This article has been accepted for publication by IEEE SMARTCOMP 2019, but has not been fully edited. Content may change in the published version of this paper. DOI will be updated once available

Developing Machine Learning based Predictive Models for Smart Policing

Lavanya Elluri*, Varun Mandalapu*, Nirmalya Roy
Department of Information Systems
University of Maryland Baltimore County
lelluri1@umbc.edu, varunm1@umbc.edu, nroy@umbc.edu

***Abstract*— Crimes are problematic where normal social issues are confronted and influence personal satisfaction, financial development, and quality-of-life of a region. There has been a surge in the crime rate over the past couple of years. To reduce the offense rate, law enforcement needs to embrace innovative preventive technological measures. Accurate crime forecasts help to decrease the crime rate. However, predicting criminal activities is difficult due to the high complexity associated with modeling numerous intricate elements. In this work, we employ statistical analysis methods and machine learning models for predicting different types of crimes in New York City, based on 2018 crime datasets. We combine weather, and its temporal attributes like cloud cover, lighting and time of day to identify relevance to crime data. We note that weather-related attributes play a negligible role in crime forecasting. We have evaluated the various performance metrics of crime prediction, with and without the consideration of weather datasets, on different types of crime committed. Our proposed methodology will enable law enforcement to make effective decisions on appropriate resource allocation, including backup officers related to crime type and location.**

Keywords- ***Crime Prediction, Weather, Temporal features, Deep Learning, Smart Policing***

I. INTRODUCTION

Crimes are fundamental social issues influencing personal satisfaction, financial development, and notoriety of a nation. Violations are one of the central points to influence essential choices in a person's life, like moving homes, avoiding dangerous neighborhoods, not going out at night. Violations influence and shape the picture of a network. Similarly, crimes influence the economy of a country by putting fiscal weight on government, requiring extra police officers, courts, and administration. Reducing crime has become challenging due to the rate at which they occur. Crime figures demonstrate a hidden, 22.6% [4] rise in the rape rate, and 273 murders are recorded till December 2018 in New York City (NYC) as shown in Table I. These figures can be lessened on the off chance we can predict the crime event and take preventive measures ahead of time. The offense rates can be altogether lessened by constant crime gauging and mass observation, which leads to a healthier lifestyle. Legitimate examination of past offense information helps in anticipating violations and boosts in diminishes the crime rate. The examination procedure incorporates an investigation of crime reports and distinguishing developing examples, arrangement, and patterns faster than a manual process would be. This investigation helps in getting ready insights, questions, and maps of interest. It likewise checks whether a crime fits in a specific known example.

*: Lavanya Elluri and Varun Mandalapu are co-first authors
Corresponding Author: nroy@umbc.edu

Crime forecasting is intrinsically troublesome. A criminal investigation has effectively affirmed violations are unequally appropriated. Besides, misconduct is a very dynamic and complex marvel driven by the general population. Researchers using various control group are still examining components for perceptive power. Knowing when and where crime is bound to happen can help to decrease urban organizers to plan more secure urban areas and police powers.

At first, criminological investigations have concentrated exclusively on socio-demographic characteristics as elements corresponding with exploitation. Researchers have seen explicit gatherings of individuals will, in general, have a higher danger of exploitation, contrasted with different gatherings – as clarified by the Lifestyle Exposure Theory. For instance, men, youthful grown-ups, and African Americans have been found to recognize the higher danger of exploitation [5].

Henceforth, in this work, we examine the capability of temporal and weather-related information for crime anticipation models. We utilize such information to show crime attractors, crime generators, and the surrounding populace in an area. We have included these components I instead of the traditional elements. We have done feature selection techniques to empirically determine if only demographic attributes play an essential role, or if the weather-related elements have any impact on the major and minor crime events happening in New York City. We have found, along with basic demographic features, weather-related attributes play less significant role in the prediction of the response variable.

In section II, we discuss work related to crime prediction. In section III we discuss the feature selection techniques and machine learning models used to find RMSE. Also, deep learning approaches will be explained in detail. The results and discussion were examined in section IV and V. In the last section, we will be concluding with the results obtained.

