Bangladesh Metropolitan Crime Area Prediction Using Decision Tree

Conference Paper · July 2021

DOI: 10.1109/ICCES51350.2021.9489227

2 authors, including:

Debasish Bhattacharjee Victor
University of Brahmanbaria
6 PUBLICATIONS **42** CITATIONS

SEE PROFILE

Proceedings of the 6th International Conference on Communication and Electronics Systems (ICCES-2021)
IEEE Xplore Part Number: CFP21AWO-ART; ISBN: 978-0-7381-1405-7

Bangladesh Metropolitan Crime Area Prediction Using Decision Tree

Debasish Bhattacharjee Victor, Ms. Subhenur Latif

Computer Science and Engineering Discipline, Daffodil International University, Dhaka, Bangladesh
victorbhattacharjee97@gmail.com, snlatif@daffodilvarsity.edu.bd

***Abstract*— Today's world faces many problems with crime, affecting the day to day livings and bringing general socio-economic progress to a standstill.. Different types of crimes happen daily and nightly. If it cannot be carefully noticed or managed, it would be a great disaster for any country. Therefore, this paper was aimed at predicting metropolitan Bangladesh at different crime rates in different times. In this paper, different types of machine learning techniques could be used, but Decision Tree was used based on crime quantity to forecast Bangladesh's metropolitan area and finally analyze the result depending on the algorithm 's result. In this paper we focused on Metropolitan Police Aare like DMP, CMP, KMP, RMP, BMP, SMP, Railway where D means Dhaka, C means Chittagong, K means Khulna, R means Rajshahi, B means Barisal, S means Sylhet, and MP means Metropolitan Police.**

Keywords: ***Crime, Metropolitan Area, Prediction, Bangladesh Police, Machine Learning, Decision Tree , Measuring Metrics, AUC , ROC .***

I. Introduction

[1] Everyone is different and crime will effect them in various ways. The essential thing is to remember that any changes in how you feel could be caused by the traumatic incident. One thing that might make a crime particularly difficult to deal with is to know that another person has done it intentionally. In contrast to an accident or disease, persons committing a crime plan to cause some kind of harm. The effects of crime can also persist for long, and the degree of seriousness of the crime does not depend on it. Some people deal with the most awful atrocities really well, but others may be extremely affected by a smaller incidence. Crime affects not only everybody's everyday life but also social, economic growth in the country. [2] The word crime comes from the Latin word krinos, which implies accusation. It covers activities contrary to social order that deserve social disapproval and condemnation. In general, crime involves violating the laws of law, regulating or damaging or destroying human society or resources, or creating human life problems. The word "crime" was not defined in the Penal Code of Bangladesh. However, it can be interpreted as an act of action or omission that is destructive to society as a whole in its wide definition. However, all acts that harm the community are not crimes unless they are punishable by law. Osborn states that crime is an act or default that tends to the prejudice of the community and the prohibited law on punishment inflicted on state lawsuits. Consequently, in its legal definition, crime encompasses crimes or defaults which are punishable under the Bangladesh Penal Code.. [3] On Tuesday, Bangladeshi Court confirmed the penalty for a key Islamic leaders over atrocities committed in the independence fight over four decades ago, opening the way for his execution. Ali Ahsan Mohammad Mojaheed, 67, Jamaat-e-Islami Party Secretary-General, was convicted in 2013 on counts of genocide, killing intellectuals, torture and kidnapping during the liberation . "This verdict pleases everyone in Bangladesh," Attorney General Mahbubey Alam told reporters. "No higher crime than eliminating intellectuals of the nation." Defenses counsels minister said he would seek judgment review and the Jamaat called for a 24-hour national strike on Wednesday in protest. Prime Minister Sheik Hasina ordered an inquiry into war crimes in 2010, opening the path for indictments by a war crimes tribunal criticized by Islamists as part of a politically driven campaign to destabilize Jamaat-e-leadership. Islami's Two Jamaat leaders were executed in December 2013, one in April. International human rights groups argue the tribunal's processes are below international norms. Government disputes the charge. East Pakistan broke away to become independent following India-Pakistan conflict. About three million people died. Some factions in Bangladesh, like the Jamaat, opposed Pakistan's separation, but the party denies charges that its leaders committed crimes. As an example [14] In

