Using Machine Learning Algorithms to Analyze Crime Data

Article *in* Machine Learning and Applications An International Journal · March 2015

DOI: 10.5121/mlaij.2015.2101

2 authors, including:

Natarajan Meghanathan
Jackson State University
242 PUBLICATIONS **1,741** CITATIONS

SEE PROFILE

USING MACHINE LEARNING ALGORITHMS TO ANALYZE CRIME DATA

Lawrence McClendon and Natarajan Meghanathan*

Jackson State University, 1400 Lynch St, Jackson, MS, USA

ABSTRACT

Data mining and machine learning have become a vital part of crime detection and prevention. In this research, we use WEKA, an open source data mining software, to conduct a comparative study between the violent crime patterns from the Communities and Crime Unnormalized Dataset provided by the University of California-Irvine repository and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com. We implemented the Linear Regression, Additive Regression, and Decision Stump algorithms using the same finite set of features, on the Communities and Crime Dataset. Overall, the linear regression algorithm performed the best among the three selected algorithms. The scope of this project is to prove how effective and accurate the machine learning algorithms used in data mining analysis can be at predicting violent crime patterns.

KEYWORDS

Machine Learning, Crime Pattern, Linear Regression, Additive Regression, Decision Stump

1. INTRODUCTION

The Federal Bureau of Investigation (FBI) defines a violent crime as an offense which involves force or threat [1]. The FBI's Uniform Crime Reporting (UCR) program categorizes these offenses into four categories: murder, forcible rape, robbery, and aggravated assault. The FBI UCR program defines each of the offenses as follows: (i) Murder - The willful (non-negligent) killing of one human being by another. The UCR does not include deaths caused by accident, suicide, negligence, justifiable homicides and attempts to murder or assaults to murder (which are scored as aggravated assaults), in this offense classification [2]. (ii) Forcible Rape - Rape is a sexual attack on a female against her will. Though attempts or assaults to commit rape by threat or force are considered crime under this category, statutory rape (without force) and other sex offenses are excluded [3]. (iii) Robbery - The taking or attempting to take anything of value from the care, custody, or control of a person or persons by force or threat of force or violence and/or by putting the victim in fear [4]. (iv) Aggravated Assault - It is the unlawful attack conducted by one person upon another to inflict severe or aggravated bodily injury. The UCR program specifies that an aggravated assault usually involves the use of a weapon or other means to produce death or great bodily harm. Attempted aggravated assaults that involves the use of guns, knifes and other weapons are considered to belong to this category because if the assault were completed, it would have lead to serious personal injury. An offense that involves both aggravated assault and larceny-theft occurring together, the offense is considered to belong to the category of robbery [5]. Unfortunately, these type of crimes seem to have become common place in the society. Law

DOI : 10.5121/mlaij.2015.2101

enforcement officials have turned to data mining and machine learning to aid in the fight of crime prevention and law enforcement.

In this research, we implemented the Linear Regression, Additive Regression, and Decision Stump algorithms using the same finite set of features, on the communities and crime un normalized dataset to conduct a comparative study between the violent crime patterns from this particular dataset and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com [6]. The crime statistics used from this site is data that has been provided by the FBI and had been collected for the year 2013 [6]. Some of the statistical data that was provided by neighborhoodscout.com such as the population of Mississippi, population distribution by age, number of violent crimes committed, and the rate of those crimes per 100K people in the population are also features that have been incorporated into the test data to conduct analysis.

The rest of the paper is organized as follows: Section 2 gives an overview of data mining and machine learning. It also provides more information about the Linear Regression, Additive Regression, and Decision Stump machine learning algorithms that were selected for this project as well as the five metrics of output evaluation. Section 3 provides information about the Communities and Crime Un normalized Dataset that had been used for the test projection, a descriptive list of all the features that were selected to conduct the comparative study, and the statistical data that was provided from neighborhoodscout.com. Section 4 presents the results from each of the algorithms and Section 5 concludes with the findings and discussion of the project results.