II. RELATED WORK

A. Urban Computing

Over the years, detecting developments and extensive scale processing foundations create an assortment of enormous information in urban spaces [7], [8], [9]: topographical information, human versatility, traffic designs,

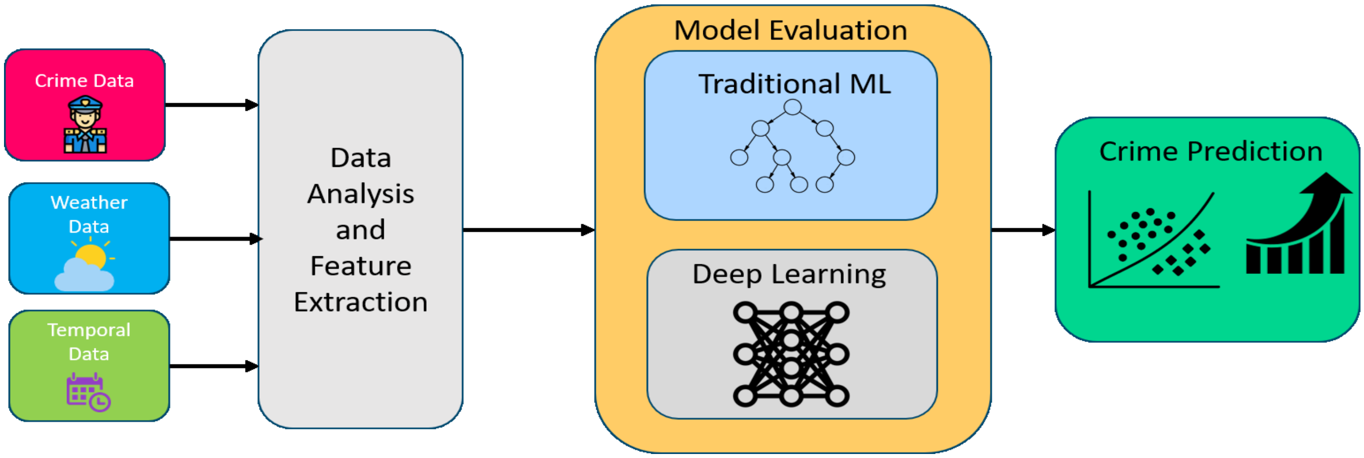

Figure 1: Architecture Flow

correspondence designs, air quality, and so forth. The vision of urban processing, a developing field authored by Zheng and partners, has opened the intensity of huge and heterogeneous information gathered in urban spaces. This is applied to measure significant issues our urban areas encounter today. They recognize seven application regions of urban processing: urban arranging, transportation frameworks, ecological issues, vitality utilization, social applications, business applications, and open comfort and security.

B. Crime Prediction

Other work which inspired us to look more into crime prediction is crime prediction based on Foursquare datasets [1]. In the paper, authors have proposed the utilization of Foursquare information for crime forecast. They employed feature selection methods to explore the intensity of various places of interest, measured by Foursquare check-ins [6], in foreseeing crime calculations in New York City over 5 years. Their examination demonstrates that the number of locations and the venue entropy is the most discriminative highlights for all incidents. This study found evidence that crime occurred is linked to the type of customers visiting that area for different activities.

Crime investigation, for example, illegal practices in the particular dimension and spatial-temporal models [11], [12], [13], [14] have been widely contemplated in recent years. Conventional crime expectation techniques incorporate grid mapping, covering ellipses, and kernel density estimation; delivering expectations dependent on the absence of uniform offense circulation. Nevertheless, these techniques typically consider time or space factors independently, furthermore, are, in this manner, extremely subtle to reality determination. This can result in expected results and prediction results that do not outperform simple linear regression. Numerous studies have used time and space factors like those mentioned in the

Foursquare analysis. A few investigations noticed the effect of environmental highlights on crime. However, few examinations have attempted to apply topographical highlights to crime anticipation. In this study, we will be closely looking at temporal features and their impact on crime category in NYC.

C. Crime Prediction with weather

Statistical crime analysis, such as whether a crime can happen, or not, based on temperature and other weather-related elements was researched by Matthew Ranson [2]. His work shows temperature has a substantial impact on crime occurrence. Based on the statistical analysis, the author states, 22000 murders can happen by 2099 in the USA. This number is too high and made us think about machine learning models which can predict whether a major or minor crime happens in a location. The author has performed extensive statistical analysis based on the time of day, he showed the disparity in occurrence, for each of the seven types of crime.

