Sam Pennington and Brianna King

Data Mining – Fall 2018

Analysis of Potential Correlation in Atlanta

Weather And Crime

**Introduction**

It goes without saying that the city of Atlanta is a dangerous place. According to a recent article published by the AJC, the Atlanta police department is tracking up to 1000 different street gangs working within the limits of metropolitan Atlanta. While law enforcement attempts to combat this issue, they are at times simply overwhelmed by the sheer numbers of the gang population and as a result, the gangs continue to grow in size and number. This myriad of gangs is a large part of why Atlanta has seen a high rate of crime over the last several years. Our project will attempt to draw a correlation between crimes and weather data for the 2017 year in Atlanta. We will evaluate features of data including type of crime committed, type of weather on a given day, temperature, rate of crime committed on a day, and location of the crime committed (Lets try to plot the locations on a map if possible)

**Prior Studies**

Many studies have been done attempting to correlate the weather with crime. A set of data published on a WeatherOps blog details statistics tracking various types of crime in Chicago. Temperatures measured ranged from -2 to 102 degrees farenheight. These graphs were split by crime type which included Assault, Battery, Narcotics, and Stalking. This study focused on crimes of opportunity and was unable to draw a definitive conclusion about temperature and crime other than that crimes of opportunity decrease significantly when the temperature is either really hot (>90 degrees) or very cold (<30 degrees). Crime is shown to steadily rise in the temperature range in between these two values for Assault and Battery while spiking around the 30 degree mark and staying rather steady until the 90 degree point.

Another study conducted through the Chicago tribune also focused on crime in Chicago. It looked at crime data from 2012-2017 including lesser crimes such as theft and more serious crimes such as homicide. The results showed that all measured crimes had a positive, linear correlation to temperature increases with the exception of homicides. Shootings had the highest rate of increase per degree temperature rise averaging an additional 5 shootings per degree rise, while narcotics had the flattest slope being almost negligible. Homicides were shown to not correlate linearly with temperature and were mistakenly graphed as a flat line.

Studies have also been done based on the premise of how humidity and rainfall correlate to crime. One such study, cataloged in the British Journal of Criminology, was unable to draw any definitive conclusions about how rain and crime were correlated as the experimented was performed on three separate datasets and drew three separate conclusions. One experiment showed a positive correlation, one showed negative, and one showed negligible correlation. Humidity was also studied with slightly more conclusive results. An experiment performed on multiple different sets of data showed a slightly negative correlation between humidity and crime rate.

**Data Set**

Our overall dataset consists of two subsets of data. The first subset is a comma separated value file detailing crimes that were committed within the Atlanta area. This dataset was found on Kaggle.com[[1]](#footnote-1) and is based off an original data set published by the Atlanta City Police Department. This ACPD data was compiled, cleaned, and republished for our usage. This dataset contains information about the type of crime committed, the location it was committed, the time is was reported, the number of subjects involved, etc.

Our second subset contains data about the weather. This data was measured by NOAA (National Oceanic and Atmospheric Association) from the Fulton County government station. Our data was obtained by submitting a request to the organization for daily weather data for Atlanta on each day in 2017. This information can be requested through their online portal[[2]](#footnote-2). This data includes maximum and minimum temperatures, precipitation information, windspeed, humidity, etc. Measurements were taken approximately every 30 minutes and the data can be averaged together on a day by day basis to do a more generalized analysis.

**Data Cleansing**

The data sets that we used came with a lot of information that we did not deem necessary. As a result, we decided it was best to not waste memory or load time within our Python notebook, and instead deleted the columns of unnecessary data entirely. After going through each of the two spreadsheets, we narrowed down the data to the following columns:

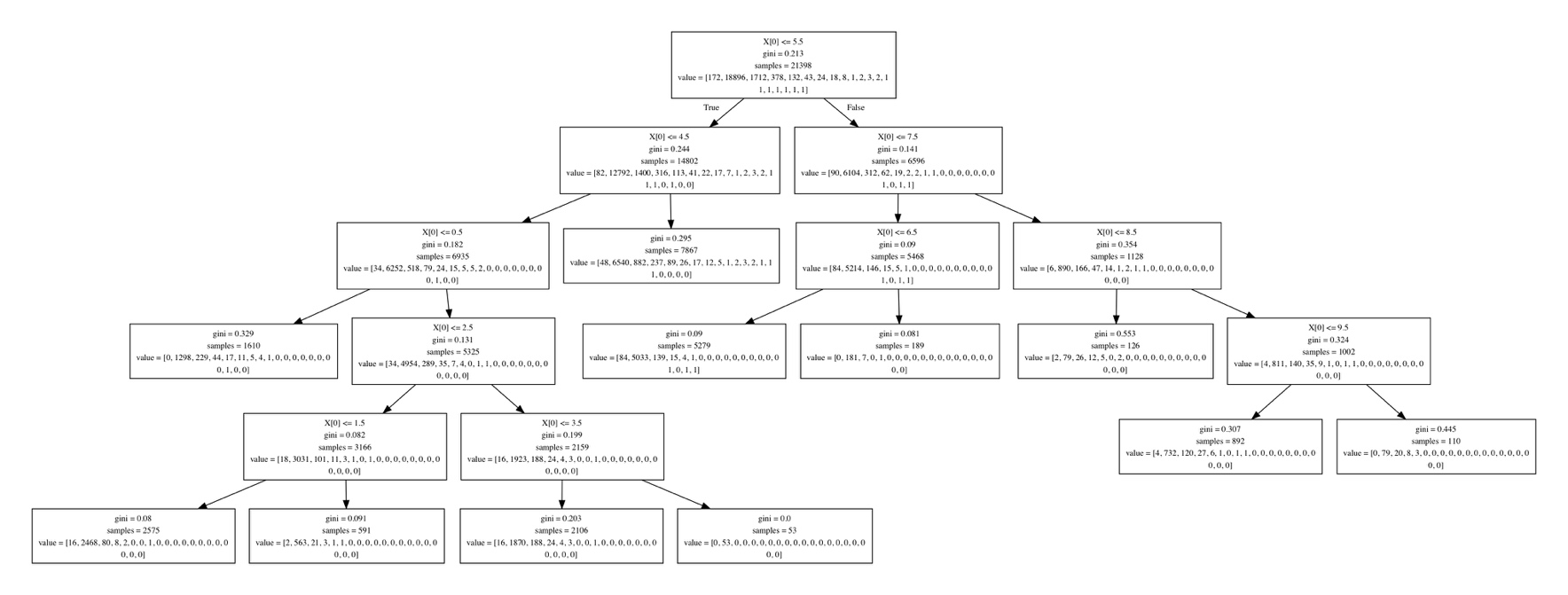
* Crime Data: Date, occurrence time, possible date, possible time, maximum number of victims, the day of the week, the string literal of the type of crime, the neighborhood, and the geographic coordinates where the crime occurred
* Weather Data: Data, visibility, dry bulb temperature in farenheight, relative humidity, wind speed, atmospheric pressure, and the sunrise and sunset times

The weather data needed further cleansing because the data was taken multiple times each day. To overcome this, we averaged the results of each category within each day. These results were then saved in place of the original data, resulting in averaged data for each day of the year.