2020 girls compared to the same period the previous year. Another example occurred in Sylhet where [15] the alleged rapists were apprehended by the locals and handed over to the cops. In Sylhet's Beanibazar district, a fifth-grade schoolgirl was gang-raped. The two alleged rapists were apprehended by locals and later handed over to police. On Wednesday, the survivor's father filed a complaint with the Beanibazar police station. Faisal Ahmed alias Petla, 30, and Minshuku Ahmed, 26, have been arrested. The 11-year-old girl was forcibly picked up by two men and taken to a secluded location when she went out to fetch water from a tube well around 9 p.m. on Tuesday, according to local sources. Thus, foresee trends in crimes and to predicts places already in the red zone and to take further measures. This work has been used for forecasting criminal territory based on knowledge base data of different crimes that took place during 2019 since 2010 , based on [4] police website supported by the government of Bangladesh's police community . Here we applied tree based on decision and this decision tree produced ninety percent of accuracy.

II. Related Works

[12] The authors used a method of economical assumption which is normal in all companies due to a variety of factors, As a result of privatizations, globalizations, and liberalizations, the businesses is forced to compete more aggressively. Data-driven decision-making was proposed in this report. A hybrid approach is being developed, which includes neural technology with tree along with vector machines .[13] The authors of this article classify crimes by focusing on sentiment and the newspaper. The authors collected information from a variety of news sources, grouped the news by mood by sentimental polarity for per day , and then analyzed the information to identify crimes. The authors of [8] used identification and clustering anomalies to forecast crime trends, and the crime resolution mechanism was followed. Authors applied Clustering algorithms with outlier detection by mapping of the regions covered by the theme scenarios spatial ellipse, thematic-grids mappings, and kernel estimation technique . In [6] authors detect main red crime zones by mapping methodology which has been used for calculate density of crimes . Authors in [7] used a classification model to predict crime patterns. The clustering techniques was applied for the creation of clusters, speed of clusters , and at last crime area was detected . state density, and state effectiveness in crime rate control. Authors got 89% accuracy for Bayesian, 0.727 for Decision Tree, and 0.678 for Support Vector Machine. [9] The writers of this paper use sentiment analysis and newspaper data to identify crimes. The authors gathered information from a variety of news outlets, categorizing it as negative, optimistic, or normal for each day, and then analyzing it to find crimes. The authors of [10] predicted crime and used the data to infer Twitter post and find out crimes . For the study, they used a semantic analysis-based automatic crime forecasting model. [5] The authors of this novel work used information based numerical data's to forecast crimes .[11] A alliance worked on sentiment analysis in 2015 in order to classify specific data characteristics from Twitter. The sentiments were analyzed using a lexicon-based approach. They then performed lexicon-based analysis using bow and obtained their results.

III. METHODOLOGY

We used Decision Tree which is a Machine learning algorithm and got our expected result.

A. *Data Gathering*

The data was acquired from the police department of Bangladesh in this study. Crimes such as Dacoity's, Murders', Women Child Repression, Kidnappings, Police Assault, Burglaries, Thefts, Other Cases. There are nineteen crimes data on the website together with one years individually . We therefore have to collect these data each time from starting to end where starting and ending date were 2010 and 2019 by altering the year on top of the webpage URL. After we have collected these data from 2010 to 2019,

	Unit	Dacoity	Robbery	Murder	SpeedyTrial	Riot
0	DMP	47	220	245	363	3
1	CMP	16	108	94	31	7
2	KMP	3	9	29	25	0
3	RMP	4	20	21	9	15
4	BMP	8	12	19	21	0

***B.** Data Preprocessing*

There were not any missing data. The dataset contains different Units as metropolitan area of Dhaka, Chittagong, Rajshahi, Khulna, Barisal, Rangpur, Railway, Mymensingh. In Total dataset contains 181 data and for work purpose related to 6 Metropolitan Area and one Railway area then it got reduced and the final dataset contains 70 data from 2010 to 2019 year where for each Metropolitan area has 10 data from 2010 to 2019. And it is fully balanced dataset.