Another Ellen G. Cohn has done a similar analysis [3] on weather and crime. In this work, she focuses on how the temperature changes impact various types of crimes. The analysis shows that an increase in temperature will increase crimes like assaults, burglary, collective violence, domestic violence, and rape [11], [13]. No correlation was found between high temperatures and crimes like robbery, larceny, and motor vehicle theft. Researchers have done extensive analysis of the impact of temperature, across various crime types. However, this kind of analysis would not be much help for a police department, as it does not predict the incident occurrence. Our work mainly focuses on predicting whether a major, or minor, crime can happen based on machine learning and deep learning models. This helps the department to be alert and send the required number of officers to the location to reduce the probability of an event occurring.

III. Methodology

In this section, we describe our methodology on how to build machine learning models and cross-validate them on New York crime data. We aim to predict the crime category based on weather attributes. Figure 1 represents the architecture flow; the four phases of our methodology are

This article has been accepted for publication by IEEE SMARTCOMP 2019, but has not been fully edited. Content may change in the published version of this paper. DOI will be updated once available

Dataset extraction, data cleansing, feature selection techniques, and machine learning models.

- **Dataset Extraction:** We have downloaded the New York Police Department (NYPD) data for the current year with all the crime types whit approximately 220K incidents this year.
- **Data Cleansing:** All the empty or null values were replaced with "NA" before we performed any statistical, machine learning, or deep learning approaches.
- **Feature selection techniques:** We have used various feature selection techniques like best subset selection, forward stepwise selection, and backward stepwise selection to identify the best subset from given dataset.
- **Machine learning models:** In this step, we have performed multiple classification methods to check the confusion matrix and observe various performance metrics.

A. *Dataset Extraction and Data Cleaning stage*

In this first stage of our system, we have downloaded various NYPD datasets [5] containing all seven crime types listed under the NYPD website. However, these data sets do not have any weather-related attributes. To satisfy our need, we have collected the weather data for the whole year and used various Microsoft Excel functionalities to order and compare the compliant dates with weather dates. Then we compiled the part day, temperature, and day of the week. We also incorporated light availability based on hourly weather data from New York City and categorized the seven crime types into major or minor crimes; to classify the response variable. Categorization is done based on the crime types listed under the NYPD website. This enabled us to evaluate the predicted results from all machine learning and deep learning models performed using the RapidMiner [17]. Table I [4] shows the various crime types, and prevalence counts for the years 2017 and 2018 from NYPD CompStat website.

TABLE I: 2018 VS. 2017 CRIME INCIDENTS

Week of 12/03/18 – 12/09/18				28-day	Year to Date	
	2018	2017	%Chg	2018	2017	%Chg
Murder	13	13	0.0%	273	270	1.1%
Rape	127	126	0.8%	1,694	1,382	**22.6%**
Robbery	941	1,148	-18.0%	12,138	13,119	-7.5%
Felony Assault	1,318	1,341	-1.7%	18,953	19,077	-0.6%
Burglary	804	982	-18.1%	10,892	11,424	-4.7%
Grand Larceny	3,422	3,495	-2.1%	40,742	40,744	0.0%
Grand Larceny Auto	369	416	-11.3%	5,167	5,341	-3.3%
Total	6,994	7,521	-7.0%	89,859	91,537	-1.6%

B. *Feature Selection Techniques*

Feature selection techniques are mainly used to identify the best subsets from N number of predictors [15], [16]. There are various feature subset selection techniques, but in this paper, we have used the best subset selection, forward stepwise selection, and backward stepwise selection. Each of these subset selections and the results will be discussed in detail in the next few sections.

i. Best Subset Selection:

- ➢ Compute the null model, noted as M0. 0 for 0 parameters.
- ➢ Choose the finest model with one predictor called M1. To get M1, we fit all p models that contain precisely one predictor.
- ➢ Fit all model (p2) and choose the best model with exactly two predictors called M2.
- ➢ Fit all model (pk) and select the best model with p predictors Mp.

The best model is derived based on the smallest residual sum of squares, or consistently, the largest r-squared value for linear regression.

