  5. Iteration through i for the hash element in array
  6. If element exists then
  7. Set the element as duplicate
  8. Remove Duplicate (exists element)
  9. End if
  10. Repeat with all chunks until you know they will not
      contain any duplicates from chunk
  11. Deleted all elements marked for deletion
  12. End for
}
    
```

Algorithm 1: Removal of Duplication in Crime Dataset

The avoidance of the duplication from the crime dataset is described as algorithmic representation in Algorithm 1.

B.Determination Layer

Feature selection (FS) is a procedure generally engaged in machine learning to resolve the high dimensionality issue. It picks a subset of significant features and eradicates irrelevant, repeated, and noisy features for simpler and more succinct data illustration. It is to model an objective output variable ‘y’, with a subset of the significant interpreter variables such as inputs. This is a common goal and some more precise objectives can be recognized.

1) Construction of Attribute Vector

The characteristic setting for machine learning (ML) is to be given a group of objects, each of which is considered by numerous dissimilar features. Features can be of altered sorts: e.g., they might be integer, float, or string data types. A feature vector (FV) comprising all of the feature values for a given data point is entitled the FV; Length of Vector is (‘d’), then every data as drawn to a ‘d-dimensional’ vector space, entitled the feature space. Design matrix (DM) is constructed through modified data points in the group. Each row and column refer to the data points of the feature. The design matrix (DM) is the simple data object on which ML algorithms operate.

An Exploration of Crime Type and Prediction Using RALASD Feature Selection Algorithm with Deep Learning Technique

| Block | IUCR | Primary Type | Arrest | Domestic | Beat | District | Ward | Community Area | FBI Code |
|-------|------|--------------|--------|----------|------|----------|------|----------------|----------|
| 13948 | 88 | 17 | 0 | 0 | 185 | 14 | 45 | 10 | 8 |
| 13838 | 152 | 30 | 1 | 0 | 107 | 8 | 15 | 57 | 17 |
| 10487 | 37 | 2 | 0 | 0 | 135 | 10 | 28 | 26 | 10 |
| 9442 | 37 | 2 | 0 | 0 | 42 | 3 | 7 | 42 | 10 |
| 1783 | 82 | 29 | 0 | 0 | 209 | 16 | 27 | 7 | 7 |
| 4994 | 48 | 2 | 1 | 1 | 60 | 5 | 17 | 70 | 10 |
| 14364 | 48 | 2 | 0 | 1 | 271 | 21 | 37 | 24 | 10 |
| 16844 | 147 | 30 | 1 | 0 | 72 | 6 | 16 | 67 | 17 |
| 361 | 37 | 2 | 1 | 0 | 214 | 16 | 2 | 7 | 10 |
| 1051 | 54 | 2 | 0 | 1 | 105 | 8 | 3 | 33 | 5 |
| 29 | 48 | 2 | 1 | 1 | 6 | 0 | 4 | 31 | 10 |
| 1354 | 36 | 2 | 0 | 0 | 2 | 0 | 42 | 31 | 10 |

Table 7: Attribute Vector

Better feature values may occur five times or more in a data set. Performance enables a model to study how this feature value communicates to the label. It has many samples with a similar discrete value offers the model a plan to see the feature in various settings, and determines when it is a better predictor for the class. Table 7 shows the Attribute Vector result after the preprocessing stage.

2) Proposed Feature Selection Algorithms

The proposed model for the attribute selection phase in crime prediction is shown in Figure 2. The figure depicts the processing flow of the feature selection strategy in crime prediction.