***C.** Model selection*

In this paper We used Decision Tree for predicting areas. Our algorithm accuracy was very good and accomplished more than 90 % accuracy for testing data. In this Paper our used this algorithm because it can handle multiclass classification and it gives a excellent result.

***D.** Data Description*

A. DMP

From this DMP we can be easily captured that the crimes rate were very extreme during early year and the most occurred crime was other cases and it was 7.2k plus and thefts were 1920 . We can see that from crime rate decreases at recent years .

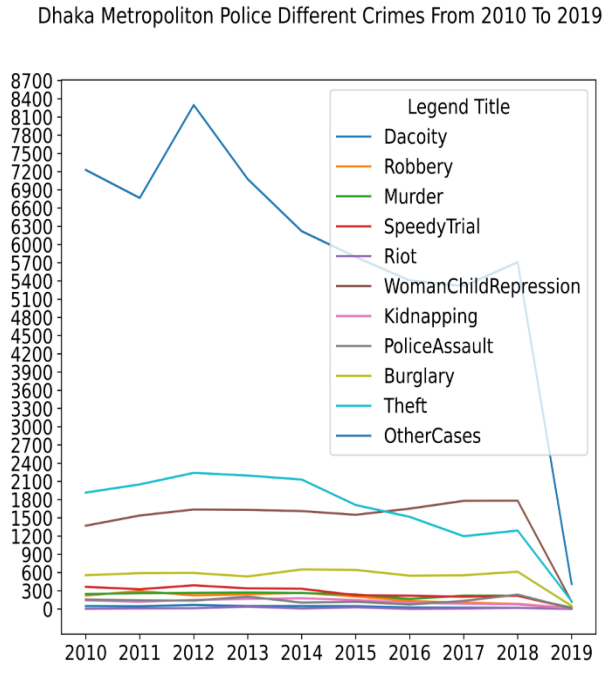

Fig 1: DMP

B. CMP

From the year of 2010 to 2019 the Dacoity crime rate was less and it was in between 0 to 16 and other cases crimes occurred maximum times at 2010 and it was 1831 and then it fluctuate different times from 2011 to 2018 and at the year of 2019 it got the value of 125. Theft rate maximum was 355 at the year of 2011. WomenChildRepression were 455 at that time also and at 2019 the most vulnerable crimes were OtherCases, WomenChildRepression and Theft.