Figure 2 shows the results obtained from the best subset selection performed in R software for NYPD dataset. Each row in this graph denotes a model; the shaded rectangles in the columns shows the variables included in the given model.

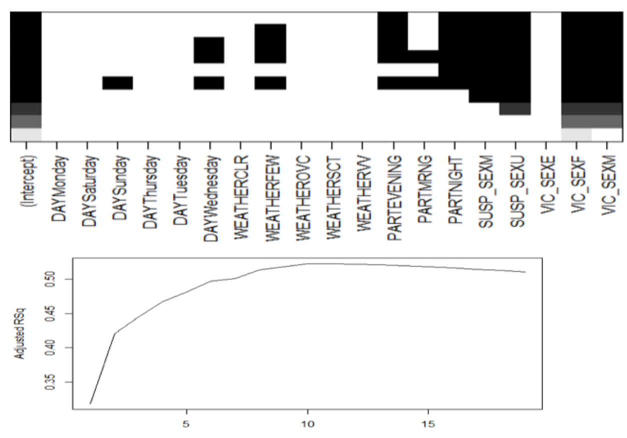

Figure 2: Best Subset selection results

ii. Forward Stepwise Selection:

- ➢ Begin with a null model. The null model has no predictors, just an intercept.
- ➢ Fit p simple linear regression models, each with one of the variables, in your variable set, and the intercept. The modeler then searches through all the single-variable models and determines the best one. The best model is defined as the one with outcomes in the lowest residual sum of squares. You pick a model and fix this one in the model.

This article has been accepted for publication by IEEE SMARTCOMP 2019, but has not been fully edited. Content may change in the published version of this paper. DOI will be updated once available

- We continue to search through the remaining, p-1, variable and find out which variable should supplement the current model, to best improve the residual sum of squares.
- This process continues until one of the ending rules is satisfied. For instance, when all outstanding variables have a p-value above a given significance threshold. Results obtained from this method are shown in Figure 3.

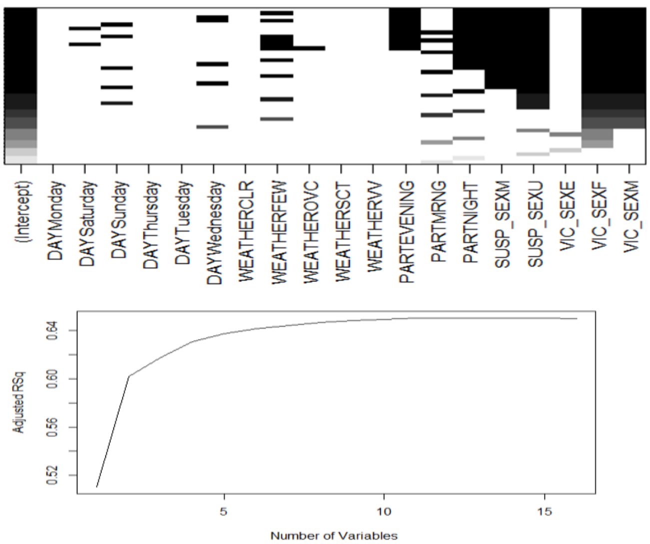

Figure 3: Forward stepwise selection results

iii. Backward Stepwise Selection:

- Begin with all variables in the model.
- Eliminate the variable with the largest p-value; the least statistically significant.
- The original (p - 1) variable model is t, and the variable with the largest p-value is removed.
- This process is continued until an ending rule is reached. For instance, we may stop when all remaining variables have an acceptable p-value, defined by a given significance threshold.

We performed this method on NYPD dataset and results are shown in Figure 4. Looking at the results of best subset selection, setting suspect and victim demographic details aside, we observed that night, evening, cloud coverage, and Sunday are significant for crimes. The stepwise forward selection shows that morning, night, Sunday, Wednesday, clear, and cloud coverage are useful in most of the models. The backward stepwise selection also shows evening, night, Sunday and cloud coverage are used in most models. Comparing all three feature selection techniques we can clearly state that nighttime, Sunday, and cloud coverage are most significant.

