**CHICAGO CRIME DATA ANALYSIS**

**INTRODUCTION** :

**DATASET SELECTED** : “Crimes in Chicago”

**ISSUE ADDRESSED** : This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2012 to 2017, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. In order to protect the privacy of crime victims, addresses are shown at the block level only and specific locations are not identified.

**WHY IS THIS A BIG DEAL?** : The predictions made from this dataset are of great relevance because they show the trend of repeated crimes at respective locations and respective time periods. This enables the reader to properly judge the criminal activities in a particular area and decide on preventive measures for the citizens staying in those areas by making them aware about their surroundings.

**CONTENT** :

**Number of columns** : 23

**Number of rows** : 1456714

**Column names and description** :

* **ID -** Unique identifier for the record.
* **Case Number -** The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.
* **Date -** Date when the incident occurred. this is sometimes a best estimate.
* **Block -** The partially redacted address where the incident occurred, placing it on the same block as the actual address.
* **IUCR -** The Illinois Uniform Crime Reporting code. This is directly linked to the Primary Type and Description. See the list of IUCR codes at <https://data.cityofchicago.org/d/c7ck-438e>.
* **Primary Type -** The primary description of the IUCR code.
* **Description -** The secondary description of the IUCR code, a subcategory of the primary description.
* **Location Description -** Description of the location where the incident occurred.
* **Arrest -** Indicates whether an arrest was made.
* **Domestic -** Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.
* **Beat -** Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts. See the beats at <https://data.cityofchicago.org/d/aerh-rz74>.
* **District -** Indicates the police district where the incident occurred. See the districts at <https://data.cityofchicago.org/d/fthy-xz3r>.
* **Ward -** The ward (City Council district) where the incident occurred. See the wards at <https://data.cityofchicago.org/d/sp34-6z76>.
* **Community Area -** Indicates the community area where the incident occurred. Chicago has 77 community areas. See the community areas at <https://data.cityofchicago.org/d/cauq-8yn6>.
* **FBI Code -** Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS). See the Chicago Police Department listing of these classifications at <http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html>.
* **X Coordinate -** The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.
* **Y Coordinate -** The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.
* **Year -** Year the incident occurred.
* **Updated On -** Date and time the record was last updated.
* **Latitude -** The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
* **Longitude -** The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
* **Location -** The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.

**METHODOLOGY :**

The systematic approach that has been followed through out the prediction of the model using various algorithms is as follows:

* **Data Extraction** : Importing the dataset in a dataframe.
* **Data Preprocessing** : Employing different imputation methods in order to clean and validate the data. Also dividing the data into train and test set.
* **Feature Selection** : Most correlated features with the target column were selected in order to get the best accuracy.
* **Exploratory Data Analysis** : Analysing different features of the dataset using heatmaps and different graphs.
* **Data modelling** : Predicting the best model by employing various modelling techniques i.e various algorithms to check the accuracy score, precision, recall and f1 score respectively.
* **Hyper Parameter Tuning** : Tuning the hyper parameters in order to get the best accuracy score.