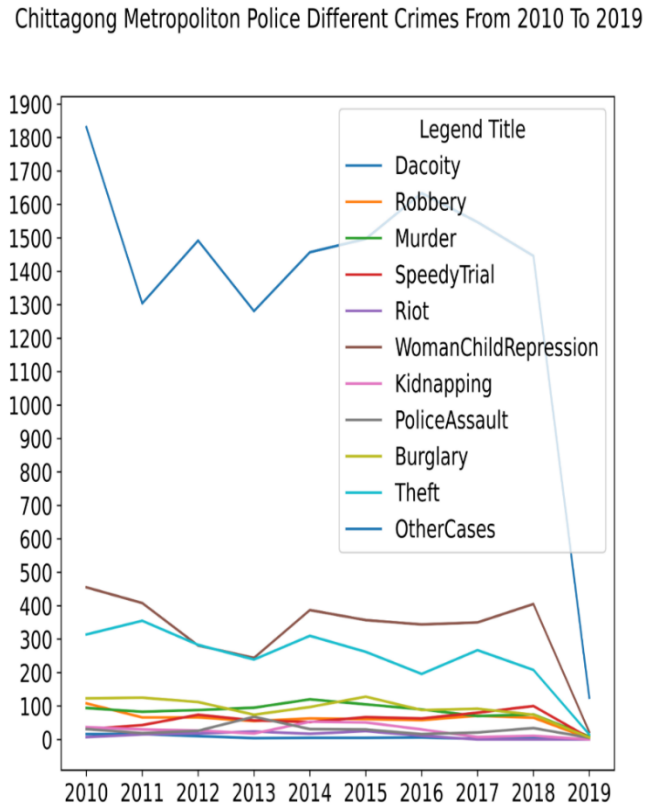

Fig 2: CMP

C. KMP

On Khulna metropolitan city WomenChildRepression crimes were in between 153 to 206 from 2010 to 2018 and no Riot crimes were occurred from 2010 to 2019 when we visualized that graph carefully .

and Dacoity were also very less and Robbery, Murder , Speedy Trail , Police Assault ,Kidnapping were less than 50 in a whole and except these crimes the other crimes rate were in between 4 to near about 600 .

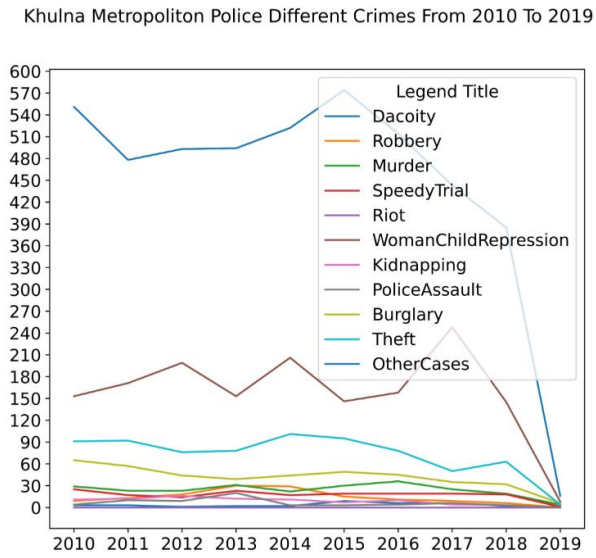

Fig 3: KMP

D. RMP:

On this Rajshahi Metropolitan Area the crime rate of other cases and theft kidnapping ,PoliceAssult were maximum . Other crimes from murder to other cases was also at alarming rate and they were at very huge rate during that period . Often At the year of 2013 and they were 791 and 166 , 23,58 and from 2010 to 2013. All those crimes gradually increased and from 2014 got decreased and got the final value at 2019 as 52,6,0,0 and remain crimes rate under 70 from 2010 to 2019 where WomenChildRepression fluctuates from 157 to 11 during that 10 periods.