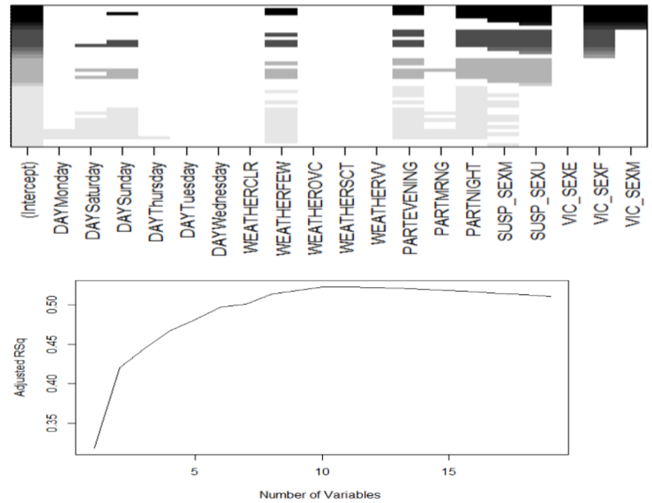

Figure 4: Backward stepwise selection results

After performing all three feature selection methods, and looking at the results, apart from basic demographic features, weather-related attributes play a significant role in predicting the response variable. Figure 5 shows the correlation matrix for all the features listed in NYPD dataset. The correlation plot mainly represents the variables positively and negatively related to each other. The yellow color in the plot represents a high positive correlation between variables, whereas the dark blue color represents a negative correlation between attributes. Also, adjusted R-squared values reach a maximum value, in all the three feature selection techniques, between five and seven variables.

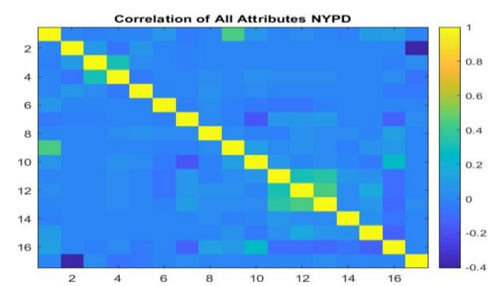

Figure 5: Correlation matrix for all the features

This study also leverages the automatic feature (AF) selection technique provided by RapidMiner data science platform [18], [19]. This technique helps us determine the strength of attributes in prediction based on four measures. They are correlation, ID-ness, stability and Missing Values. Correlation is the amount of resemblance between the attribute and the target variable. Stability is the percentage of identical values present in the attribute. ID-ness is the amount of uniqueness in the attribute values. Missing values is the number of missing values present in the attribute. The analysis of the selected attributes is listed in Table II below. We removed the Missing Values measure as there are no missing values in the selected data.

TABLE II: STRENGTH OF ATTRIBUTES ON PREDICTION

Attribute Name	Attribute Measures			Impact
	Correlation (%)	ID-ness (%)	Stability (%)	
VIC_SEX	6.05	0.01	40.14	High
VIC_RACE	0.03	0.01	28.84	High
VIC_AGE_GROUP	0.02	0.02	34.86	High
SUSP_SEX	0.04	0.01	45.14	High
SUSP_RACE	0.04	0.01	28.46	High
SUSP_AGE_GROUP	0.98	0.02	27.88	High
PATROL_BORO	0.05	0.01	21.59	High
OFNS_DESC	20.61	0.08	18.46	High
JURIS_DESC	0.08	0.03	89.06	High
DAY_PART	0.14	0.01	33.44	High
PREM_TYP_DESC	0.00	0.11	28.72	Med
LOC_OF_OCCUR_DESC	0.00	0.01	52.74	Med
HOURLY VISIBILITY	0.00	0.02	69.80	Med
WEEKEND_Vs_WEEKDAY	0.00	0.00	56.88	Med
DAY	0.00	0.02	15.22	Med
SKY_WEATHER	0.01	0.01	44.95	Med
BORO_NM	0.01	0.01	29.68	Med

Attributes are classified as the medium impact if their correlation is less than 0.01% or greater than 50% [21]. We did not deal with low impact attributes, as the attributes used in this test are selected from feature selection techniques discussed in the above sections. From the above Table II, we can observe demographic information of both victim, suspect and part of the day has a significant impact on predictions. We also see weather, visibility and day have a significant impact on predictions, as mentioned in above Table II.