2. DATA MINING AND MACHINE LEARNING ALGORITHMS

2.1. Data Mining

Data mining is part of the interdisciplinary field of knowledge discovery in databases [7]. Data mining consist of collecting raw data and, (through the processes of inference and analysis); creating information that can be used to make accurate predictions and applied to real world situations such as the stock market or tracking spending habits at the local Wal-Mart. It is the application of techniques that are used to conduct productive analytics. Data mining software packages such as the Waikato Environment for Knowledge Analysis (WEKA), the data mining software package used in this project, are used to conduct analysis of data sets by utilizing machine learning algorithms [8]. The five tasks that these types of software packages are designed for are as follows: (i) Association - Identifying correlations among data and establishing relationships between data that exist together in a given record [7, 9]. (ii) Classification - Discovering and sorting data into groups based on similarities of data [7]. Classification is one of the most common applications of data mining. The goal is to build a model to predict future outcomes through classification of database records into a number of predefined classes based on a certain criteria. Some common tools used for classification analysis include neural networks, decisions trees, and if-then-else rules [9]. (iii) Clustering - Finding and visually presenting groups of facts previously unknown or left unnoticed [7]. Heterogeneous data is segmented into a number of homogenous clusters. Common tools used for clustering include neural networks and survival analysis [9]. (iv) Forecasting - Discovering patterns and data that may lead to reasonable predictions [7]. It estimates the future value based on a record's pattern. It deals with continuously valued outcome. Forecasting relates to modeling and the logical relationships of the model at some time in the future [9]. (v) Visualization - Enabling researchers to rapidly and efficiently locate vital information that is of interest [7]. It also refers to presenting data in such a way that users can view complex patterns within the data. It is used in conjunction with other data mining

models to provide a clear understanding of the patterns or relationships discovered. Some of the tools used in visualization include 3D graphs, "Hygraphs" and "SeeNet" [9].

2.2. Machine Learning

Arthur Samuel, a pioneer in machine learning and artificial intelligence defined machine learning as a field of study that gives computers the ability to learn without being explicitly programmed [10]. In essence, machine learning is a computer system's method of learning by way of examples. There are many machine learning algorithms available to users that can be implemented on datasets. However, there are two major types of learning algorithms: supervised learning and unsupervised learning algorithms. Supervised learning algorithms work by inferring information or "the right answer" from labeled training data. The algorithms are given a particular attribute or set of attributes to predict. Unsupervised learning algorithms, however, aim to find hidden structures in unlabeled class data. In essence, the algorithms learn more about the dataset as it is given more examples to be implemented on. There are five types of machine learning algorithms that are used to conduct analysis in the field of data mining: (i) Classification Analysis Algorithms - These algorithms use the attributes in the dataset to predict values for one or more variables that take discrete values. (ii) Regression Analysis Algorithms - These algorithms use the attributes of a dataset to predict values for one or more variables that take continuous values (e.g., profit/loss). It a statistical tool used in the process of investigating the relationships between variables [11]. (iii) Segmentation Analysis Algorithms - Divide data into groups or clusters of items that have similar properties. (iv) Association Analysis Algorithms - Find correlations between different attributes in a dataset. Typical application of such type of algorithms involves creation of association rules, which can be used in market basket analysis. (v) Sequence Analysis Algorithms - Summarize frequent sequences or episodes in data, such as Web path flow. Sequence analysis works by discovering the identification of associations or patterns over time [9].

2.3. Algorithms Selected for Analysis

WEKA provides many machine learning algorithms from eight different categories for users to implement and conduct analysis on datasets: Bayes, Functions, Lazy, Meta, Multi-Instance (MI), Miscellaneous, Rules, and Trees. The following algorithms were selected to conduct analysis of the Communities and Crime Un normalized Data set over the course of this research project.