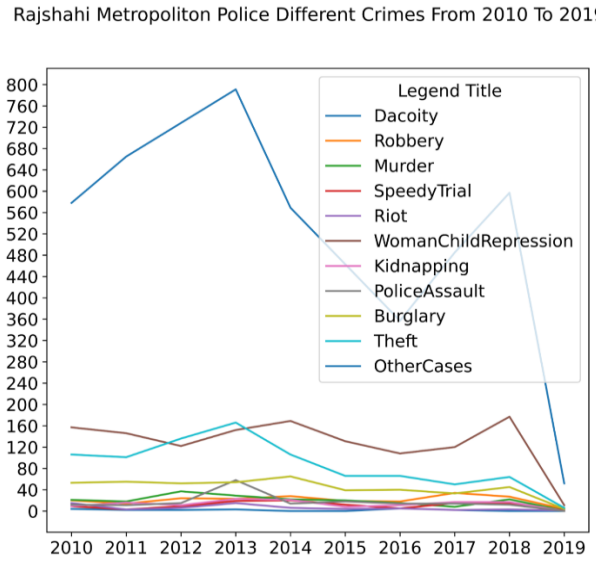

Fig 4: RMP

E. BMP

In Barisal Metropolitan area the ratio of Dacoity was in range of 18 to 0, Robbery was from 16 to 0, Murder was 20 to 0 range where the there were almost no Riot crime. The ratio of kidnapping, police assault, burglary up and down in between 88 to 0. The number of maximums occurred crime was other cases. At the year of 2019 the other cases were 49 and at 2010 it was 557 and then it got its maximum value as 600 at 2012 and the third highest number was 444 regarding the year of 2018. WomenChildRepression 147 at 2012 which is the maximum value and at 2019 it got the value of 14 which is the second most occurred crime in 2019. Then other major crimes also placed in under control .

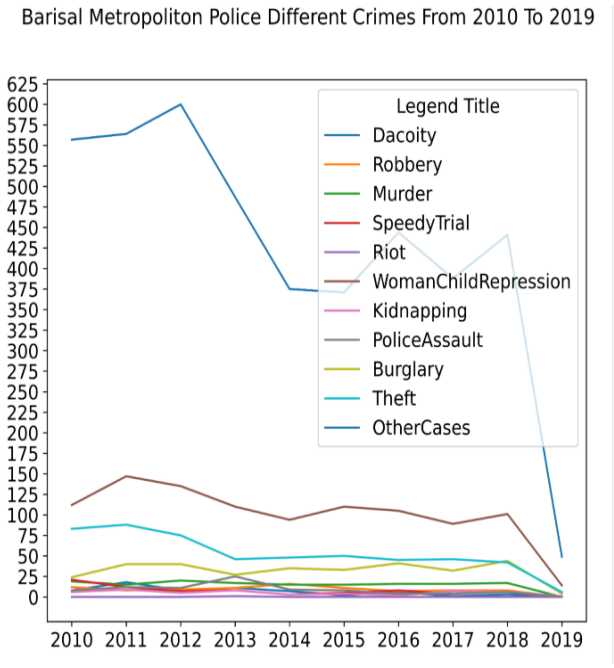

Fig 5: Barisal City

F. SMP

There are many crimes occurred in Sylhet Metropolitan Area as well as other Metropolitan cities and if we just visualize the plotting graphs of all metropolitan area along with Railway Area with Sylhet Cities then we can say that the crime radios are less with the comparison to them.

Or we can see with the data points also the same scenario for Sylhet Metropolitan Area .

Other cases, WomenChildRepression crime rate were more than 100 at 2010 where other cases were less than 40 on that particular year and from the year of 2011 to 2018 the range of Other cases were 857 to 731 and WomenChildRepression was 235 to 120. On 2019 all the crimes got decreased and got the range in between 0 to about 60 in Sylhet Metropolitan Area.

