USING MACHINE LEARNING AND RECURRENT NEURAL NETWORKS FOR EFFICIENT CRIME DETECTION AND PREVENTION

[1]Souglman Kabirou Kanlanfeyi, [2]Keshav Kishore
[1]Post Graduate Student, [2]Assistant Professor
[1]Departement of Computer Science and Engineering,
[1]AP Goyal Shimla University, Shimla (HP), India

Abstract: Defined as an action or omission which constitutes an offence and is punishable by law, the crime is a phenomenon that is the heart of the authorities and whose measures are very adequate are taken to reduce or stop. Some predictions can be made in order to detect and prevent by using machine learning algorithms such as Regression algorithms and Classification or by using Deep Learning with Recurrent Neural Networks. In this document, after doing a multivariate analysis of different important features affecting occurrence of crime, we will first use a classic machine learning classification algorithm which is Random Forest and then introduce a classification model based on Neural Networks with the Keras framework and its KerasClassifier wrapper provided by Google. In order to increase the quality of the network by find the optimum combination of Hyperparameters, Grid Search CV method will be used. The Neural Network architecture that will be considered is the Recurrent Neural Network with the Long Short Term Memory technique. The objective is to compare the performances of these techniques used for the prediction of crimes by giving a comparison table and difference confusion matrices. The output of our study showed that Deep Learning Architectures especially LSTM techniques takes the lead over other ones on predicting type of crime by considering date and location information.

Keywords: Machine Learning, Deep Learning, Crime Prediction, Neural Networks, Recurrent Neural Networks, LSTM

I. Introduction

We are focusing only on supervised learning especially on classification algorithms because the data we are going to use is already labeled and processed, we are just going to classify each input data in a specify category. We can take the commonly used example where algorithms are used to classify a new mail received as a spam or not. These techniques can be applied for more real-world problems like classify a picture as a dog or a cat. As classification algorithms, we have Naïve Bayes, Support Vector Machine, Random Forest… In order to deal with big amount of data and to have good results for prediction, we will use Deep Learning which is a subset of Machine Learning which works with Artificial Neural Networks. Since there are different types of architectures in Deep Learning design for some specific purposes, we will use Recurrent Neural Networks which are used for time series problems.

A. Machine Learning

Machine Learning is a subset Artificial Intelligence which consist on learning from experience (data) without being programmed. As defined previously, we can remark that the essential element of machine learning is data which is trained commonly trained for a future use. Since ML is based on learning from data, we have 3 types of learning in ML:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

In our study, we will focus on Supervised Learning which can be defined as learning from data which is already known. In other words, supervised learning is learning from data which is already labelled. In this type of learning we have 2 type of techniques which are: regression and classification.

Regression consist on predicting a number from previous data. For example based on an existing dataset of salaries, we can predict what an employee can receive according to many factors like level of study, number of years of experience, post and on… Briefly, regression consists on predicting numerical value. In another hand, classification consist on predicting the class of an input based on previous observations. An example is to classify the class of an animal image whether to predict if it's a dog or a cat in case of binary classification or to predict the class of an image

The supervised technique used in this paper is classification especially multiclass classification where are predicting the class (type of crime) by providing different types of input.

B. Deep Learning

DL is a subset of AI based on functional of human neural networks called Artificial Neural Networks. DL was adopted due to the fact that some ML techniques became to perform well on some complex problems and also due to the increase of data amount nowadays. As said previously, most DL architectures are based on ANN they are also designed for some specific task like CNN for Image based problems, RNN for time series… How does ANN work?

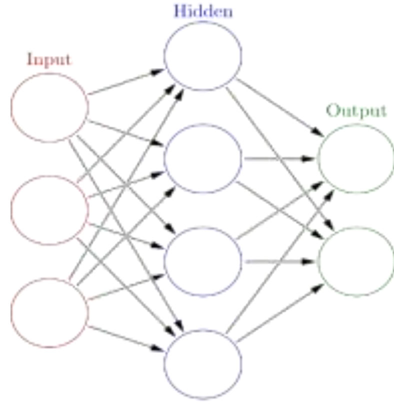

Fig 1 ANN architecture

We have three types of layers as shown on the figure Fig.1 which are:

Input layer: receives input data. An example of input can be the number of rooms, facilities, and distance to the downtown... in the case of price prediction of a house. The input layer passes the inputs to the first hidden layer.