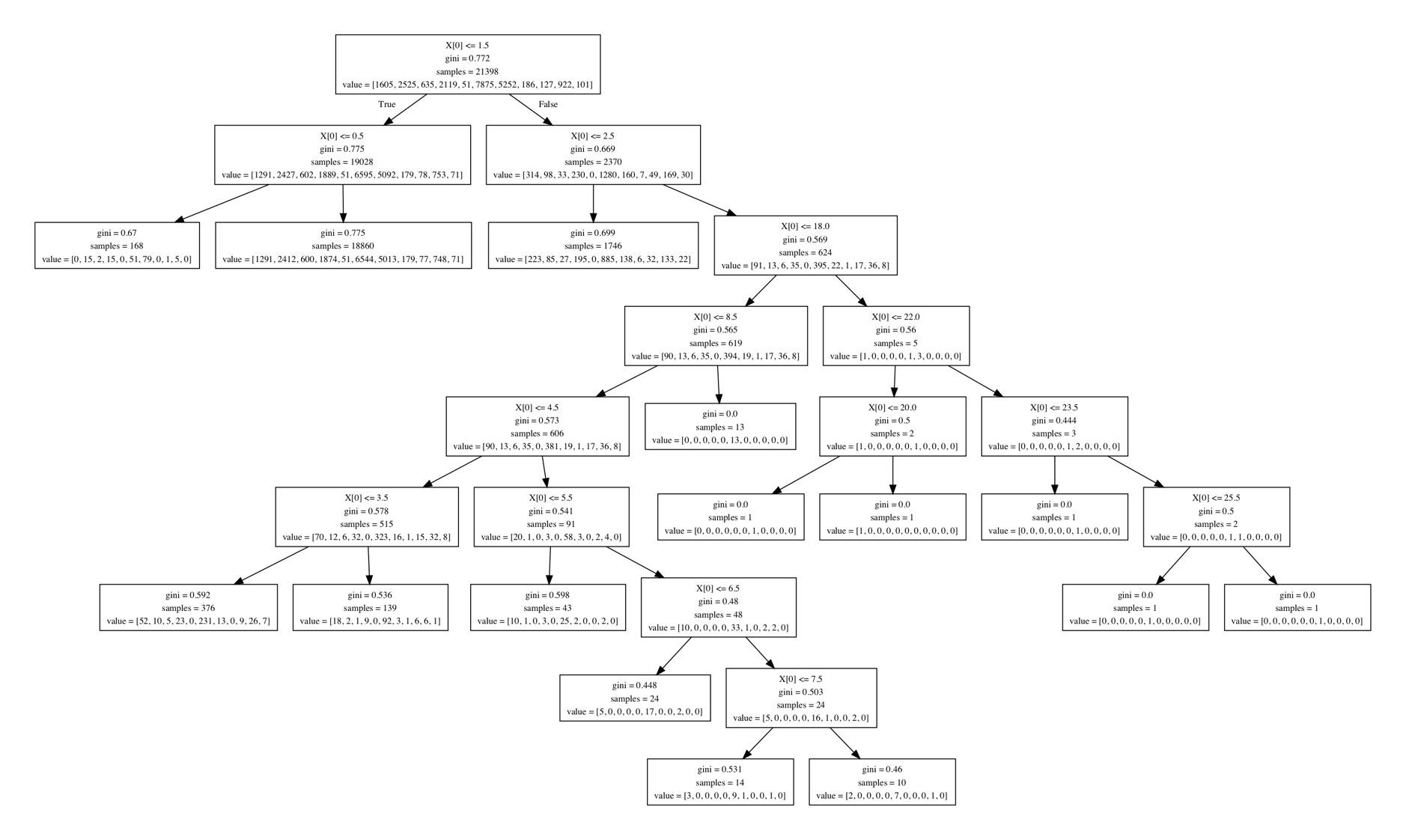
Within the crime data set, we used a code creator to assign a code to each of the unique string objects within each data feature. This was done in order to perform our classification models and tests. There were also some blank entries within certain rows and columns. To smooth out the data, we decided to take the average of all other values in the column and fill in the missing values with that new value. This is a much safer way to handle missing data rather than filling in “0” values. It also maximizes the amount of data points in the set by not deleting all rows with a single missing value.

**Exploratory Analysis**

**Decision Tree Results**



X[0] =



X[0] =

\*Other decision trees were created but were too large to show clearly within a document.

Sources

<https://www.ajc.com/news/crime--law/georgia-gang-situation-getting-worse-investigators-say/hf4OxArZ9YL8UaUQ9yP6jJ/>

<https://blog.weatherops.com/how-does-the-weather-affect-crime-rates>

<https://www.chicagotribune.com/news/data/ct-crime-heat-analysis-htmlstory.html>

1. https://www.kaggle.com/priscillapun/crime-in-atlanta-2017?fbclid=IwAR2P\_xMijBLjw4UtxWwqKuoFtARy1WqCG5GqU\_n47iLV1SmLVAjGomVV9YQ#COBRA-YTD2017.csv [↑](#footnote-ref-1)
2. https://www.ncdc.noaa.gov/cdo-web/datatools/lcd [↑](#footnote-ref-2)