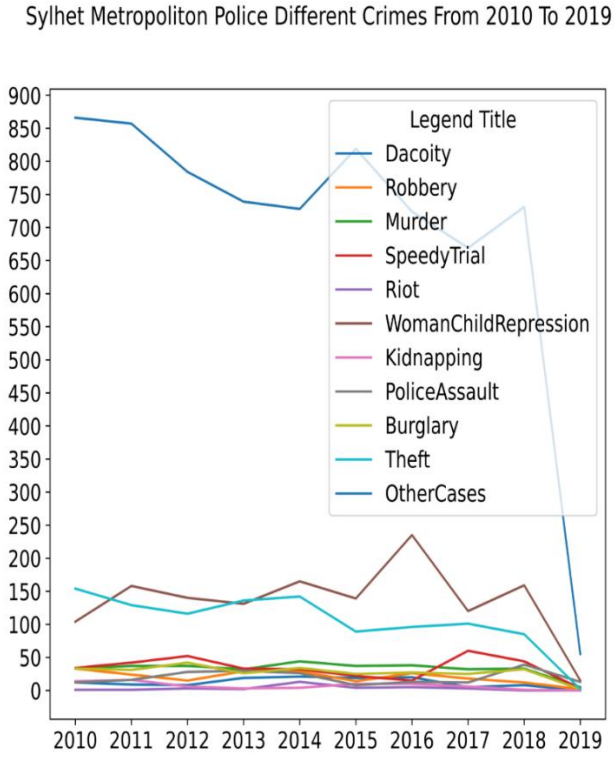

Fig 6: SMP

G. Railway

On Railway Area we can see that thee are variations of curve line in this area . Actually this scenario has not occurred in most major Metropolitan Areas . This signifies that there happened crime quantities is not determined by saying from 10 to 80 or such like that. In every tear crime for same name occurred with a various range to Railway area .And in a nutshell we could say that those hazardous lines implies different amount .

The range of dacoity ,riot ,speedy trail ,police assault, burglary ,kidnapping were less than 10, range was in between 20 for robbery , WomenChildRepression rate less than 9 from 2010 to 2015 but it got maximum value of 160 at the year of 2016 and the ratio of theft was less than in 120 range for all of this 10 years.

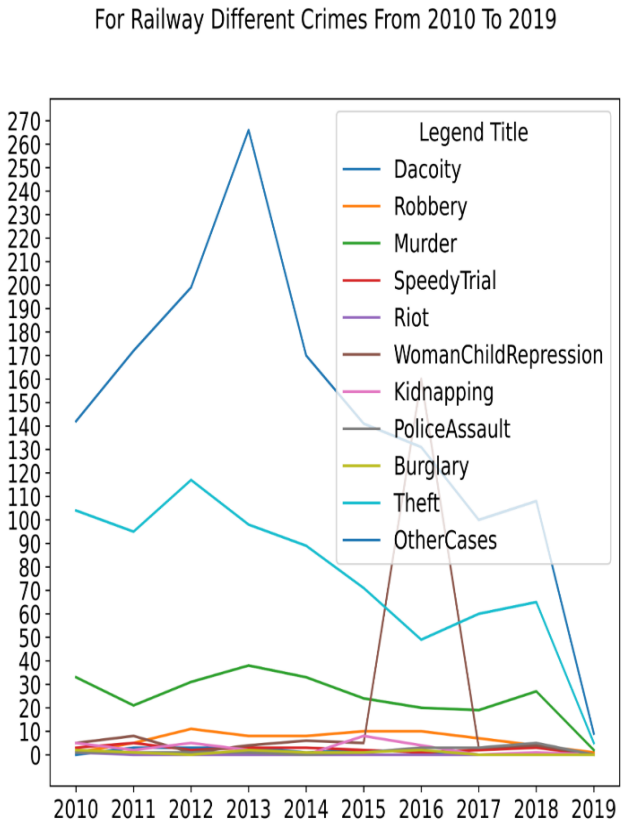

Fig 7: Railway Area

V. Result and Evaluation

In this paper we selected Decision Tree rather than others types of classifiers. In this paper we noticed that region based on others algorithm such as LR, SVM, KNN accuracy was dropped as expected and their accuracy was near about 80 to 84 percent and by analyzing that we decided to use Decision Tree to use in our research work.

Why we used Decision Tree

If there is any option in pursuing several choices, then decision trees are a great way to go about it. Decision Trees are great tools for helping to determine the multiple courses of action we need to pursue. There are enough reasons such as No feature scaling required, can handles non-linear parameters efficiently and Decision Tree can automatically handle missing values and robust for outliers and does not require many time for training period and handle less data with

not by affected by noise. Research is vital in helping us choose the right path; they offer a high-reduced decision-making framework that allows us to design experiments and examine alternative possibilities. We choose this which gave best result. Without the hyperparameter boosting it gave best result and as test we used 15% data with random state=10, max_depth=8, criterion = "gini" and get our result as expected as almost 91% for our test data. Further we measured F1 score, precision and recall for RF. Line Plots For 7 different Metropolitan Area separately from 2010 to 2019.

a) **TrueWithPositive:** It took the number of texts belonging to the data for evaluation purpose and measured correct label data with regarding to that certain target.

b) **TrueWithNegative**: It took the number of texts belonging to the data for evaluation purpose and measured correct label data with not regarding to those certain targets

c) **FalseWithPositive:** It denotes that quantity of data from the evaluation sets which is not correct label for that data, which has been encountered with the algo corresponding to that target.

d) **FalseWithNegative**: It denotes that quantity of data from the evaluation sets which is not correct label for that data encountered with the algo and not corresponding to that target.