Hidden layers: perform mathematical computations on our inputs. One of the challenges in creating neural networks is deciding the number of hidden layers, as well as the number of neurons for each layer.

Output layer: returns the output data. In our case, it gives us the prediction of the price of the house.

The process of prediction the price of the house is done as followed:

- Each connection between neurons is associated with a weight. This weight dictates the importance of the input value.
- The initial weights are set randomly.
- When predicting the price of a house, the number of rooms should be one of the heavier factors.
- Hence, the number of rooms' neuron connections will have a big weight.

Here is an illustration of the key difference between ML and DL

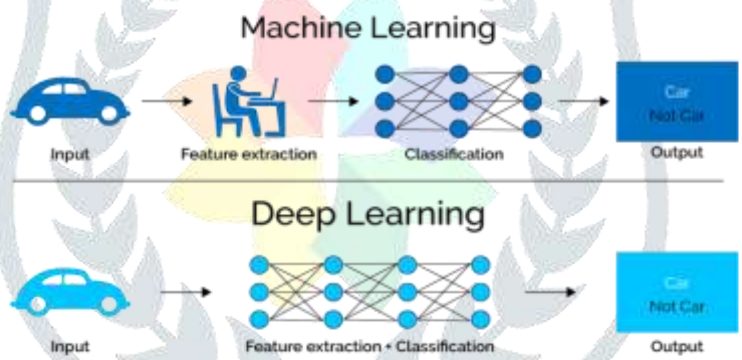

Fig.2 Machine Learning Vs Deep Learning

C. Ensemble Learning

As it's said commonly, Unity is Strength, the intuition behind this technique can be guessed by the name; ensemble learning consists basic on making combination of many based models to have a good performance by improving the results of based models. While Random Forest is known as the most Ensemble Learning technique used in Data Science, we have other various types but this last one is the best one why trying to explain what Ensemble Learning is. Where a decision tree is a technique based on questions and conditions to predict a certain value or class, in Random Forest these trees are combined together to form a Forest of Trees called Random Forest. We have to say that this Ensemble technique can be used for both Regression and for Classification called RF Regression and RF Classification.

D. Hyperparameters Tuning

Commonly used in DL but also in ML, this technique is used to identify the best combination of Hyperparameters needed for the training of a model but what are Hyperparameters? Defined as parameters that are already known and initialized before the training step. The main difference between these parameters and the other one is that these ones are used to get others. For example, in DL we have the optimizer, number of epochs, batch size.

E. Callbacks

The callbacks are very essential in Deep Learning architectures, they are used to improve the efficiency of a networks by returning information on training process and by providing powerful visualization. We have different callbacks when they popular ones are: EarlyStop, ModelCheckPoint, LearningRateScheduler, TensorBoard.

Other callbacks are: RemoteMonitor, CSVLogger, BaseLogger and History, LambdaCallback.

F. RNN with Long Short Term Memory

It has been said previously that RNN are facing some issue of vanishing of gradient descent which is basically explained as the difficulty of RNN to deal with long dependency states. For example, it's very easy for classic RNN to predict a current state by using the previous one but when it comes a little difficult of handling long-term dependencies. There comes LSTM networks which are specialized on remembering information for a long time and to solve problem of exploding and vanishing of gradient.
LSTM is used to tackle problem of long dependency by retaining information for a long time which basically contains four types of NN with different memory block called cells.

II. Related Work

Since the application of AI in many fields such as crime prediction, many researches have focused on prediction of crime classification using different factors such as time and location as shown in papers [1] and [2] by using different ML classifications algorithms like SVM, RF, Decision Trees, KNN… Due to the complexity of some tasks and preprocessing done by the engineer and the huge amount of data nowadays, ML techniques have been shown to performs bad on these kind of data where DL techniques performs well so it has been applied to these kind of data and perform better than ML as shown in papers [6] and [14]. Since crimes an occurrence of crime can be related to previous one and happen over times, researchers noticed that it will be preferable to predict crime occurrence by using a DL architecture designed for these kinds of problems called RNN as shown in papers [4] and [7]. In order to optimize training process by choosing the optimal combination of Hyperparameters that will be used to train our data as shown in paper [11]. Except time and location information, it has been shown that other factors can affect crime occurrence such as unemployment and illiteracy as shown in paper [10].