- ***Linear Regression*** - The algorithm uses linear regression for prediction and uses the Akaike criterion to select models; the algorithm could work with weighted instances. This method of regression is simple and provides an adequate and interpretable description of how the input affects the output. It models a variable Y (a response value) as a linear function of another variable X (called a predictor variable); Given n samples or data points of the form (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), where $x_i \in X$ and $y_i \in Y$, predictive regression can be expressed as $Y = \alpha + \beta X$, where α and β are regression coefficients. Assuming that the variance of Y is a constant, the coefficients can be solved using the least squares method. This minimizes the error between the actual data point and the regression line.

$$\beta = \frac{\left[\sum (x_i - mean_x)(y_i - mean_y)\right]}{\left[\sum (x_i - mean_x)^2\right]} \quad \text{and } \alpha = mean_y - \beta * mean_x$$

where $mean_x$ and $mean_y$ are the mean values for random variables X and Y given in a data training set. The X variable is the input value (independent) and Y is the response output value (dependent) that depends on X.

- ***Additive Regression*** - This is a meta classifier algorithm that could enhance the performance of a regression base classifier. Each iteration of the algorithm fits a model for the residuals from the previous iteration of the classification process. Prediction is accomplished by adding the predictions of each classifier. Reducing the shrinkage (learning rate) parameter helps to prevent over-fitting and has a smoothing effect but increases the learning time. Each input feature makes a separate contribution to the output, and they are just added together. It is denoted by the following equation.

$$\mathbf{E} = \left[Y \mid \overrightarrow{X} \models \vec{x} \right] = \alpha + \sum_{j=1}^{p} f_j(x_j)$$

- ***Decision Stump*** - This algorithm is a class for building and uses a decision stump along with a boosting algorithm. The algorithm does regression (based on mean-squared error) or classification (based on entropy). The missing values are treated as separate values. Decision trees have a robust nature that allows them to work well with large datasets and helps algorithms to make better decisions about the variables. Decision trees typically have multiple layers consisting of three types of nodes as shown in Figures 1-2 [12] and explained below:

  - ✓ Root node - has incoming edges and zero or more outgoing edges
  - ✓ Internal node - each of which has one incoming edge and two or more outgoing edges
  - ✓ Leaf node - commonly referred to as an end node, each of which has exactly one incoming edge and no outgoing edges [12].

The decision stump is basically a decision tree, however, with a single layer as shown in Figure 2. A stump stops after the first split. They are typically used in population segmentation for large data and in smaller datasets to aid in making decisions in simple yes/no models .

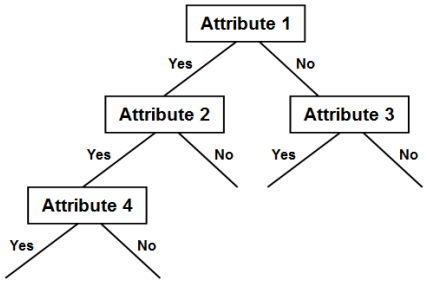

Figure 1: A Sample Decision Tree Model

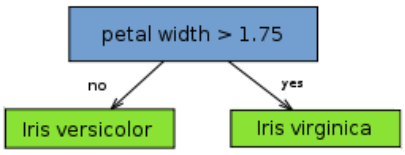

Figure 2: Decision Stump Model

2.4. Evaluation Metrics

After implementation of the algorithms, WEKA outputs five metrics that evaluate the effectiveness and efficiency of the algorithms: Correlation coefficient, Mean absolute error, Root mean squared error, Relative absolute error, and the Root relative squared error. The results for

these five metrics will be used in the comparative evaluation of the crime statistics. The objective of this research is to present how effective the algorithms can be in determining patterns of criminal activities.