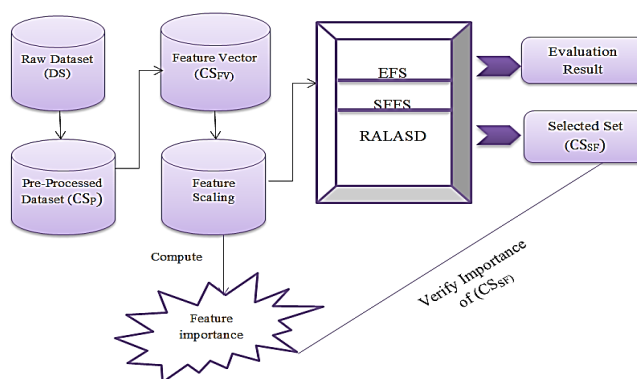


Fig2: Proposed Model for Crime Attribute Selection

i) Exhaustive Feature Selection (EFS): This technique searches all related features from the subset. It locates the best-performing crime feature subset. It creates all the subsets of features from one to ‘N’, with ‘N’ being the number of the feature vector, and for every subset, it accumulates a Deep Learning calculation and chooses the subset with the best performance. The parameters, which can be portrayed as the **minimum** and the **maximum number of features**. It reduces this present technique's computation period if it picks sensible counts for these parameters. The lower correlation between any two features in the crime dataset could help the model to perform high results. The Pre-Processed crime dataset is extracted and converted into the “N x M” Feature vector. This vector gives input to this algorithm to reduce the unwanted feature and select the important feature. The selected feature that could help to achieve higher performance of the classifiers will be discussed in the next chapter. At initial, selecting the non-empty subset from the pre-processed crime dataset.

$$SS = \text{Non } \emptyset \text{ Subset from in CrimeSet}$$

At next, selecting the random tuple (set) from the Non-empty subset. Then, Evaluate the Minimum and Maximum features from the Random Tuples(RT).

$$RT = \text{SelectRandomTuplefrom}(SS)$$

$$MMS = \text{MinimumandMaximum}(RT)$$

Computing the score of between the Minimum and Maximum selected (MMS) features and selects the optimum subset from this. The optimum subset put into the selected subset and the result will be returned

$$\text{Optimum}(\text{Set}) = \text{FindoptimumsetfromMMS}$$

$$\text{FinalSet} = \text{Insertion}(\text{OptimumSet})$$

ii) Sequential Forward Floating Search Algorithm (SFFS): Sequential feature selection algorithms search for an efficient subset of features by aggregating the best features or eliminating the worst features iteratively. At the initial stage, Crime Data extracted from the pre-processed crime Data set (CS_P) in the matrix format (M_{xy}). The next process is converting the Set into the Crime Feature Vector (CS_{FV}). The subset is selected as Random Format and computes the performance for the subset of features.

$$SS = \text{Random Set in}(CS_{FV})$$

Every feature is evaluated with the performance of the classifier.

$$ER = \sum_{i=0}^{SS} \text{Precision of Logistic Regression}(F_i)$$

The Evaluation Result (ER) is based on the scoring the precision of the Logistic Regression Classifier. The highest score of the feature is choosing at starting one to select the next feature.

$$F1 = \text{Top Precision Score of the Evaluation Result}$$

This process is searching to select the next combination of best features in the Set. Again, it computes the Performance and selects the next best one from the pair.

$$\text{Next Feature}(F) = \text{Search Best Performance from the Set}$$

$$\text{Selected Feature} = \text{Next Feature}(X) | \text{Next Feature}(Y)$$

This process while continuing when the desired number of features is chosen by this algorithm. After selecting the number of features (K) this algorithm will stop the work.

iii) Random Assessment of Leveling Attributes based on Synchronized Deviation (RALASD): The proposed feature selection procedure that has involved a significant novel method is the Random Assessment of Leveling Attributes based on the Synchronized Deviation (RALASD). RALASD is effective for working in a large quantity of data difficulties and it's offering a good solution for a comparatively low dimension in the Crime Dataset. Since the operative ascent estimation, this procedure is appropriate for large dimensional issues where many positions are being determined in the optimization procedure. It is clear, in any case, that few algorithms may process better than others may on specific classes of issues because of being able to abuse the difficult structure.