III. Proposed Work

The main purpose of our study will be to efficiently predict Crime by using Machine Leaning and Deep Learning techniques such as machine learning classification algorithms and Recurrent Neural Networks.

1. Workflow

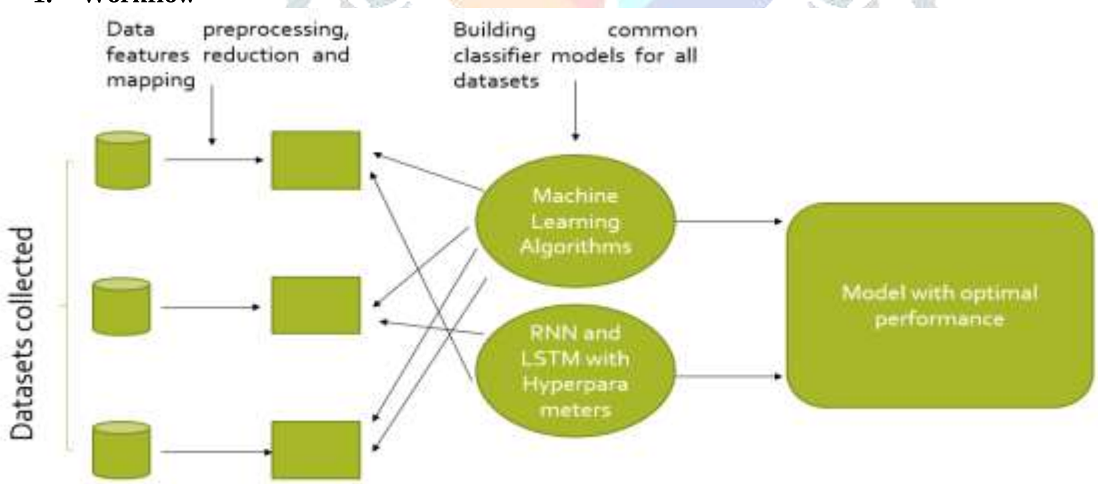

Fig.3 Workflow

2. Different Steps

Step1: Data Collection and Analysis

The first step of this process is to collect data from different open data platform such as Kaggle or data.gov… To ensure integrity and equality, most of attributes of these different dataset are commonly same and are from the same type. These attributes are: Category of crime, Type of crime, Occurrence Date, Location of Crime (Latitude and Longitude), Neighborhood Identification…

Step 2: Data Preprocessing

This step consists on preparing data for training process. This step includes dropping missing data, data formatting, data conversion, data reduction. After converting data into appropriate on for training process, data is normalized and scaled using different techniques depending on the architecture. For example:

- Standar Scaling is used for ML and ANN tasks
- For RNN it's preferable to use MinMax Scaling technique

Step 3: Building of different classification models

After the sets created, we will build common classification models for all the datasets by using Machine Learning and Deep Learning techniques.

Step 4: Measurement of models performances

After classification models built, we will fit to them the training sets that were created previously. Testing set will used for prediction to see the performance of every algorithm on the testing sets. The different performance metrics that will be consider in our study are: accuracy, precision, f1 score and the recall. All these performance metrics can be calculated by using the confusion matrix which is used to summarize the prediction results of a classifier by giving the number of correct and incorrect predictions by a classification model. It can also be considered as a table which is divided into 4 main parts in a case of a binary class (0 or 1) prediction as follows:

- True Positive: Prediction is 1 and Observation is 1
- True Negative: Prediction is 0 and Observation is 0
- False Positive: Prediction is 1 but Observation is 0
- False Negative: Prediction is 0 and Observation is 1

Let's summarize it in a table

	Class 1 predicted	Class 2 predicted
Class 1 observed	TP	FN
Class 2 observed	FP	TN

Table 5.1 Confusion Matrix 2x2 Example

Note that in a case of multiclass classification, we will get N x N confusion where N corresponds to the number of classes that we want to classify.