- ***Correlation Coefficient*:** The correlation coefficient measures the strength of association between two variables. The value of a correlation coefficient will range between -1 and 1. The larger the absolute value of the correlation coefficient, the more stronger is the relationship between the variables. The strongest relationships are indicated by coefficient values of -1 or 1. The weaker relationships are indicated by a value of 0. If the correlation is positive, it means that as one variable becomes larger, the other variable tends to become larger. For example, as an individual's level of education increases (x-axis), the wage per hour that they will make per hour increases (y-axis). A negative correlation means that if one of the variables grow larger, the other usually gets smaller. Again for example, the more failing grades a student receives, the lower the probability of that student passing becomes. Strong correlations on the scatter plots are indicated by the data points plotted just as a straight line whether positive or negative. The more random the data points, the weaker the correlations between the variables.

The correlation coefficient is evaluated by the equation: $C_i = \frac{Cov(T,P)}{\sigma_t * \sigma_p}$, where $Cov(T,P)$ is the covariance of the target and model outputs; and σ_t and σ_p are the standard deviations calculated as follows:

$$Cov(T,P) = \frac{1}{n}\sum_{j=1}^{n}\left(T_j - \overline{T}\right)\left(P_j - \overline{P}\right) \qquad \sigma_t = \sqrt{\frac{\sum_{j=1}^{n}\left(T_j - \overline{T}\right)^2}{n}} \qquad \sigma_p = \sqrt{\frac{\sum_{j=1}^{n}\left(P_j - \overline{P}\right)^2}{n}}$$

where P_j is the value predicted by a machine learning algorithm for sample case j (out of n sample cases); T_j is the target value for sample case j; $\overline{T}$ and $\overline{P}$ are respectively the means of the target values and predicted values for the test samples.

- ***Mean Absolute Error*:** The mean absolute error (*MAE*) measures the absolute difference between the predicted values and observed values of the target feature. *MAE* is evaluated by the following equation:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | x_{pred} - x_{obs} |$$

The absolute value of the difference from the predicted values and known values are taken and then divided by the number of observations in the dataset. In essence, it measures the average magnitude of the error. The values of the mean absolute error can range anywhere from 0 to infinite. It also presents negatively-oriented values, which means the lower the value the more accurate will be the algorithm.

- ***Root Mean Squared Error*:** The root mean squared error (*RMSE*) is the square root of the average of the square of total error. The root mean squared error is commonly used to measure the accuracy of errors for numerical predictions. It is defined by the equation below, where x_{pred} and x_{obs} are the predicted and observed values.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_{pred} - x_{obs}\right)^2}$$

- ***Relative Absolute Error*:** The relative absolute error (*RAE*) is a total of the absolute error. It works by taking the total absolute error and normalizing it by diving by the total absolute error of the prediction. It is denoted as follows, where P_j and T_j are respectively the predicted and targeted values for test sample j and $\overline{T}$ is the mean of the targeted values for the test samples.

$$RAE = \frac{\sum_{j=1}^{N} | P_j - T_j |}{\sum_{j=1}^{N} | T_j - \overline{T} |}$$

- ***Root Relative Squared Error*:** The root relative squared error (*RRSE*) is denoted by the equation below, where P_j and T_j are respectively the predicted and targeted values for test sample j and $\overline{T}$ is the mean of the targeted values for the test samples.

$$RRSE = \sqrt{\frac{\sum_{j=1}^{N}\left(P_j - T_j\right)^2}{\sum_{j=1}^{N}\left(T_j - \overline{T}\right)^2}}$$

3. DATASET ANALYSIS

3.1. Communities and Crime Un normalized Dataset

The dataset selected to conduct this research is a Communities and Crime Un normalized dataset. It consists of socio-economic data from the 1990 Census, law enforcement data from 1990 Law Enforcement Management and Admin Stats survey, and crime data from the 1995 FBI UCR. It also consists of 2215 instances or crimes that had been reported from across the country and 147 total attributes, commonly referred to as features. The dataset contains 4 non-predictive features, 125 predictive features, and 18 potential goal features which are listed below. The associated task for this particular dataset is regression; so the regression algorithms mentioned in the previous section were used for the predictions and comparison [13]. The non-predictive (identifying variables) features could possibly hinder implementation or even prevent certain algorithms from being used: (i) community name: Community name - not predictive - for information only (string); (ii) state: US state (by 2 letter postal abbreviation) (nominal); (iii) county code: numeric code for county - not predictive, and many missing values (numeric); (iv) community code: numeric code for community - not predictive and many missing values (numeric). The predictive features include those that involve the community and law enforcement. These particular set of features in the dataset were selected because they are believed to have plausible connections to one of the 18 potential crime goals.

