Crime Density using News Article Analysis

Shailendra Kr. Gupta - 2016CSB1059
Shreyanshu Shekhar - 2016CSB1060

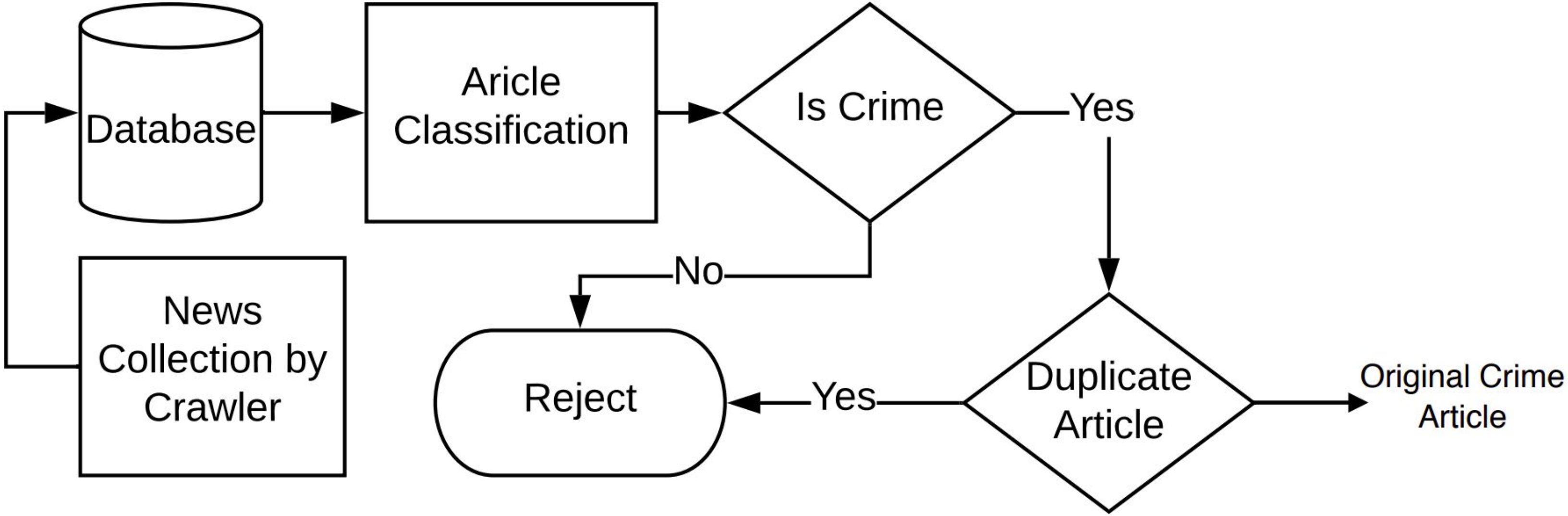
Supervisors:-

Neeraj Goel
Mukesh Saini

Problem Statement

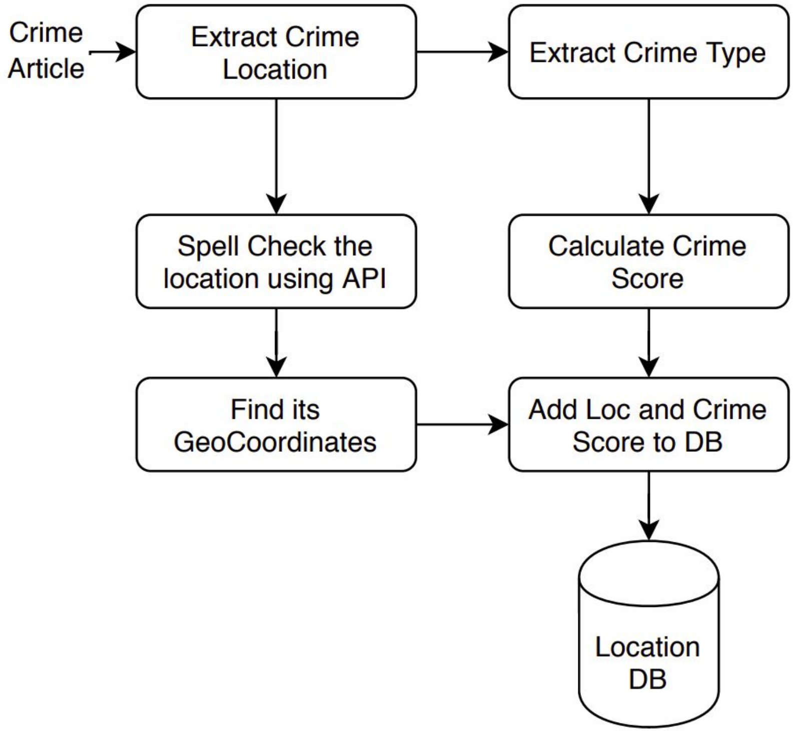
- Main objective:
  - Generate a heat map
  - Based on crime rate
- Use Case:
  - Finding safest route/place
  - Dynamically organising police force
  - Predicting the future occurrence of crime