The Precision, Recall, F1 -Score Report

ROC, AUC Curve

There are many reasons for choosing AUC , ROC such as in general or for different thresholds, the curves of different models can be directly compared. The model's skill can be summarized using the area under the curve (AUC).

Lower false positives and higher true negatives are indicated by lower values on the x-axis of the plot. Higher true positives and lower false negatives are indicated by higher values on the y-axis of the plot. The AUC function, like the roc curve() function, takes the true outcomes (0,1) from the test set as well as the predicted probabilities for the 1 class. It gives an AUC score ranging from 0.0 to 1.0 for no skill and perfect skill, respectively.

Area Under the Curve (AUC) is a measuring tool which compute the ability of a classifier and distinguish between classes which used as a summary of the ROC curve. The higher the AUC, the better the performance of the model for differentiating between the positive and negative classes.

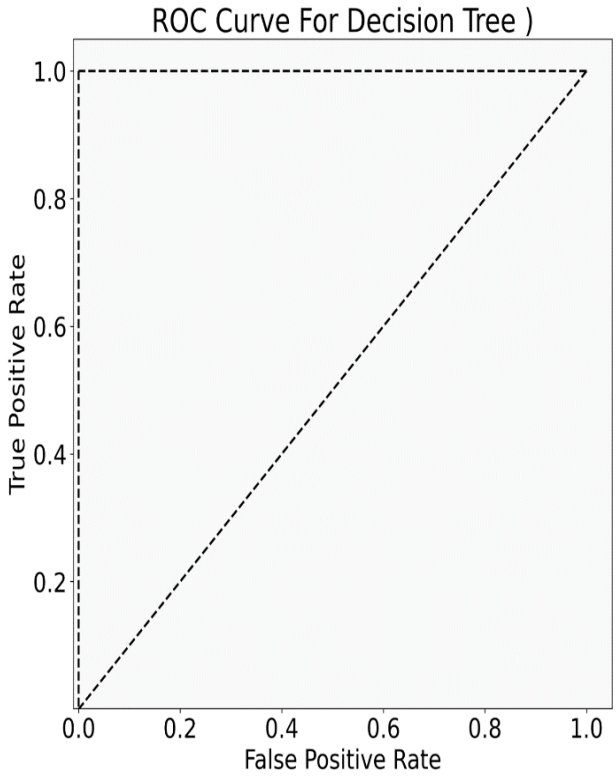

Fig : AUC, ROC Curve

V . Conclusion and Future Scope

Area	Precision	Recall	F1-Score	Support
DMP	1.00	1.00	1.00	1
CMP	0.00	0.00	0.00	0
KMP	050	1.00	0.67	1
RMP	1.00	1.00	1.00	3
BMP	0.00	0.00	0.00	1
SMP	0.00	0.00	0.00	0
Railway	1.00	1.00	1.00	5.00
Micro avg	0.91	0.91	0.92	11
Macro avg	0.50	0.57	0.52	11
Weighted avg	0.86	0.91	0.88	11

In This paper we worked on 7 Area such as DMP, KMP, RMP, CMP, Railway, BMP, SMP. We would extend our work for ATU, GMP, RPMP . And our limitation is that Decision Tree gave good precision , recall and f1 values in some metropolitan area where it fails to give accurate score for CMP, KMP, SMP . Another limitation is that our accuracy was not so high because of data size limitation. And by using that methodology the Bangladeshi Governments and others departments would be able to know that actually which areas are in hot zone and which steps needed to reduce those offence and it would be beneficial for other researchers also to find crime trends , hot zones if they have necessary data .