Since the focus of this project is towards analyzing the crime patterns of the four violent crime categories, the features that will be analyzed are the murders, murdPerPop, rapes, rapesPerPop, robberies, robbbPerPop, assaults, assaultPerPop, and ViolentCrimesPerPop. WEKA has a built in

visualization tool that presents a plot matrix. The plot matrix displays scatter plots that shows the correlations between two features as shown in Figures 3 and 4. This feature along with the results from the output evaluation were also used in analysis process.

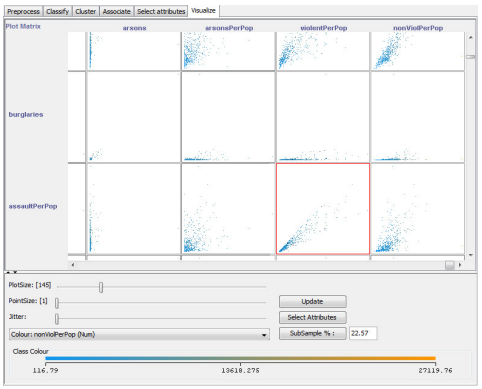

Figure 3: Plot Matrix

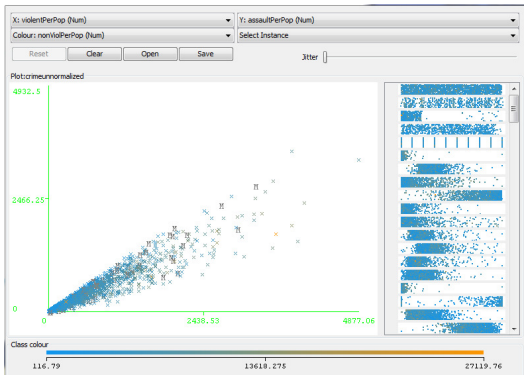

Figure 4: Scatter Plot of Violent Crimes per 100K and Assaults per 100K

The "NomialToBinary -R first-last" filter was applied to the dataset. This, instead of listing "state" as a feature, presented the two letter abbreviation of all the states that make up the United States homeland. All the other states were removed as features. The community name feature is of the String data type and would not allow the regression algorithms that were chosen for this project to be used so it was also removed from the list of features.

3.2. Mississippi 2013 Crime Dataset

The following data are reported from neighborhoodscout.com and are used in comparative studies. Table 1 displays the total of all crimes (classified broadly as violent and property crimes) and Table 2 displays a breakdown of the violent crimes. Figure 5 displays the number of crimes reported in different cities of the state, with the intensity of the blue color used as a measure to indicate the magnitude of the number of crimes. The total population of the state is 2,991,207.

Table 1: Mississippi Crime Totals

Annual Crimes		
Violent	Property	Total
8,214	81,500	89,714
Annual Crimes per 1,000 residents		
2.75	27.25	29.99

Table 2: A Breakdown of the Violent Crimes

	Murder	Rape	Robbery	Assault
Report Total	195	930	2,409	4,680
Rate per 1,000	0.07	0.31	0.81	1.56