Overall Framework Flow - I



Overall Framework Flow - II

- Extracted all crime locations
- Calculated their crime score
- Store them into the DB



Changes After The Paper (New Objectives)

- Crime Score of unknown location
- Updates in the database
- Duplicate Detection

Location Extraction

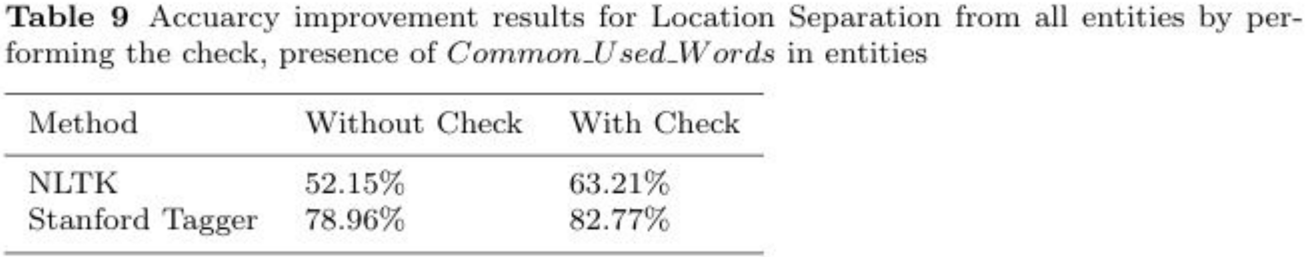


Table 9 Accuarcy improvement results for Location Separation from all entities by performing the check, presence of *Common_Used_Words* in entities

Method	Without Check	With Check
NLTK	52.15%	63.21%
Stanford Tagger	78.96%	82.77%

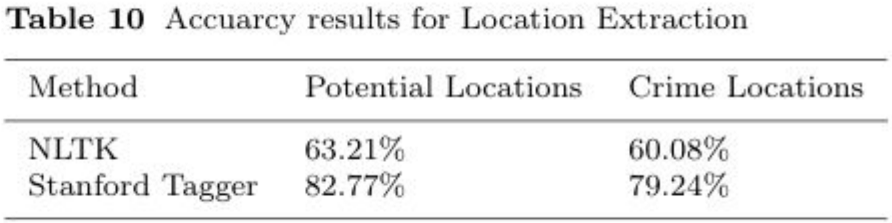


Table 10 Accuarcy results for Location Extraction

Method	Potential Locations	Crime Locations
NLTK	63.21%	60.08%
Stanford Tagger	82.77%	79.24%

Duplicate Detection

- Using both tf-idf and simhash method
- Simhash for better results
- Tf-idf for handling cases of large and small documents comparisons