IV. Datasets

Our study will focus on two different datasets from Denver and Seattle all cities from USA. The models with the same parameters built will be applied on both datasets preprocessed. We ensure that most attributes of the datasets are common and of the same type such as time and location attributes. We have:

Denver Crime Dataset provided in a csv format by Kaggle and that can be also find on the Denver Open Data Catalog which contains the report of all the incidents and crimes that occurred in the city of Denver in USA from 2013 to now.

Seattle Police Reports Dataset: also provided in a csv format, this data represents crime reported to the Seattle Police Department (SPD) and which can be accessed on Kaggle. Each row contains the record of a unique event where at least one criminal offense was reported by a member of the community or detected by an officer in the field.

V. Algorithms and Technologies used

By choosing Python 3.67 as programming language because its stability, we decided to go for libraries like scikitlearn for ML techniques and TensorFlow 2.0 in which Keras is already included for DL architectures. For tuning the Hyperparameters, we used GridSearchCV and to optimize our different networks we applied different callback such as EarlyStop, LearningRateScheduler and ModelCheckPoint. The different algorithms used are:

- Random Forest

Considered as an ensemble learning algorithm because it's a combination of based model which is decision tree algorithm to produce an optimal prediction result. It's used for both regression and classification know respectively as Random Forest Regression and Random Forest classification. Unlike some machine learning classification algorithms, Random Forest can be used for multi class classification.

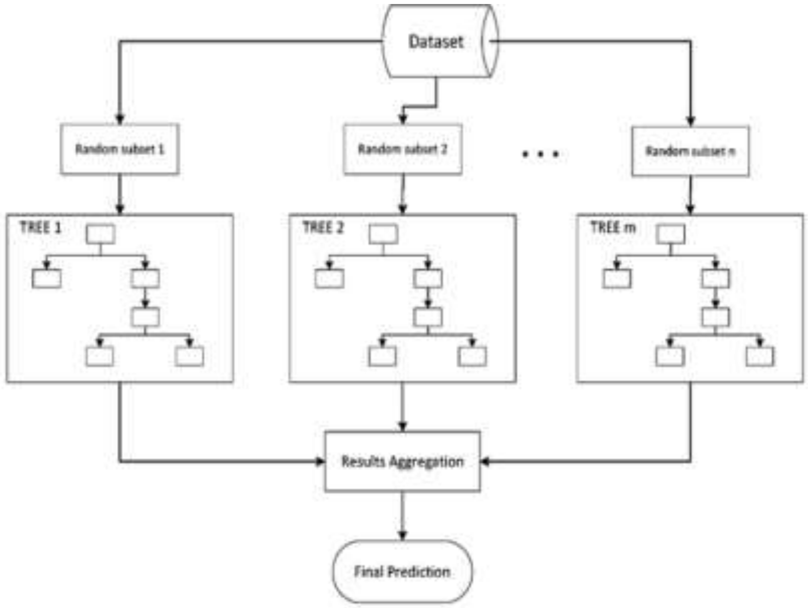

Fig.4 Random Forest Structure

Here we can remark that first we have the original dataset which is splitted in many random subsets on which decision trees algorithm is applied on each subset. At the end, an aggregation function is done on all the results of each subset previously got.

- Artificial Neural Networks

ANN is the simplest architecture of Deep Learning where we have three types of layers (input, hidden, output layers) which have each a numbers of neurons. The number of neurons in the input layers is equal to the number of independent variables and the number of neurons in the output one is equal to the number of instances or class to predict (the dependent variable). In the case of hidden layer, we have to say that we can have more than one hidden layer. It's called Deep Neural Network when we have more than one hidden layer. The number of neurons in these layers is arbitrary.

- LSTM

It has been said previously that RNN are facing some issue of vanishing of gradient descent which is basically explained as the difficulty of RNN to deal with long dependency states. For example, it's very easy for classic RNN to predict a current state by using the previous one but when it comes a little difficult of handling long-term dependencies. There comes LSTM networks which are specialized on remembering information for a long time and to solve problem of exploding and vanishing of gradient.
LSTM is used to tackle problem of long dependency by retaining information for a long time which basically contains four types of NN with different memory block called cells.