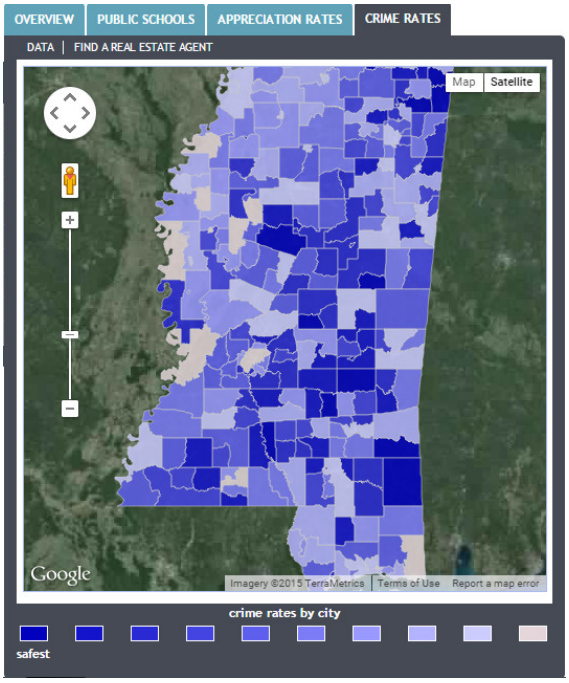

Figure 5: Crime Rates in Different Cities of Mississippi

4. RESULTS

This section presents all of the results from the implementations of the Linear Regression, Additive Regression, and Decision Stump algorithms. The algorithms were run to predict each of the following features in the datasets: murders, murdPerPop, rapes, rapesPerPop, robberies, robbbPerPop, assaults, assaultPerPop, and ViolentCrimesPerPop. Note that perPop refers to for every 100K of people. The algorithm that gives the lowest error values for each feature and the highest correlation coefficient is highlighted in the results presented in Tables 3 through 11.

Table 3: Results for Murder [Total Number of Instances - 2215]

Algorithm	*Correlation Coefficient*	*Mean Absolute Error*	*Root Mean Squared Error*	*Relative Absolute Error*	*Root Relative Squared Error*
Linear Regression Model	0.99	3.0	6.4	26%	11%
Additive Regression Model	0.98	3.5	11.1	30%	19%
Decision Stump Model	0.83	7.6	32.3	65%	55%

Table 4: Results for Murder per 100K of Population [Total Number of Instances - 2215]

Algorithm	*Correlation Coefficient*	*Mean Absolute Error*	*Root Mean Squared Error*	*Relative Absolute Error*	*Root Relative Squared Error*
Linear Regression Model	0.83	3.5	5.2	56%	56%
Additive Regression Model	0.88	2.6	4.4	41%	48%
Decision Stump Model	0.67	3.9	6.8	61%	74%

Table 5: Results for Rape
[Total Number of Instances - 2007]

Algorithm	*Correlation Coefficient*	*Mean Absolute Error*	*Root Mean Squared Error*	*Relative Absolute Error*	*Root Relative Squared Error*
Linear Regression Model	0.98	7.6	17.3	22%	16%
Additive Regression Model	0.96	12.4	27.9	36%	26%
Decision Stump Model	0.76	23.8	68.1	69%	64%

Table 6: Results for Rape per 100K of Population [Total Number of Instances - 2007]

Algorithm	*Correlation Coefficient*	*Mean Absolute Error*	*Root Mean Squared Error*	*Relative Absolute Error*	*Root Relative Squared Error*
Linear Regression Model	0.73	15.7	23.3	61%	68%
Additive Regression Model	0.83	13.2	19.2	52%	56%
Decision Stump Model	0.64	18.2	26.3	71%	77%

Table 7: Results for Robberies
[Total Number of Instances - 2214]

Algorithm	*Correlation Coefficient*	*Mean Absolute Error*	*Root Mean Squared Error*	*Relative Absolute Error*	*Root Relative Squared Error*
Linear Regression Model	0.99	53.4	109	15%	5%
Additive Regression Model	0.98	109	333	30%	15%
Decision Stump Model	0.88	208	1081	55%	48%

Table 8: Results for Robberies per 100K of Population [Total Number of Instances - 2214]