The Pre-processed crime dataset (CSP) is converted in a standardized format and forming as a feature vector. Feature vector formed as a matrix which includes "N x f" that builds a random non-empty subset SF for the important selection of the features.

$$N \in \text{No. of Features and } f \in \text{Features}$$

$$S_F = \text{GenerateRandomSubsetfromFeatureVector}$$

After selecting the subset this algorithm will compute the Assessment value from the random dataset. Then compute the synchronized deviation for the random solution of Ran_Assess () .

$$\text{Ran_Assess}() = \text{ComputetheRandomAssessmentFromtheSubset}(S_F)$$

$$\text{Sync_Dev}() = \text{CalculatetheSynchronizedDeviationof}(\text{Ran_Assess}())$$

Then validating mutually independent features among the subset for flattening the features to the subset selections (CSSF) are computed.

$$\text{if}(\text{Sync_dev}) = \text{MutuallyIndependent}\{\text{return } 1\}\text{else}\{\text{return } 0\}$$

The independent attributes forming level and add to the selected subset CSSF with the accuracy of the classifier. The algorithmic representation for the RALASD has appeared in Algorithm 2.

*Accuracy = Classification(MutuallyIndependentFeature)
SelectedSubset = High \uparrow AccuracyoftheFeature*

RALASD() Algorithm

```
{  
Input: CSFV – Crime Feature Vector  
Output: CSSF – Selected Feature Set  
Step-1: Read the feature vector CSFV  
Step-2: The vector CSFV includes n x f for the dataset // ‘f’  
features and ‘n’ interpretations  
Step-3: For each f` in f  $\forall$  n  
Step-4: Initialize the non-empty subset SF  $\in$  CSFV  
Step-5: The Subset SF includes feature set with ‘j’ features  
Step-6: Compute RAS = Ran_Assess (f`, j) // random  
assessment for the features in the subset  
Step-7: Initialize the random solution RAS  
Step-8: Then compute Sync_Dev (RAS (i), j) // computation of  
synchronized deviation for random solution  
Step-9: Validate that it is mutually independent or not  
Step-10: Each features of SF must generate from a symmetric  
zero  
Step-11: Then set the probability deviation for the features  
Step-12: The deviation must have a finite inverse to other  
features for leveling  
Step-13: The deviation with ‘1’ must be a mutually  
independent with other  
Step-14: The independent attributes are leveled and add to the  
selected subset CSSF with the accuracy of the classifier  
Step-15: Return CSSF  
Step-16: End for  
}
```

Algorithm 2: RALASD Algorithm

3) Prediction Layer

Deep learning is a kind of ML technique that makes a computer do the work as humans, such as speech identification, predict images, and crime. Deep learning is one of the methods in AI technique. Deep Learning frameworks have expressively more info accessible to accumulate neural networks with numerous deep layers for predicting crime. Fig3. shows the proposed model for the classifier and it is described as follows.

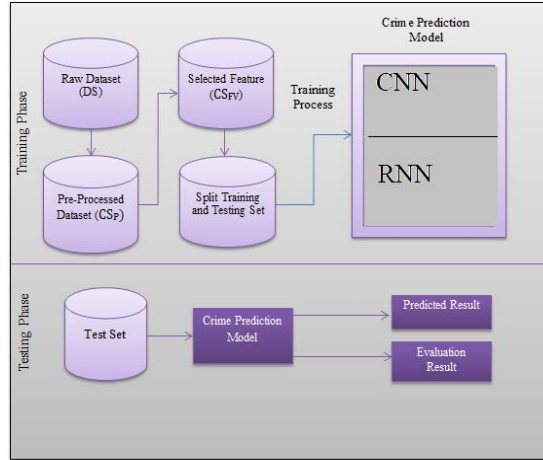


Fig 3. Proposed Model for the Crime Prediction

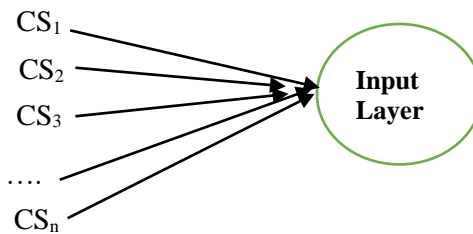
i) Convolution Neural Network: Classification as a well-known data mining supervised learning method is used to extract meaningful data from huge datasets and can be strongly used to predict unknown classes. Neural Networks, as another classification strategy, are a nonlinear model, which can model real-world complex connections. Neural networks could evaluate the posterior probabilities, which give the reason for setting up classification rules and accompanying statistical analysis. This classifier level was constructed with Convolutional Neural Network (CNN) algorithm furnished with a modular structure. By empirical test, the system found that even the traditional CNN is more promising when contrasted with traditional methodologies.