C. Crime Prediction

The focus of this study is to predict the type of crime based on the weather, temporal data, and relevant crime data. For this purpose, we relabeled samples in the dataset based on their severity. For example, we labeled violations as Minor crimes and Felony & Misdemeanor as Major crimes in the dataset. We obtained 17 features of crime, including weather and temporal for algorithm training and testing. We use RapidMiner data science platform to train and test traditional algorithms. For deep learning, we utilized Keras and TensorFlow in python. We also developed similar deep learning architecture for all four. The structure is a 4-layer architecture with one input layer containing: 17 input neurons based on 17 best features, one CNN/RNN layer based on the type of algorithm with 32 neurons, one fully connected layer with 256 neurons, and finally an output layer with two neurons for two classes

IV. RESULTS

In this study, to avoid overfitting, we adopt dropout in our deep learning algorithms. We apply dropout after each connected layer. For the CNN/RNN layer, we apply 20 percent dropout, and for the fully connected layer, we applied a dropout of 50 percent. For a deep neural network, models are trained in mini batches with a batch size of 32. All models are trained on a multitude of epochs, as too few or too high will have an impact on model training, like underfitting and overfitting. We used binary cross entropy for binary classifications and Adam optimizer for optimization of the network.

In this study to evaluate machine learning models we primarily compare four performance metrics. These are the Area Under Curve (AUC), Accuracy & Cohen's Kappa. For stable and accurate results, we trained and tested all ten models using 5-fold cross-validation. In this way, hyperparameter settings are consistent across all folds of cross-validation. We also divided the algorithms into traditional and deep learning based on the type and reported results accordingly.

A. Traditional Algorithms

For this study, we trained and tested AutoMLP (Multilayer Perceptron), Decision Tree, Logistic Regression, Random Forest [10], Neural net & SVM. From the results mentioned in below Table III, we observe Decision tree,

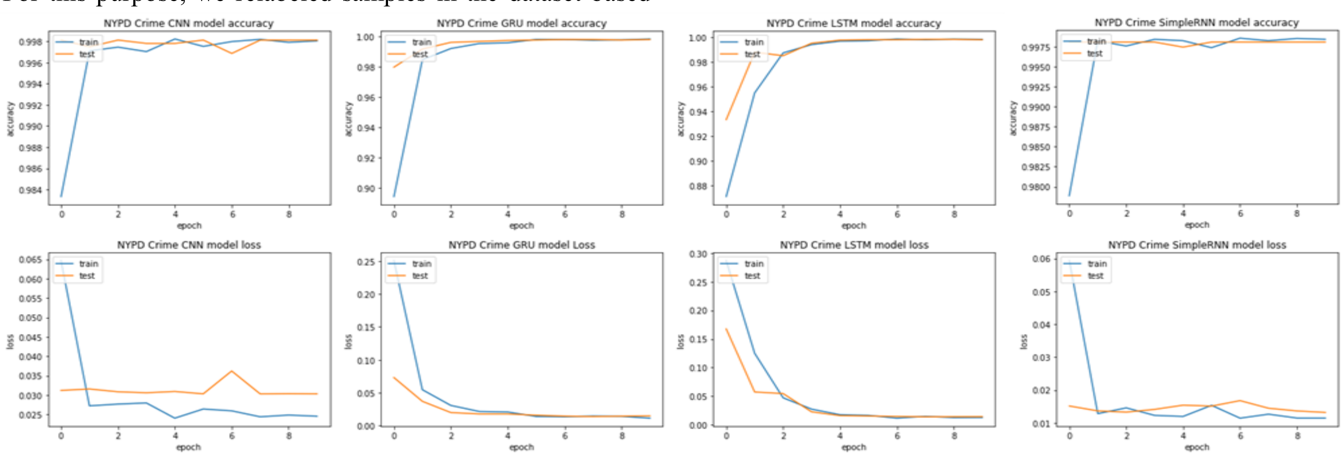

Figure 6: Plots shows the accuracy and loss of training and validation of all 4 Deep learning algorithms

This article has been accepted for publication by IEEE SMARTCOMP 2019, but has not been fully edited. Content may change in the published version of this paper. DOI will be updated once available

AutoMLP, Neural net, & Logistic regression performed well, with the highest AUC of 1, kappa 0.99 and lowest RMSE of 0.035. This is consistent with our expectation, because the features in the dataset do not have an explicit relationship with each other, the neural nets are trained to calculate the weights among features and eventually identify remarkable features for classification.