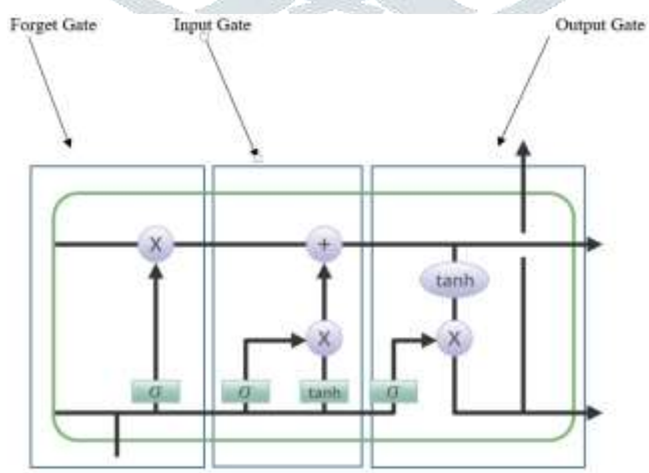

Fig.5 Different types of gates in LSTM

Retention of information and the manipulation of memory is managed by three kind of gates which are:

- Forget Gate

It's used to remove the information which is already used in the first cell state or which is no longer useful. At this step, we have two types of inputs x_t and h_t-1 which are respectively the input at the current time and the output of the previous cell which are been given to the gate and multiplied with weight matrices and then an addition of bias. Information is forgotten if a particular cell output is

0 and retained if it is 1 after the resultant of the previous inputs (x_t and h_t-1) is has been passed through an activation which gives a binary result.

- Input Gate

This gate is used to add the information to the cell gate which is being regulated through a sigmoid activation function by filtering values that are going to be remembered like with the forget gate with the same input (x_t and h_t-1). We then have the creation of a vector by using the tanh function which gives output in a range of -1 to 1 containing possible values of x_t and h_t-1. Finally, a multiplication of the vector's values and the regulated values with the sigmoid function is done to produce a useful information.

- Output Gate

It is used for extracting useful information from current cell state in order to present it as output.

VI. Result Analysis

The results of each classifiers built we be compared based on these performance metrics which are calculated using the confusion matrix.

1. The different performance metrics chosen are:

- Accuracy

It is equal the ratio of the total number of correct predictions (TP + TN) to the total number of samples. In the case of multi-class classification, it's preferable to use the function *accuracy_score* to get the overall accuracy on the whole samples. The formula to calculate it is:

Accuracy = TP + TN / (TP + TN + FP + FN)

- Precision

Precision is considered as the ability or the precision of the classifier to predict positive instances.

Precision = TP / TP + FP

- Recall

Recall of is defined as how many of True Positives were correctly predicted

Recall = TP / TP + FN

- F1 Score

Used to measure test accuracy, it's defined as the weighted harmonic mean of the test's precision and recall.

F1 Score = 2 (Precision * Recall) / (Precision + Recall)

2. Classification Results

This section will be divided into two parts since we are using two different datasets.
Each dataset were splited as followed:

Technique	Training Set	Validation Set	Testing Set
Machine Learning	0.75	NA	0.25
Deep Learning	0.5625	0.1875	0.20

Table.1 Splitting of the different Datasets

✓ Denver Dataset

Algorithm	Accuracy (%)	Precision (%)	F1 Score (%)	Recall (%)
ANN	99.24	97.34	98.01	98.87
RNN (LSTM)	99.7	99.25	99.28	99.35
RF	83.45	99.20	88.47	85.04

Table.2 Classifiers results on Denver Dataset

On this dataset, we remark that RNN classifier performs just little better than ANN classifier with an accuracy of 99.7% against 99.24% for ANN. Random Forest classifier comes last with an accuracy of 83.45%. This ranking is the same for all the performance metrics except the precision where RF performs better than the other classifiers with a precision of 99.20%.