**ALGORITHMS USED :**

RESULTS OF DIFFERENT ALGORITHMS AND RESPECTIVE METRICS**:**

**RANDOM FOREST CLASSIFIER REPORT :**

precision recall f1-score support

BURGLARY 1.00 1.00 1.00 1120

NARCOTICS 1.00 1.00 1.00 1836

THEFT 1.00 1.00 1.00 4460

DECEPTIVE PRACTICE 1.00 1.00 1.00 943

ROBBERY 1.00 1.00 1.00 802

CRIMINAL DAMAGE 1.00 1.00 1.00 2192

BATTERY 1.00 1.00 1.00 3674

ASSAULT 0.99 1.00 1.00 1259

WEAPONS VIOLATION 1.00 1.00 1.00 243

OFFENSE INVOLVING CHILDREN 0.99 0.89 0.94 152

CRIMINAL TRESPASS 1.00 1.00 1.00 526

OTHER OFFENSE 0.98 1.00 0.99 1232

MOTOR VEHICLE THEFT 1.00 1.00 1.00 850

INTERFERENCE WITH PUBLIC OFFICER 0.95 1.00 0.98 105

PROSTITUTION 1.00 0.99 0.99 100

PUBLIC PEACE VIOLATION 1.00 0.98 0.99 202

SEX OFFENSE 1.00 0.91 0.95 65

CRIM SEXUAL ASSAULT 0.94 0.93 0.94 89

OTHERS 0.90 0.85 0.87 110

HOMICIDE 1.00 0.95 0.97 40

micro avg 1.00 1.00 1.00 20000

macro avg 0.99 0.97 0.98 20000

weighted avg 1.00 1.00 1.00 20000

**NEURAL NETWORK REPORT** :

precision recall f1-score support

BURGLARY 1.00 1.00 1.00 1120

NARCOTICS 1.00 0.96 0.98 1836

THEFT 1.00 1.00 1.00 4460

DECEPTIVE PRACTICE 0.93 0.86 0.89 943

ROBBERY 1.00 1.00 1.00 802

CRIMINAL DAMAGE 1.00 1.00 1.00 2192

BATTERY 0.99 0.97 0.98 3674

ASSAULT 0.95 0.98 0.97 1259

WEAPONS VIOLATION 0.97 0.92 0.95 243

OFFENSE INVOLVING CHILDREN 0.84 0.54 0.66 152

CRIMINAL TRESPASS 1.00 0.81 0.89 526

OTHER OFFENSE 0.76 0.98 0.85 1232

MOTOR VEHICLE THEFT 1.00 0.95 0.98 850

INTERFERENCE WITH PUBLIC OFFICER 0.64 1.00 0.78 105

PROSTITUTION 0.87 0.77 0.81 100

PUBLIC PEACE VIOLATION 1.00 0.80 0.89 202

SEX OFFENSE 0.86 0.58 0.70 65

CRIM SEXUAL ASSAULT 0.80 1.00 0.89 89

OTHERS 0.80 0.65 0.72 110

HOMICIDE 0.58 0.95 0.72 40

micro avg 0.96 0.96 0.96 20000

macro avg 0.90 0.89 0.88 20000

weighted avg 0.97 0.96 0.97 20000

**KNN REPORT :**

precision recall f1-score support

BURGLARY 1.00 1.00 1.00 1120

NARCOTICS 1.00 1.00 1.00 1836

THEFT 1.00 1.00 1.00 4460

DECEPTIVE PRACTICE 1.00 1.00 1.00 943

ROBBERY 1.00 1.00 1.00 802

CRIMINAL DAMAGE 1.00 1.00 1.00 2192

BATTERY 1.00 1.00 1.00 3674

ASSAULT 1.00 1.00 1.00 1259

WEAPONS VIOLATION 1.00 1.00 1.00 243

OFFENSE INVOLVING CHILDREN 1.00 1.00 1.00 152

CRIMINAL TRESPASS 1.00 1.00 1.00 526

OTHER OFFENSE 1.00 1.00 1.00 1232

MOTOR VEHICLE THEFT 1.00 1.00 1.00 850

INTERFERENCE WITH PUBLIC OFFICER 0.99 1.00 1.00 105

PROSTITUTION 1.00 1.00 1.00 100

PUBLIC PEACE VIOLATION 1.00 1.00 1.00 202

SEX OFFENSE 1.00 0.98 0.99 65

CRIM SEXUAL ASSAULT 0.99 1.00 0.99 89

OTHERS 0.97 0.98 0.98 110

HOMICIDE 1.00 0.95 0.97 40

micro avg 1.00 1.00 1.00 20000

macro avg 1.00 1.00 1.00 20000

weighted avg 1.00 1.00 1.00 20000

**FINAL RESULT :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classifier Name** | **Accuracy\_Score** | **Precision** | **F1 Score** | **Recall** |
| Random Forest Classifier | 0.99935 | 0.9993529954316175 | 0.99935 | 0.99935 |
| Neural Network | 0.96495 | 0.9701403614545303 | 0.96495 | 0.96495 |
| KNeighborsClassifier | 0.99935 | 0.9993529954316175 | 0.99935 | 0.99935 |