B. *Deep Learning*

In this study, we mainly focus on four deep learning algorithms [20]. From the performance metrics, we observe, deep learning performs very well when predicting the type of crime. We observe, LSTM model predicts with a higher AUC of 0.997 and a kappa value of 0.995. By comparing Tables III and IV, we can observe, neural networks perform better compared to traditional machine learning algorithms. The RNN variants LSTM, GRU and SimpleRNN vary based on their total number of parameters and flexibility, which impact their computational complexity.

TABLE III: TRADITIONAL MACHINE LEARNING MODELS PERFORMANCE FOR CRIME PREDICTION

Algorithm	**With Weather Attributes**			**Without Weather Attributes**		
	AUC	**Accuracy (%)**	**Kappa**	**AUC**	**Accuracy (%)**	**Kappa**
AutoMLP	0.995	99.87	0.996	0.994	99.87	0.996
SVM	0.635	78.53	0.001	0.605	77.36	0.029
Decision Tree	1	99.85	0.995	0.997	99.87	0.996
Neural Net	0.996	99.86	0.996	0.996	99.87	0.996
Logistic Regression	0.996	99.65	0.990	0.996	99.63	0.989
Random Forest	0.998	92.26	0.688	0.999	99.87	0.996

TABLE IV: DEEP MACHINE LEARNING MODELS PERFORMANCE FOR CRIME PREDICTION

Algorithm	**With Weather Attributes**			**Without Weather Attributes**		
	AUC	**Accuracy**	**Kappa**	**AUC**	**Accuracy**	**Kappa**
CNN	0.996	99.55	0.995	0.996	99.85	0.995
Simple RNN	0.990	98.91	0.971	0.996	99.86	0.996
LSTM	0.997	99.85	0.995	0.993	99.86	0.989
GRU	0.996	99.86	0.995	0.995	99.86	0.993

V. DISCUSSION

To analyze the performances of deep neural networks, we set aside 10 percent of the data for validation during training. We then plotted the accuracy and loss of both training and validation sets to observe if there are any discrepancies. From the below plots, in Figure 6, we observe, the algorithm performances on both training and testing sets are highly accurate. We also observed, there is no problem of either overfitting or undertraining as the training stopped when performance loss and accuracy of the validation set are stable; without changes.

To analyze the performance of algorithms and understand the highly accurate performance of deep learning and neural networks we visualized the dataset using TSNE, visualization technique [21]. The visualization below, in Figure 7, shows the distribution of Major and Minor crime samples in a 3-dimensional plot. The samples in green are related to major crime, and the samples in blue are related to minor crime. We can observe that the data related to both classes are distributed into groups of samples. This will enable the algorithms to calculate local minima which improve algorithms train and test performance [22]. The ability of neural networks to tune to different parameters and weights give them high flexibility to fit huge datasets. Main issues with deep learning are its complex models which are hard to interpret and computationally expensive. Decision tree algorithm performed better, and their usage is highly recommended as this is easy to interpret and also has less computational complexity [23].

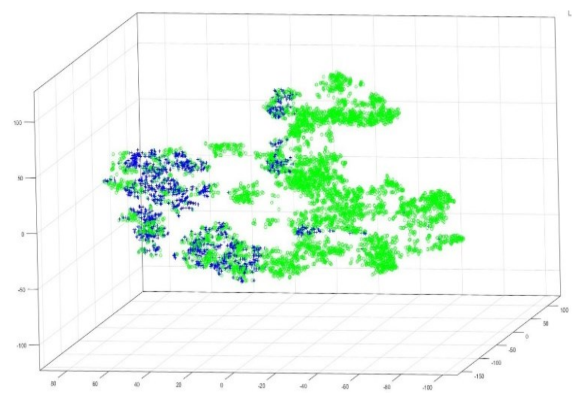

Figure 7: Visualization of crime dataset with the weather and temporal features

VI. CONCLUSION

In this work, we have used 2018 NYPD crime and New York city weather data set to check if the weather-related attributes play a significant role, or not, by performing various feature selection techniques. We then finalized the most essential features in the dataset, that influence predictions, and applied various machine learning and deep learning algorithms to compare their performances in predicting crime data. These models can be used to send high or low crime alerts to police officers. One interesting observation is the negligible impact of weather-related attributes on algorithm predictions even they seemed relevant based on feature selection techniques and in contrast with earlier studies. In the next stage, we want to combine population density based on the location with the current features and observe if this factor plays a significant role in predicting crimes. We want to perform similar prediction methods on all the major cities, with significant crime impacts in the USA.