A CNN is a multilayered neural network with a unique design to identify complex features in crime data. CNN's have been utilized in crime detection, image recognition, robotics, and self-driving vehicles. When a CNN is fabricated, it tends to be utilized to order the substance of various perspectives. All the techniques needed to be done do is feed crime information into the model. Just like ANNs, CNNs are motivated by the functions of the human brain. CNN can classify crime by detecting features, like how the human mind recognizes highlights to recognize objects. To construct a multilayer convolutional network from the model that generated so far and require trailing certain phases such as weight initialization, convolution and pooling, Initial and next convolution layer, train and evaluate.

The preprocessed dataset (CSP) output is given as an input for the Feature Selection (FS) algorithm after applying that it returns the best-selected result. The output of the selected features (CSSF) is converted into the training set (CS_{Train}) along with the labels (Class). Training set with the Class variable is sliced into training and testing set with the ratio of 80% and 20%. Then apply for the prediction using CNN. The CNN is initialized as three-layer in the following manner.

a) Input Layer: It is the first layer of the Crime Neural Network and it is called as starting Neurons of Crime Prediction also. It takes input values (Crime Features) and passed them on to the next layer. Fig 4. shows the Input layer model.

$$\text{Input Layer}(I) = CS_{Train} \in CS_1, CS_2, CS_3, \dots CS_n$$



Input Neurons

Fig 4. Input Layer of CNN

*b)Hidden Layer:*This layer separates the network into precise transformations of the crime data. Each hidden layer function is specialized to produce a defined output. In this work this layer process to identify the crime type and location that utilized in conjunction by subsequent hidden layers to predict the crime in the Data set. Fig 5. shows the hidden layer.

$$\text{Hidden Layer} = \frac{\text{No. of Input Layer} + \text{No. of Output Layer}}{2}$$

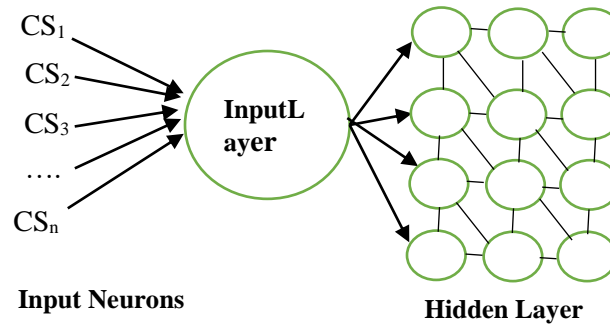


Fig 5. Hidden Layer of CNN

*c) Output Layer:*This layer produces the output for the Testing result. There must be a minimum of one output layer in this network but here two output layers are available.