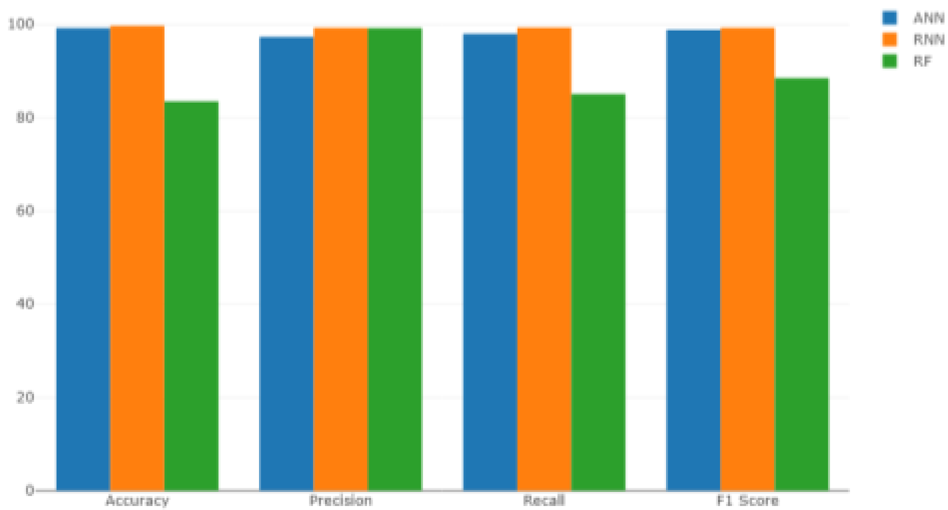

Fig.6 Comparison of performance of different classifier (Denver)

✓ Seattle Dataset

Algorithm	Accuracy (%)	Precision (%)	F1 Score (%)	Recall (%)
ANN	92.48	92.36	85.33	84.60
RNN (LSTM)	93.31	93.56	93.34	93.31
RF	75.56	96.43	68.95	59.17

Table.3 Classifiers results on Seattle Dataset

On this dataset, we remark the same situation as on Denver dataset where RNN performs better than other classifiers with an accuracy of 93.31% against 92.48 and 75.56% for ANN and RF. We remark also that for the precision, RF performs better than others with a precision of 96.43% against 93.56% and 92.36% for RNN and ANN.

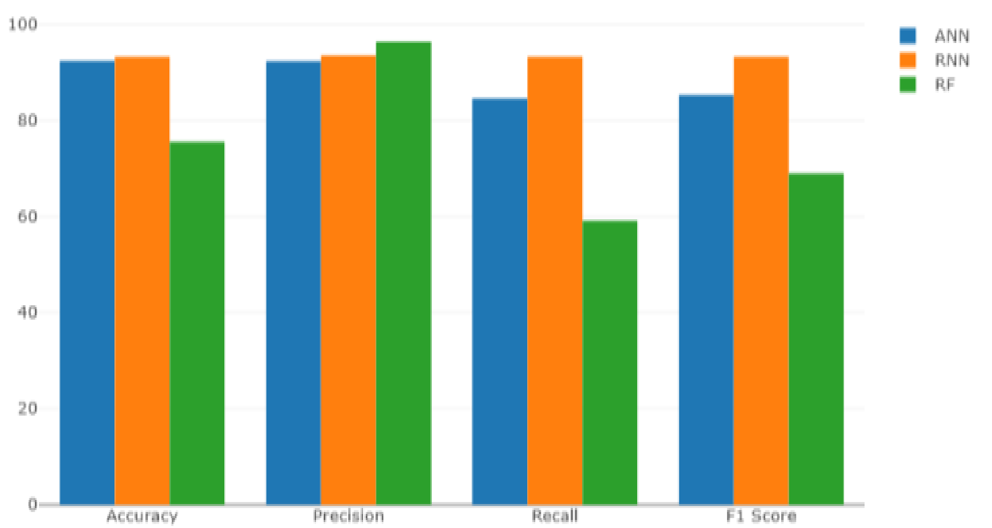

Fig.7 Comparison of performance of different classifier (Seattle)

VII. Conclusion and Future Scope

This study helped us to figure out how AI can be applied in most of the fields in order to predict many facts such as crime prediction. It has shown in many papers that ANN is known to be very efficient in this type of situation but since RNN have been adopted for crime prediction, most papers and studies in this domain have shown that the different types of architectures of DL designed for time series problems take the lead on ANN. We also remark the same situation in our implementation part were LSTM models performs well over other models by considering our different models respectively on Denver and Seattle datasets with an accuracy of 99.7% and 93.31%. In second place comes ANN models which score an accuracy of 99.24% and 92.48%. RF model comes last in predict type of crime in our comparative analysis with an accuracy of 83.45% and 75.56%. In conclusion we can say that, DL architectures in global take the lead on machine learning algorithms in these kinds of situation where we have large amount of data and with complex data.