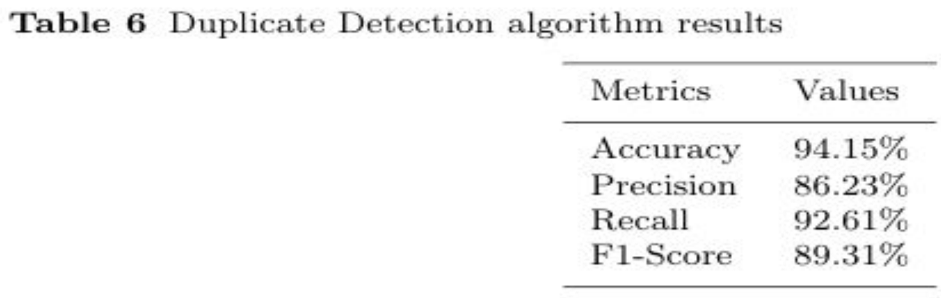


Table 6 Duplicate Detection algorithm results

Metrics	Values
Accuracy	94.15%
Precision	86.23%
Recall	92.61%
F1-Score	89.31%

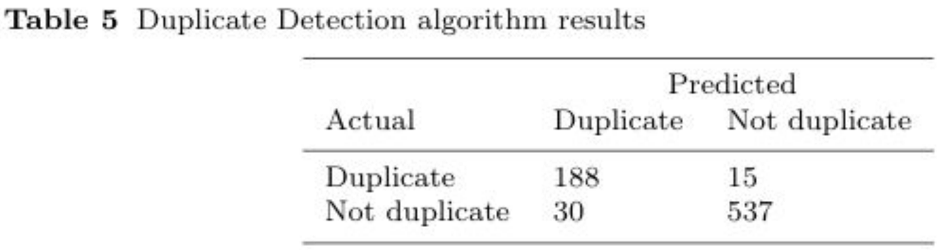


Table 5 Duplicate Detection algorithm results

	Predicted	
Actual	Duplicate	Not duplicate
Duplicate	188	15
Not duplicate	30	537

Duplicate Detection

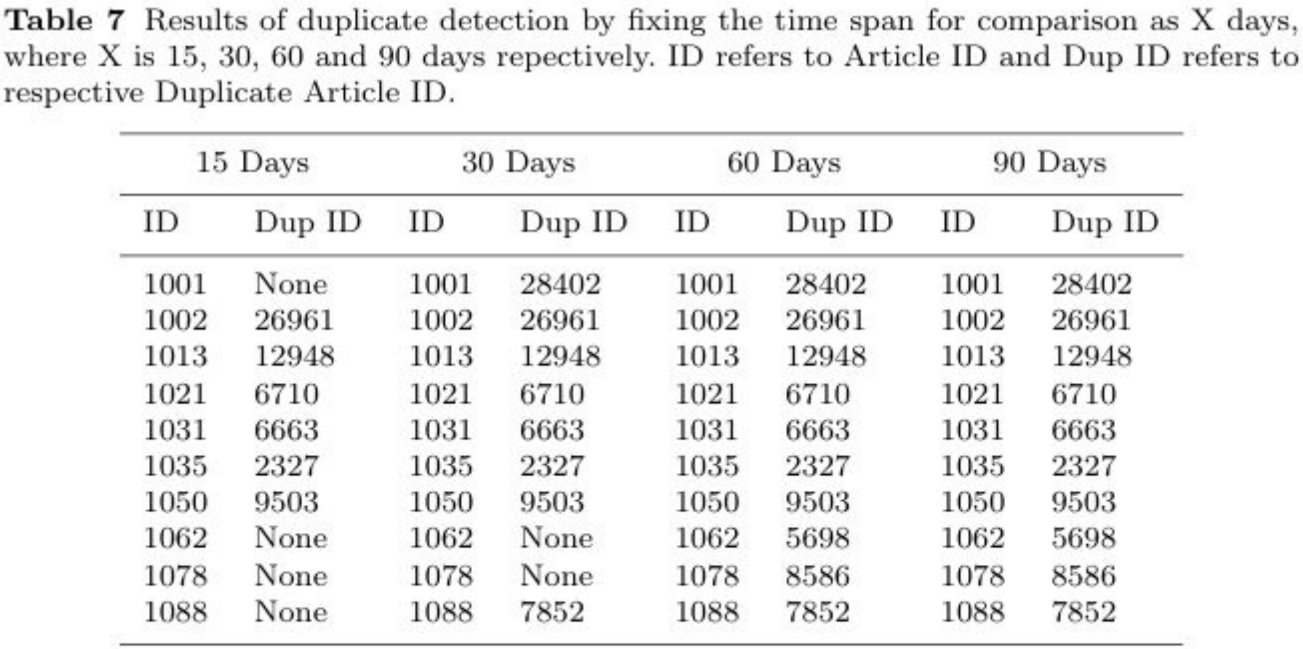


Table 7 Results of duplicate detection by fixing the time span for comparison as X days, where X is 15, 30, 60 and 90 days repectively. ID refers to Article ID and Dup ID refers to respective Duplicate Article ID.