$$\text{Output layer} = 1 \text{ or } 0$$

The output layer receives the input from the hidden layer and it produces the result. Fig 6. shows the output layer of CNN

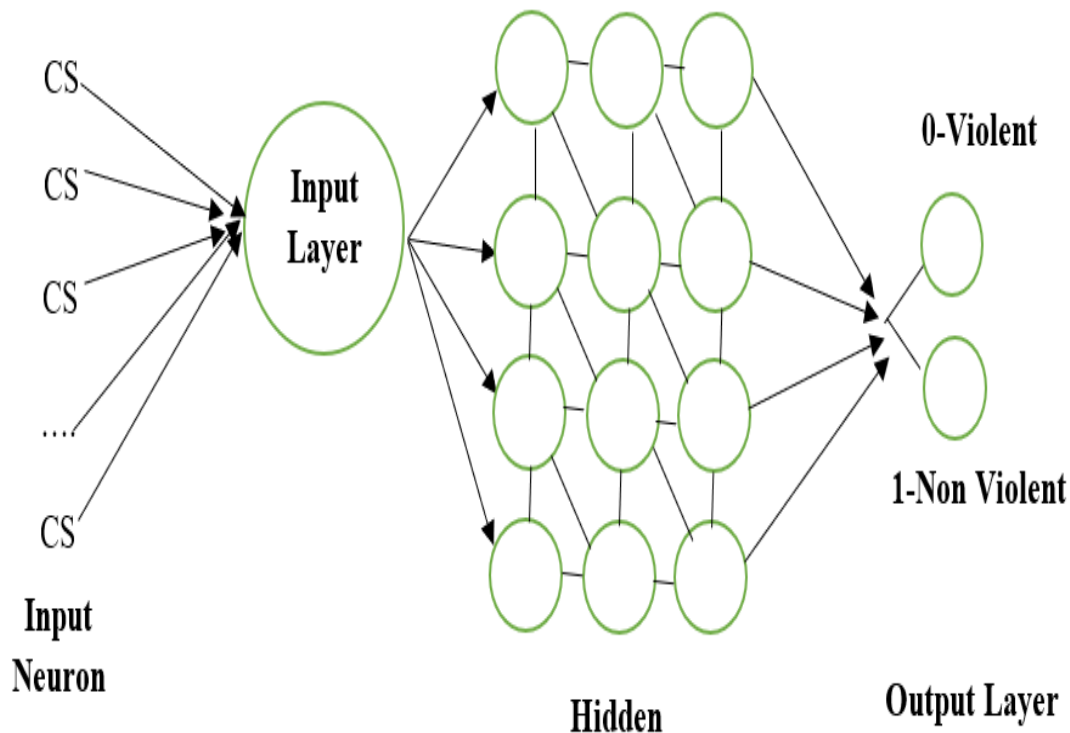


Fig 6. Output Layer of CNN

```

Algorithm CNN ()
{
1: Load the data CSSF
2: Load Training Set CSTrain and Testing Set CSTest
3: Initialize the convolutional layer with Conv_ID, kernel, AF,
input and batch_norm (3,'RELU',128)
4: Initialize the pooling layer 1 for the training set
(Max_Pooling_ID,pool_size='2')
5: Initialize the weight with random value for CNN based
learning network
6: Initialize the pooling layer 2 for the training set
(Max_Pooling_ID,pool_size='2')
7: Then randomly select the training set
8: Perform the flatten with the activation function
9: Dropout()
10: Perform the dense with num_class and AF='Softmax'
11: Compute the output for each layer
12: Find the activation rate of output nodes
13: Performing the optimizer 'adam' and lossfunction 'Binary
Cross Entropy'
14: Recalibrate process till merging criterion is met
15: Repeat the process till the merging criterion is met
16: Return Class Labels
}

```

Algorithm 3: CNN Algorithm

Finally, the method needs to fit the model to the training dataset and test its performance with the test set. The CNN achieves this by calling the fit generator function on the classifier object. The first argument it takes is the training set. The second argument is the number of arguments in the training set that want to use to train the CNN, and finally, it returns the output.