Algorithm	*Correlation Coefficient*	*Mean Absolute Error*	*Root Mean Squared Error*	*Relative Absolute Error*	*Root Relative Squared Error*
Linear Regression Model	0.95	47	69.6	31%	30%
Additive Regression Model	0.91	59	99.5	39%	42%
Decision Stump Model	0.69	102	169	67%	72%

Table 9: Results for Assaults
[Total Number of Instances - 2202]

Algorithm	*Correlation Coefficient*	*Mean Absolute Error*	*Root Mean Squared Error*	*Relative Absolute Error*	*Root Relative Squared Error*
Linear Regression Model	0.99	107	224	25%	11%
Additive Regression Model	0.98	136	354	31%	18%
Decision Stump Model	0.91	262	810	60%	41%

Table 10: Results for Assaults per 100K of Population [Total Number of Instances - 2202]

Algorithm	*Correlation Coefficient*	*Mean Absolute Error*	*Root Mean Squared Error*	*Relative Absolute Error*	*Root Relative Squared Error*
Linear Regression Model	0.93	92.8	163	30%	37%
Additive Regression Model	0.88	128	210	42%	48%
Decision Stump Model	0.71	197	310	65%	71%

Table 11: Results for Violent Crimes per 100K of Population [Total Number of Instances - 1994]

Algorithm	***Correlation Coefficient***	***Mean Absolute Error***	***Root Mean Squared Error***	***Relative Absolute Error***	***Root Relative Squared Error***
Linear Regression Model	1	0.004	0.006	0.0009%	0.001%
Additive Regression Model	0.97	116	168	26%	27%
Decision Stump Model	0.78	276	379	62%	62%

The overall accuracy of the algorithms is based on the error values. The algorithm that had the greatest correlation coefficient value also generated the lowest error values among the three algorithms. The decision stump model was the least accurate of the algorithms for each of the nine goal features. The additive regression model seemed to be most prominent for all of the crimes per 100K population features except for the RobberiesPerPop, AssaultsPerPop, and ViolentCrimesPerPop features. One could infer that this is so due to the fact that the linear regression algorithm also factors in all other present features into its implementation. These other features aid in increasing the effectiveness of the algorithm. Also, aforementioned, the additive regression model adds the weights of each feature and factors that into its output evaluation. There are some values in the RobberiesPerPop, AssaultsPerPop, and ViolentCrimesPerPop features that are either missing or are equal to zero. Overall, given the associated task for this particular dataset and the features that have been provided, the linear regression algorithm is the most accurate of the three.

Finally, a comparison has been done between the model produced by WEKA and the crime statistical data provided by neighborhoodscout.com. Aforementioned, regression makes the best prediction for an upcoming event by using the mean value of past events. WEKA provides attribute information for the features that are highlighted in the attribute select viewer window as shown in Figure 6.

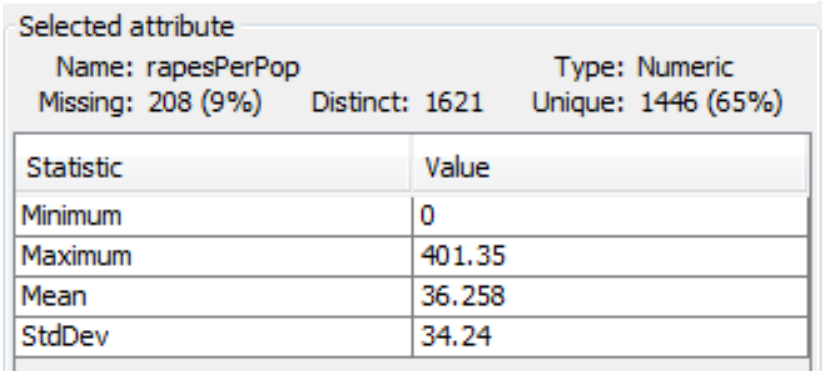

Figure 6: Selected Attribute Viewer Information Window

The mean values that were provided by WEKA were modeled for crimes that occurred for every 100K people in the population. The data that was provided by neighborhoodscout.com was modeled for occurrences for 1000 people in the population. For consistency, the crimes statistics from neighborhoodscout.com were scaled to 100K per population. The product of this rate and the mean values of the four crime per population features were used to scale the values with the neighborhoodscout.com 100K population crime totals. These projections for the MurderPerPop and RapePerPop features are relatively close, unlike the RobbPerPop and AssaultPerPop features whose residuals are greater. This is due to the fact that the error values from the robbery and assault features are greater than that of the murder and rape features.