ACKNOWLEDGMENT

This article has been accepted for publication by IEEE SMARTCOMP 2019, but has not been fully edited. Content may change in the published version of this paper. DOI will be updated once available

This research is partially supported by the NSF CAREER Award # 1750936 and ONR under grant N00014-15-1-2229.

REFERENCES

[1] Kadar, C., Iria, J., & Cvijikj, I. P. (2016). Exploring Foursquare-derived features for crime prediction in New York City. In The 5th International Workshop on Urban Computing (UrbComp 2016). ACM.

[2] Ranson, M. (2014). Crime, weather, and climate change. Journal of environmental economics and management, 67(3), 274-302.

[3] Cohn, E. G. (1990). Weather and crime. The British Journal of Criminology, 30(1), 51-64.

[4] NYPD CompStat 2.0. (n.d.). Retrieved December 18, 2018, from https://compstat.nypdonline.org/2e5c3f4b-85c1-4635-83c6-22b27fe7c75c/view/89

[5] Calgary, O. (n.d.). NYPD Complaint Data Current (Year To Date). Retrieved from https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243/data

[6] https://foursquare.com/developers/explore#req=venues/49eeaf08f964a52078681fe3?

[7] Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 5(3), 38.

[8] Salim, F., & Haque, U. (2015). Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and Internet of Things. International Journal of Human-Computer Studies, 81, 31-48.

[9] Barbosa, J. L. V. (2015, December). Ubiquitous computing: Applications and research opportunities. In Computational Intelligence and Computing Research (ICCIC), 2015 IEEE International Conference on (pp. 1-8). IEEE.

[10] Del Río, S., López, V., Benítez, J. M., & Herrera, F. (2014). On the use of MapReduce for imbalanced big data using Random Forest. Information Sciences, 285, 112-137.

[11] Wang, X., & Brown, D. E. (2012). The spatio-temporal modeling for criminal incidents. Security Informatics, 1(1), 2.

[12] Lee, I., Jung, S., Lee, J., & Macdonald, E. (2017). Street crime prediction model based on the physical characteristics of a streetscape: Analysis of streets in low-rise housing areas in South Korea. Environment and Planning B: Urban Analytics and City Science, 2399808317735105.

[13] ToppiReddy, H. K. R., Saini, B., & Mahajan, G. (2018). Crime Prediction & Monitoring Framework Based on Spatial Analysis. Procedia Computer Science, 132, 696-705.

[14] Perry, W. L. (2013). Predictive policing: The role of crime forecasting in law enforcement operations. Rand Corporation.

[15] Hocking, R. R., & Leslie, R. N. (1967). Selection of the best subset in regression analysis. Technometrics, 9(4), 531-540.

[16] Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. British Journal of Mathematical and Statistical Psychology, 45(2), 265-282.

[17] Mierswa, I., & Klinkenberg, R. (2019). RapidMiner Studio (9.1) [Data science, machine learning, predictive analytics]. Retrieved from https://rapidminer.com/

[18] Mierswa, I. (2009). Non-convex and multi-objective optimization in data mining (Doctoral dissertation).

[19] Mierswa, I., & Wurst, M. (2006, July). Information preserving multi-objective feature selection for unsupervised learning. In Proceedings of the 8th annual conference on Genetic and evolutionary computation (pp. 1545-1552). ACM.

[20] Goodfellow, Ian, et al. Deep learning. Vol. 1. Cambridge: MIT press, 2016.

[21] Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(Nov), 2579-2605.

[22] Kawaguchi, K. (2016). Deep learning without poor local minima. In Advances in Neural Information Processing Systems (pp. 586-594).

[23] Bhargava, N., Sharma, G., Bhargava, R., & Mathuria, M. (2013). Decision tree analysis on j48 algorithm for data mining. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3(6)