ii. Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) indicates to a particular design of an artificial neural network that functions admirably for subjective grouping datasets. A particularly RNN comprises of cyclic associations that allow the neural network to all the more likely model arrangement information contrasted with a customary feed forward RNN. The vital thought of these cyclic associations or loops is that they permit data to endure while training. At the end of the day, those loops and cyclic associations allow the organization to pass data starting with one-phase then onto the next iteration. A RNN deals with the standard of saving the yield of a specific layer and taking care of this back to the contribution to request to anticipate the yield of the layer.

The RNN work in three phases. In the main stage, it pushes ahead through the LSTM layer and makes a prediction. In the subsequent stage, it contrasts it's predict and the optimal value utilizing the loss function (LF). LF shows how well a model is performing. The lower the value of the LF, the better is the model. In the last stage, it utilizes the accuracy in back-propagation, which further ascertains the inclination for each point. The gradient is the value used to change the loads of the network at each point.

The RNN classifier model is designed with the given crime dataset. The CSF selected feature set is employed to the model. The set is split into training and testing set for network layer construction. The LSTM layer is build with the input units (15) and dropout (0.2) with the input layer. The dense layer is constructed with the back-propagation of the LSTM layer. The optimizer (ADAM) is employed with loss function to obtain the predicted class label.

```

Algorithm RNN ()
{
  1: Obtain the data  $CS_{SF}$ 
  2: Split Training Set  $CS_{Train}$  and Testing Set  $CS_{Test}$ 
  3: Construct the labels and features for the input
  4: The layers are built with Embedding, LSTM and Dense Layers
  5: The embedding layer are load with the input sequences
  6: The LSTM layers are added to the network with input units (15) and dropout (0.2)
  7: Performing the dense layer for the given units of input sequence
  8: Then train the model with  $CS_{Train}$  in sequence
  9: The predictions are made by passing the sequence to the model with  $CS_{Test}$ 
  10: Return the predicted class label
}
    
```

IV. Performance Evaluation

This section includes the performance evaluation for the employed feature selection strategies and Classification with the performance metrics such as accuracy, sensitivity and specificity.

A) Elapsed Time of Feature Selection Algorithms

Elapsed time is the measure of time that passes from the earliest starting point of the choice to its end. The assessment of the elapsed time taken by the FS methods while the selection is shown in Table 8. The table shows that the RALASD method ingests less time while choosing.

| ID | FeatureSelection Methods | Elapsed Time |
|----|--------------------------------|--------------|
| 1 | Exhaustive Feature Selection | 10 Sec |
| 2 | Step-Forward Feature Selection | 9 Sec |
| 3 | RALASD | 7 Sec |

Table 8: Evaluation of Elapsed Time

The evaluation of the elapsed time (in a sec) for the feature selection methods is shown in Fig 7.

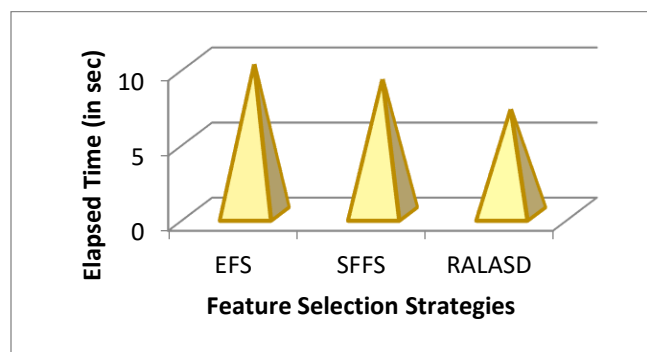


Fig 7. Elapsed Time

B) Expected and Predicted Result of Classifiers

Predicted values are simulations that take the estimation uncertainty and the fundamental uncertainty into account. They are in the same metric as the dependent variable. Expected values average over the fundamental uncertainty and thus only represent the estimation uncertainty.