In the future we will consider not only time and location factors but also provide other factors that can affect crime occurrence such as social factors like the sex, race, age, religion, marital status, financial status even demographic and meteorological data.

REFERENCES

[1] Ying-Lung Lin, Liang-Chih Yu, Tenge-Yang Chen: Using Machine Learning to Assist Crime Prevention. 6th IIAI International Congress on Advanced Applied Informatics, 2017.
[3] Ashwini B. Bais, Rajesh Babu, Jayant Adhikari: An Intelligent Analysis of Crime through Newspaper Articles – Clustering and Classification. International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 2, 2017.
[2] Nishat Shama and Dr. Mahbub Alam Majumdar: A Machine Learning Approach to Predict Crime Using Time and Location Data, 2015.
[4] Yong Zhuang, Matthew Almeida, Melissa Morabito, and Wei Ding: Crime Hot Spot Forecasting: A Recurrent Model with Spatial and Temporal Information. IEEE International Conference on Big Knowledge, 2017.
[5] Amit Gupta, Ali Syed, Azeem Mohammad, Malka N.Halgamuge: A Comparative Study of Classification Algorithms using Data Mining: Crime and Accidents in Denver City the USA. (IJACSA) International Journal of Advanced Computer Science and Applications, 2016.
[6] Trung T. Nguyen, Amartya Hatua, and Andrew H. Sung: Building a Learning Machine Classifier with Inadequate Data for Crime Prediction. Journal of Advances in Information Technology Vol. 8, No. 2, May 2017.
[7] Panagiotis Stalidis, Theodoros Semertzidis and Petros Daras: Examining Deep Learning Architectures for Crime Classification and Prediction. IEEE, 2018.
[8] Hyeon-Woo Kang, Hang-Bong Kang: Prediction of crime occurrence from multimodal data using deep learning. PLoS ONE 12(4):https://doi.org/10.1371/journal.pone.0176244, 2017.
[9]Ying-Lung Lin, Meng-Feng Yen and Liang-Chih Yu: Grid-Based Crime Prediction Using Geographical Features. International Journal of Geo-Information, 2018.
[10] Luiz G.A. Alves, Haroldo V. Ribeiro, Francisco and A. Rodrigues: Crime prediction through urban metrics and statistical learning. ELSEVIER, 2018.
[11]Tianxiang Gao: Hyper-parameter Optimization for Deep Learning, 2016.
[12] Nikhil Ketkar: Deep Learning with Python, DOI 10.1007/978-1-4842-2766-4_7, 2017.
[13] Shen Ting Ang, Weichen Wang, Silvia Chyou: San Francisco Crime Classification. Fall CSE 255 Assignment 2 Report, 2015.
[14] Anish Krishnan1, Aditya Sarguru, A. C. Shantha Sheela: Predicitve Analysis of Crime Data Using Deep Learning. International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018.
[15] **"A SURVEY OF CRIME PREDICTION USING MACHINE LEARNING AND RECURRENT NEURAL NETWORKS"**, International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN: 2349-5162, Vol 6, Issue 6, page no.629-634, June 2019, Available: http:www.jetir.org/papers/JETIR1906H66.pdf
Introduction to Recurrent Neural Network
https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/
Ensemble Methods in Machine Learning: What are they and Why Use Them?
https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f
Understanding LSTM Networks
http://colah.github.io/posts/2015-08-Understanding-LSTMs/
Overview of Recurrent Neural Networks And Their Applications
https://www.analyticsindiamag.com/overview-of-recurrent-neural-networks-and-their-applications/
Deep Learning | Introduction to Long Short Term Memory
https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/
Definition of crime:
https://www.en.oxforddictionaries.com/definition/crime
Definition of Machine Learning:
https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471
Denver Crime Data
https://www.kaggle.com/paultimothymooney/denver-crime-data
Seattle Police Reports
https://www.kaggle.com/sam/seattle-crime/downloads/seattle-crime.zip/3