Table 12: WEKA Crime Totals

	Murder	Rape	Robbery	Assault
Mean per 100K	6	36	62	178
Projected Crime Totals	179	1,076	2,485	4,635

Table 13: neighborhoodscout.com Crime Totals

	Murder	Rape	Robbery	Assault
Mean per 100K	7	31	81	156
Estimated Crime Totals	195	930	2,409	4,680

5. CONCLUSIONS

We observe the linear regression algorithm to be very effective and accurate in predicting the crime data based on the training set input for the three algorithms. The relatively poor performance of the Decision Stump algorithm could be attributed to a certain factor of randomness in the various crimes and the associated features (exhibits a low correlation coefficient among the three algorithms); the branches of the decision trees are more rigid and give accurate results only if the test set follows the pattern modelled. On the other hand, the linear regression algorithm could handle randomness in the test samples to a certain extent (without incurring too much of prediction error). Data mining has become a vital part of crime detection and prevention. Even though the scope of this project was to prove how effective and accurate machine learning algorithms can be at predicting violent crimes, there are other applications of data mining in the realm of law enforcement such as determining criminal "hot spots", creating criminal profiles, and learning crime trends. Utilizing these applications of data mining can be a long and tedious process for law enforcement officials who have to sift through large volumes of data. However, the precision in which one could infer and create new knowledge on how to slow down crime is well worth the safety and security of people.

REFERENCES

[1] Violent Crime. http://www2.fbi.gov/ucr/cius2009/offenses/violent_crime/.
[2] Murder. http://www2.fbi.gov/ucr/cius2009/offenses/violent_crime/murder_homicide.html.
[3] Forcible Rape. http://www2.fbi.gov/ucr/cius2009/offenses/violent_crime/forcible_rape.html.
[4] Robbery. http://www2.fbi.gov/ucr/cius2009/offenses/violent_crime/robbery.htm
[5] Assault. http://www2.fbi.gov/ucr/cius2009/offenses/violent_crime/aggravated_assault.html.
[6] Mississippi Crime Rates and Statistics - NeighborhoodScout. Mississippi Crime Rates and Statistics - NeighborhoodScout. Accessed February 17, 2015. http://www.neighborhoodscout.com/ms/crime/.
[7] S. M. Nirkhi, R.V. Dharaskar and V.M. Thakre. "Data Mining : A Prospective Approach for Digital Forensics," International Journal of Data Mining & Knowledge Management Process, vol. 2, no. 6 pp. 41-48, 2012.
[8] Weka 3: Data Mining Software in Java. http://www.cs.waikato.ac.nz/ml/weka/.

[9] E. W. T. Ngai, L. Xiu, and D. C. K. Chau. "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification," Expert Systems with Applications, pp. 2592–2602, 2008.
[10] J. McCarthy, "Arthur Samuel: Pioneer in Machine Learning," AI Magazine, vol. 11, no. 3, pp. 10-11, 1990.
[11] R. Bermudez, B. Gerardo, J. Manalang and B. Tanguilig, III. "Predicting Faculty Performance Using Regression Model in Data Mining," Proceedings of the 9th International Conference on Software Engineering Research, Management and Applications, pp. 68-72, 2011.
[12] S. Sathyadevan and S. Gangadharan. "Crime Analysis and Prediction Using Data Mining," Proceedings of the 1st International Conference on Networks and Soft Computing, pp. 406-412, 2014.
[13] Communities and Crime Un normalized Dataset. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Communities and Crime Unnormalized.