Crime = 0 ∈ Positive Class & No – Crime = 1 ∈ Negative Class

| | |
|--|---|
| <p>True Positive (TP): Reality: A Crime found in Training Set. Testing Result: "Crime." Outcome: Correct Result.</p> | <p>False Positive (FP): Reality: A Crime not found in Training Set. Testing Result: "Crime." Outcome: The result is wrong.</p> |
| <p>False Negative (FN): Reality: A Crime found in Training Set. Testing Result: "No Crime." Outcome: The result is confusion.</p> | <p>True Negative (TN): Reality: A Crime not found in Training Set. Testing Result: "No Crime." Outcome: The result is fine.</p> |

Table 10: Confusion Matrix for Crime Prediction Dataset

The crime prediction dataset is represented as a confusion matrix is depicted in Table 10. The table describes TP, FP, FN, and TN for the crime prediction model.

C) Accuracy (ACC)

Accuracy is main metrics in the classification algorithms to find the best in a test of crime dataset. Accuracy calculates based on the correct prediction is divided by the total prediction.

$$ACC = \text{Correct Prediction} / \text{Total Prediction}$$

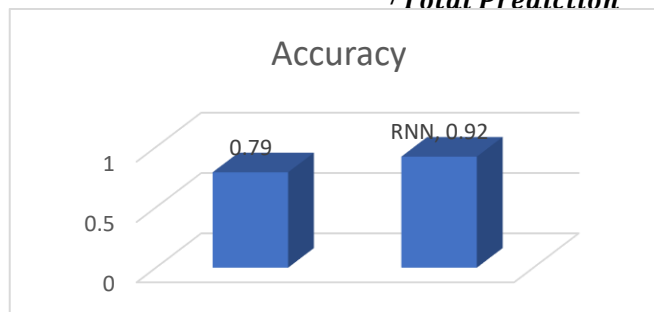


Fig 8. Accuracy Evaluation

The accuracy evaluation for the classifiers is depicted in Fig 8. The figure illustrates that the proposed CNN provides a high rate of accuracy while predicting.

D) Classification Error (CE):

This classification error computes the count of wrong predictions in the test set which is separated by the whole predicts constructed in test of crime set.

$$CE = 1 - Accuracy \equiv \frac{FP + FN}{TP + TN + FP + FN}$$

The less classification error denotes the best prediction of the crime set.

Best Prediction = Less Classification Error

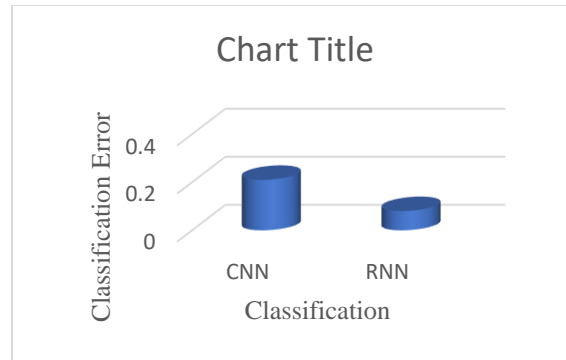


Fig 9. Classification Error Evaluation

The classification error evaluation for the classifiers while prediction is shown in Figure 9. The proposed CNN possesses less error rate than other methods.

E) Sensitivity

The sensitivity of a crime test is defined as the proportion of the dataset with the crime, which has a positive result (Crime is Available).

$$CrimeSensitivity = CS_{Train} \equiv Crime \wedge Outcome \equiv Crime$$

A test that is 100% sensitive will identify the dataset, which has the crime. Any clinical test is rarely 100% sensitive. A highly sensitive test can be useful for ruling out a crime if a dataset has a negative result.

$$Sensitivity = True\ Positive\ Rate \equiv \frac{TP}{FN + TP}$$

F) Specificity

The specificity of a Crime test set is the proportion of data without crime that has a negative result (Crime is Not Available).

$$Crime\ Specificity = CS_{Train} \equiv No\ Crime \wedge Outcome \equiv No\ Crime$$

A test that has 100% specificity will identify 100% of the dataset, which does not have the crime. Tests with high specificity are most useful when the result is positive. A highly specific test can be useful for ruling in the dataset, which has a certain crime.

