CRIME DATA ANALYSIS AND PREDICTION

OBJECTIVES

- OUR AIM IS TO EXPLORE THE GEOGRAPHIC , HISTORIC INFORMATION OBSERVED OVER A LONG TIME PERIOD IN ORDER TO PREDICT THE PROBABILITY OF CERTAIN TYPE OF CRIME AT A GIVEN LOCATION AND TIME.
- TO CREATE HOTSPOTS
- CLASSIFY CRIME BASED ON THE LOCATION
- UNDERSTAND CRIME PATTERN

WHAT IS CRIME PREDICTION AND ANALYSIS

- IT INVOLVES SYSTEMATIC ANALYSIS FOR IDENTIFYING AND ANALYSING PATTERNS AND TRENDS IN ORDER TO PREDICT THE CHANCES OF CRIME TYPE AT A PARTICULAR REGION AND TIME.
- IT ANALYSES CRIME REPORTS, ARREST REPORTS ETC., TO IDENTIFY EMERGING PATTERNS OF CRIME IN A CERTAIN LOCATION.

ENVIRONMENT AND SOFTWARE

- GOOGLE COLAB
- PYTHON (3.7)
- JUPYTER NOTEBOOK
- PANDAS (0.22.1)
- NUMPY (1.14.2)
- SKLEARN (0.19.1)
- MATPLOTLIB (3.1.3)

DATASET WALKTHROUGH

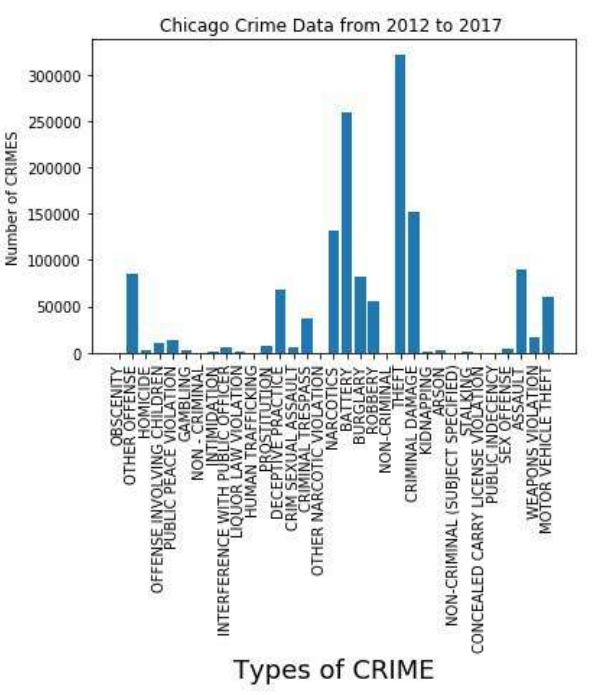
- THE DATA IS TAKEN FROM KAGGLE (A DATA SCIENCE PLATFORM). ORIGINALLY, DATA IS EXTRACTED FROM THE CHICAGO POLICE DEPARTMENT'S CLEAR (CITIZEN LAW ENFORCEMENT ANALYSIS AND REPORTING) SYSTEM.
- THE DATASET REFLECTS REPORTED INCIDENTS OF CRIME THAT OCCURRED IN THE CITY OF CHICAGO FROM 2012 TO 2017.
- INCLUDES INFORMATION REGARDING CRIME OCCURRENCE DATE,LOCATION,TYPE.
- THE DATASET CONTAINS 1418365 OBSERVATIONS AND 23 COLUMNS (16 `COLUMNS AFTER PRE-PROCESSING)