References

[1] "How can crime affect you? - Victim Support", Victim Support, 2021. [Online]. Available: https://www.victimsupport.org.uk/help-and-support/coping-crime/how-can-crime-affect-you/.

[2] "Some Aspect of Crime in Bangladesh - Assignment Point", Assignmentpoint.com, 2021. [Online].Available:https://www.assignmentpoint.com/arts/sociology/some-aspect-of-crime-in-bangladesh.html.

[3] "Yahoo is now a part of Verizon Media", In.news.yahoo.com, 2021. [Online]. Available: https://in.news.yahoo.com/weather/bangladesh-upholds-death-sentence-top-islamist-war-crimes-085135708.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAApraSXIpS-H20wRbWOuUBBzwYhMV5TNG7MfOYXbkrwUd_BwRJXhzH_oXksuNpnurJypc2bYNGfcQK0GEm-rzWDFdpELfapudYJ9Ay4SiN3IXolrFFjGUhHDKsCy680vvpr66mCJXdE6_uKnGZjSJvOJXDsbXgvaukx2Kiio76ZW.

[4] Police.gov.bd. 2020. Bangladesh Police. [online] Available at: [Accessed 8 October 2020].

[5] M. A. Awal, J. Rabbi, S. I. Hossain and M. M. A. Hashem, "Using linear regression to forecast future trends in crime of Bangladesh," 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, 2016, pp. 333-338, doi: 10.1109/ICIEV.2016.7760021.

[6] C. H. Yu, M. W. Ward, M. Morabito, W. Ding, "Crime Forecasting Using Data Mining Techniques", In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW' 11), pp. 779-786, 2011.

[7] A. H. Wahbeh, Q. A. Al-Radaideh, M. N. Al-Kabi, E. M. Al-Shawakfa, "A Comparison Study between Data Mining Tools over some Classification Methods", International Journal of Advanced Computer Science and Applications, The SAI Organization, Special Issue on Artificial Intelligence, pp. 18-26, 2011.

[8] J. Han, M. Kamber, J. Pei, "Data Mining: Concepts and Techniques", vol. 5. Morgan Kaufmann Publishers, USA, 2012.

[9] S. M. M. H. Chowdhury, Z. N. Tumpa, F. Khatun and S. K. F. Rabby, "Crime Monitoring from Newspaper Data based on Sentiment Analysis," 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2019, pp. 299-304, doi: 10.1109/SMART46866.2019.9117513.

[10] Wang, X., Gerber, M. S., Brown, D. E.: Automatic Crime Prediction Using Events Extracted from Twitter Posts, In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Pages 115- 125, DOI: 978-3-642-29047-3_28, USA (2014)

[11] Zhang, L., Ghosh, R., Dekhil, M., HSU, M., Liu, B.: Combining lexicon based and learning based methods for twitter sentiment analysis, In: National Conference on Advanced Technologies in Computing and network, page: 89-91, ISSN: 2277-9477, Maharashtra, India (2015) Moradabad, India, 2019, pp. 299-304, doi: 10.1109/SMART46866.2019.9117513.

[12] Kumar, T. Senthil. "Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm." Journal of Artificial Intelligence 2, no. 03 (2020): 185-193

[13] S. M. M. H. Chowdhury, Z. N. Tumpa, F. Khatun and S. K. F. Rabby, "Crime Monitoring from Newspaper Data based on Sentiment Analysis," 2019

8th International Conference System Modeling and Advancement in Research Trends (SMART),

[14] Human Rights Watch. 2021. “I Sleep in My Own Deathbed”. [online] Available at: [Accessed 21 April 2021].

[15] Dhaka Tribune. 2021. Schoolgirl gang-raped in Sylhet, two held. [online] Available at: [Accessed 21 April 2021]