$$Specificity = True\ Negative\ Rate \equiv \frac{TN}{TN + FP}$$

the sensitivity and specificity ratio evaluation for the classifiers. The sensitivity and specificity ratio for the proposed CNN is high then other methods.

| Method | Precision | Recall | Sensitivity | Specificity | Accuracy | CE |
|--------|-----------|--------|-------------|-------------|----------|------|
| CNN | 0.7 | 0.7 | 1.0 | 0.9 | 0.79 | 0.21 |
| RNN | 0.9 | 0.9 | 1.0 | 0.9 | 0.92 | 0.08 |

Table 11: Evaluation Metrics for Classifier

V. CONCLUSION

Crime analysis incorporates for identify and predict crimes and its types and their relationships with criminals. The high volume of crime datasets and the complication of relationships among these varieties of data have through criminology field for applying data mining techniques. Crime analysis is the analytical process of interpreting the specific features of a crime and related crime scenes. The crime sequence and the patterns which several criminals follow when committing a crime make it easy for analyzing the crime. This process includes several procedures to be followed to identify the criminals and getting more information based only on the clues or

information given by the local people. The pre-processing steps like missing value removal, duplicate removal, encoding, etc., are described in this chapter. Then the chapter includes a clustering model for the crime profile construction is designed. Then the crime profile is constructed with the crime types and the subset selection is performed to obtain the significant crime variable for the prediction. Then the training set is included in the testing model for the test phase for the prediction.

REFERENCES

1. Chen, H., Chung, W., Xu, J., Wang, G., Qin, Y., Chau, M.: Crime data mining: a general framework and some examples. *Computer* 37(4), (2004)50–56.
2. Adedayo M. Balogun ; Tranos Zuva , "Criminal Profiling in Digital Forensics: Assumptions, Challenges and Probable Solution", *International Conference on Intelligent and Innovative Computing Applications (ICONIC)2018*
3. Matthew Tonkin, Jessica Woodhams, Ray Bull, John W. Bond, and Emma J. Palmer. Linking different types of crime using geographical and temporal proximity. *Criminal Justice and Behavior*, 38(11):1069–1088, 2011.
4. Exploring Foursquare-derived features for crime prediction in New York City Cristina Kadar, José Iria, Irena Pletikosa Cvijikj
5. Mr. Ravikumar B, Mohamed Fahad Ali Abbas J, Githendra Vishal B, Vasudevan T, "Crime Incidents Detection Using Support Vector Algorithm", *VDGOOD Journal of Computer Science Engineering*
6. Ahishakiye E, Taremwa D, Opiyo E, Niyonzima I (2017) Crime prediction using decision tree (J48) classification algorithm. *Int J Comput Inf Technol*. 06(03) ISSN: 2279-0764
7. Charuni Rajapakshe ; Shashikala Balasooriya ; Hirumini Dayarathna ; Nethravi Ranaweera ; Namalie Walgampaya , "Using CNNs RNNs and Machine Learning Algorithms for Real-time Crime Prediction",2019 International Conference on Advancements in Computing (ICAC)
8. Mashnoon Islam ; Redwanul Karim ; Kalyan Roy ; Saif Mahmood ; Sadat Hossain ; Rashedur M. Rahman, "Crime Prediction Using Multiple-ANFIS Architecture and Spatiotemporal Data",2018 International Conference on Intelligent Systems (IS),1541-1672.
9. J. Prakash and P. K. Singh, "Particle swarm optimization with k-means for simultaneous feature selection and data clustering," in *2015 Second International Conference on Soft Computing and Machine Intelligence (ISCM)*, November 2015, pp. 74–78
10. A. Buczak and C. Gifford, 'Fuzzy association rule mining for community crime pattern discovery', in *ACM SIGKDD Workshop on Intelligence and Security Informatics*, Washington, D.C., 2010, pp. 1–10.