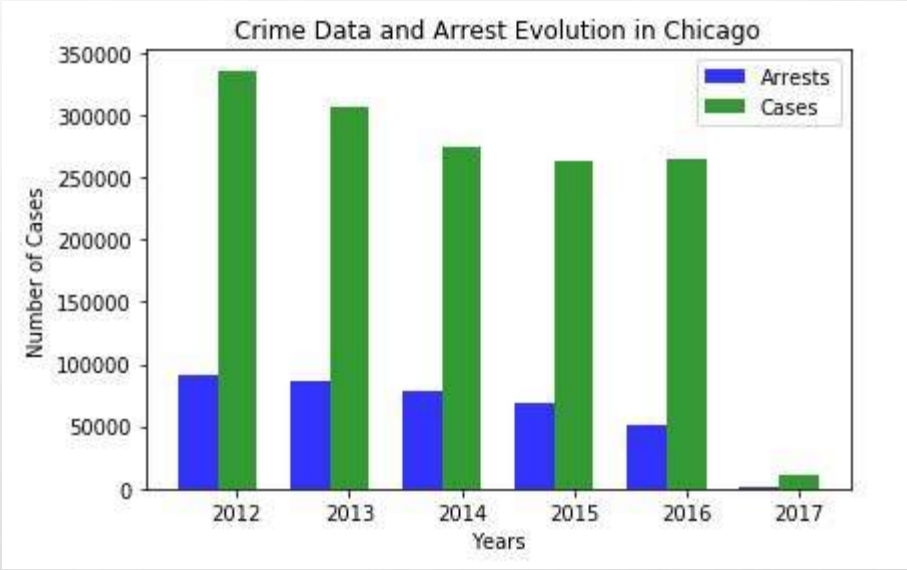
Dataset Link

WHAT OUR MODEL DOES?

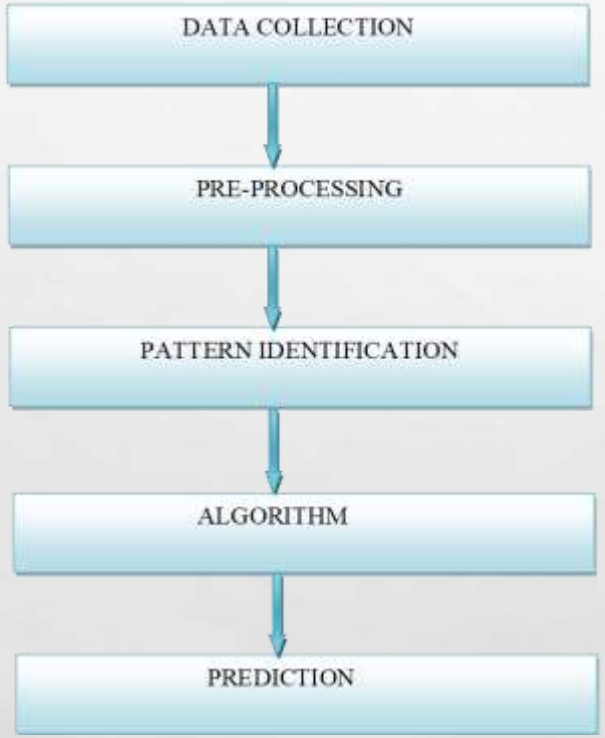
- GIVEN THE LOCATION AND TIME OUR MODEL PREDICTS THE POSSIBILITY OF A CERTAIN CRIME TYPE.
- CREATE HOTSPOTS IN A CITY OR A REGION FOR EFFICIENT PATROLLING.
- HAVE USED DIFFERENT CLASSIFIERS TO LOCATE HOTSPOTS OF CROOK ACTIVITIES PRIMARILY BASED ON THE TIME OF DAY

GRAPHS GENERATED





OUR APPROACH



WHY THIS APPROACH?

- NEAR-REPEAT THEORY IN CRIMINOLOGY RESEARCH STATES THAT THE CRIME EVENTS ARE MORE LIKELY TO HAPPEN IN THE VICINITY OF PAST CRIME EVENTS.
- IN THE LIGHT OF THIS,THE MODEL ASSIGNS THE PROBABILITY OF OCCURRENCE OF CRIME IN A PARTICULAR REGION BASED ON HISTORICAL AND GEOGRAPHICAL DATA.
- WITH THE ADVENT OF THE BIG DATA ERA AND THE AVAILABILITY OF FAST, EFFICIENT ALGORITHMS FOR DATA ANALYSIS, UNDERSTANDING PATTERNS IN CRIME FROM DATA IS AN ACTIVE AND GROWING FIELD OF RESEARCH.

RESULTS

Algorithm Used	Accuracy
KNN	85.47%
Random Forest	99.99%
XG Boost	91.18%
Bagging	92.51%
Logistic Regression	32.09%(until reaching limit)

FUTURE WORK

- TIME-SERIES MODELING OF THE DATA TO UNDERSTAND TEMPORAL CORRELATIONS IN IT, WHICH CAN THEN BE USED TO PREDICT SURGES IN DIFFERENT CATEGORIES OF CRIME
- ADDITIONAL PARAMETERS CAN BE USED FOR MORE ACCURATE PREDICTION
- GENERATING PARAMETERS FROM EXISTING DATA BY PERFORMING VARIOUS ALGEBRAIC OPERATIONS.
- ADDING NON-LINEARITY TO DATA.

TEAM DETAILS

- BINSHUMESH SACHAN (171210021)
- DIVYANSHI SINGH (171210022)
- GDV SUJITH SAGAR (171210023)
- T J VAMSI KESAV (171210055)

Thankyou