**CHAPTER 6**

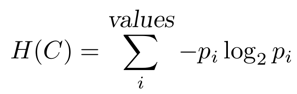
**PREDICTION OF EXISTING CRIMINALS**

The Supervised Machine Learning Algorithm ,Decision trees are used to initially split the dataset into two since some criminals may be arrested and possibly inside captivity ,so such tuples must be avoided from the dataset. This reduces the size of the dataset and reduces access time which contributes greatly towards error reduction.

**6.1 Decision Trees**

The Dataset is classified based on the Iterative Dichotomiser 3 metric 'Entropy'. On results the attribute 'Arrest' is found to have the minimum entropy. Thus the dataset is split based on the attribute 'Arrest'.

**Entropy Formula:**



**Algorithm:**

GrowTree (D,F)

Input: data D; set of features F.

Output: feature tree T with labeled leaves

1. If Homogeneous (D) then return Label (D);

2. S 🡨 BestSplit (D,F);

3. Split D into subset Di according to the literals in S;

4. For each i do

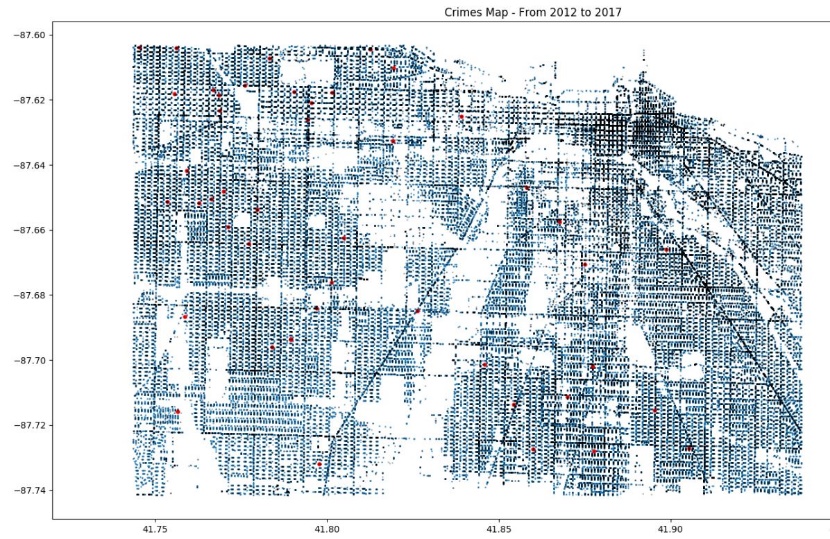
* + 1. If Di ≠ ⦸ then Ti 🡨 GrowTree (Di, F) else Ti is a leaf labeled with Label Di

7. End

8. Return a tree whose root is labeled with S and whose children are T(i).

|  |  |
| --- | --- |
| **Attribute** | **Entropy value** |
| Arrest | 0.591441 |
| Domestic | 0.650123 |
| Primary Type | 0.900451 |
| Ward | 0.951147 |

**Table 2. Net Entropy Results**



**Fig 1. Crime Heat Map**

After splitting the dataset, the distinct types of crimes are categorized and the particular area of criminal heat are recognized. This serves as a base for the prediction of existing criminal activity.

**6.2 Regressive Fine Tuning**

After obtaining the addresses of the different type of crimes and heat locations the corresponding areas are segregated and classified further using Naïve Bayes classifier.

**6.2.1 Naïve Bayes**

**Naïve Bayes** is done in order to tell the addresses which are prone to specific type of crimes. The accuracy was not well achieved but the classification serves as the basis for prediction. Regressive Fine Tuning is adopted because Naive Bayes is particularly applicable for textual data and our addresses serve as the input to this classifier.

Training set of Documents ,D={d1,d2,d3,...,d4}

Fixed set of classes, C={c1,c2,c3,...,c4}

Training set of labeled documents, Dlabeled={(d1,c1),...,(di,ci)}

This classification categorizes the address data into different location based on frequency of certain criminal activity.

|  |  |
| --- | --- |
| **Crime Type** | **Accuracy (%)** |
| Theft | 32.11 |
| Stalking | 27.36 |
| Public Peace Violation | 29.31 |
| Robbery | 45.21 |
| Battery | 31.11 |
| Arson | 61.22 |

**Table 3. Accuracy of Naïve Bayes Classifier**

**6.3 Rules Based Prediction**

For predicting a crime committed by an existing criminal is predicted using Rule-Based Prediction (Decision Trees and Naïve Bayes). But for predicting a “crime” which is going to happen in a completely different area, where crime did not take place at all, we would be going for Apriori Algorithm and RFM prediction.