15 Days		30 Days		60 Days		90 Days	
ID	Dup ID	ID	Dup ID	ID	Dup ID	ID	Dup ID
1001	None	1001	28402	1001	28402	1001	28402
1002	26961	1002	26961	1002	26961	1002	26961
1013	12948	1013	12948	1013	12948	1013	12948
1021	6710	1021	6710	1021	6710	1021	6710
1031	6663	1031	6663	1031	6663	1031	6663
1035	2327	1035	2327	1035	2327	1035	2327
1050	9503	1050	9503	1050	9503	1050	9503
1062	None	1062	None	1062	5698	1062	5698
1078	None	1078	None	1078	8586	1078	8586
1088	None	1088	7852	1088	7852	1088	7852

Duplicate Detection

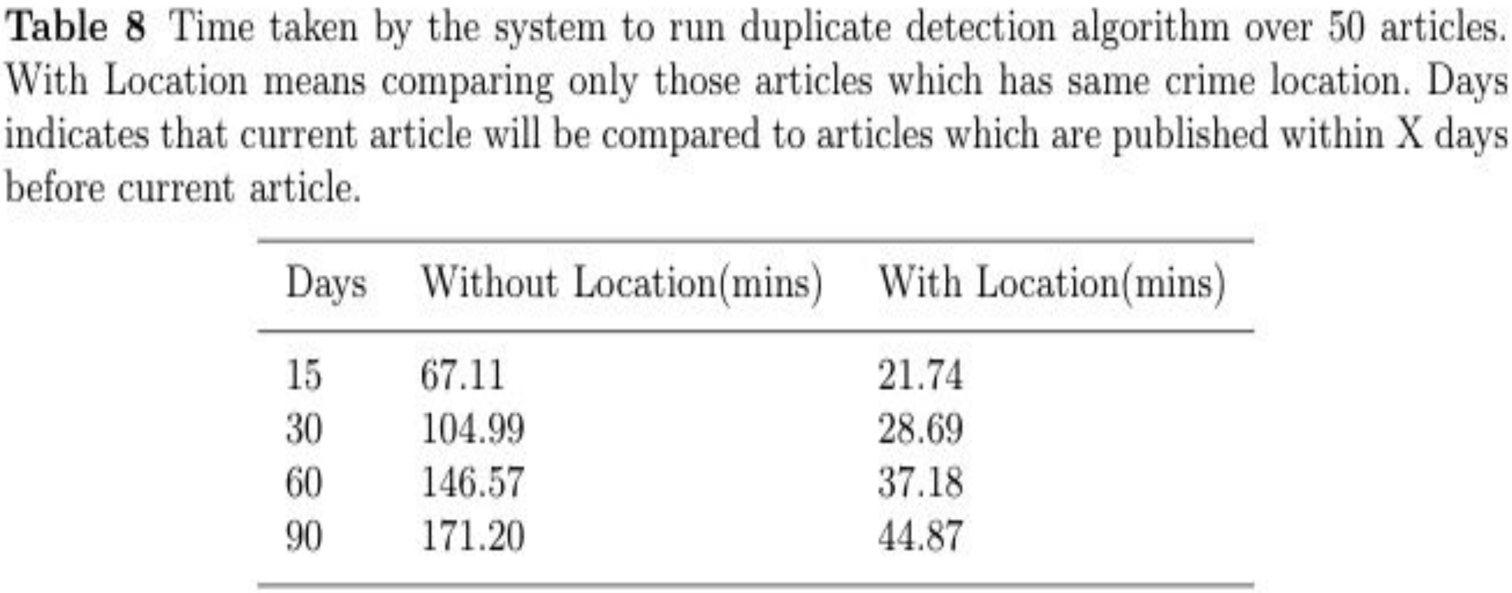
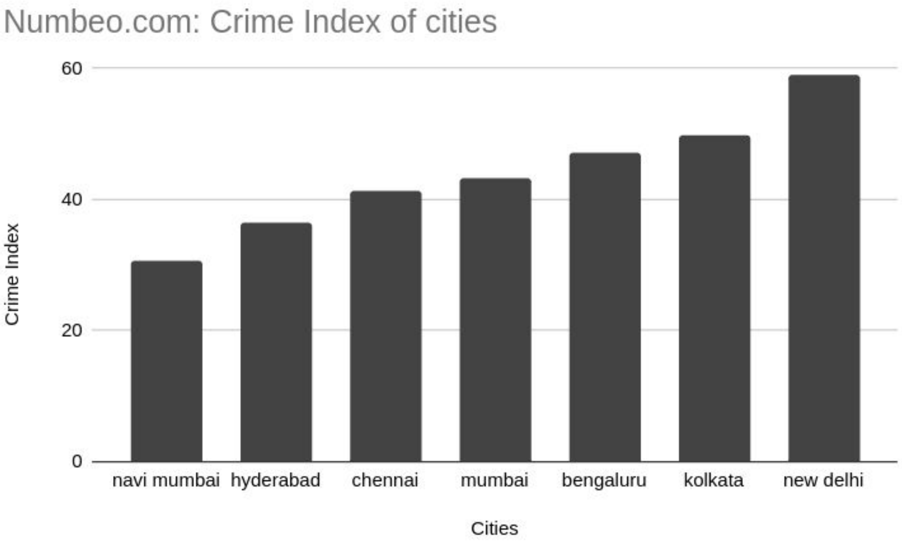
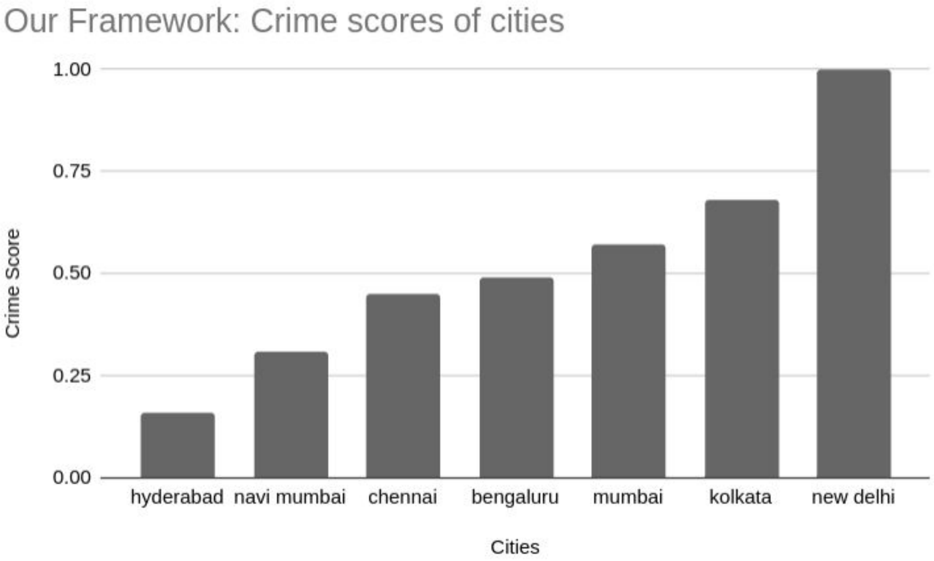


Table 8 Time taken by the system to run duplicate detection algorithm over 50 articles. With Location means comparing only those articles which has same crime location. Days indicates that current article will be compared to articles which are published within X days before current article.

Days	Without Location(mins)	With Location(mins)
15	67.11	21.74
30	104.99	28.69
60	146.57	37.18
90	171.20	44.87

Crime Score Verification


Numbeo.com: Crime Index of cities


Our Framework: Crime scores of cities

Crime Classification (ML Technique)

Data Partition ration	SVM	Naive Bayes
0.1	45.53	45.53
0.2	54.02	54.02
0.3	51.48	51.48

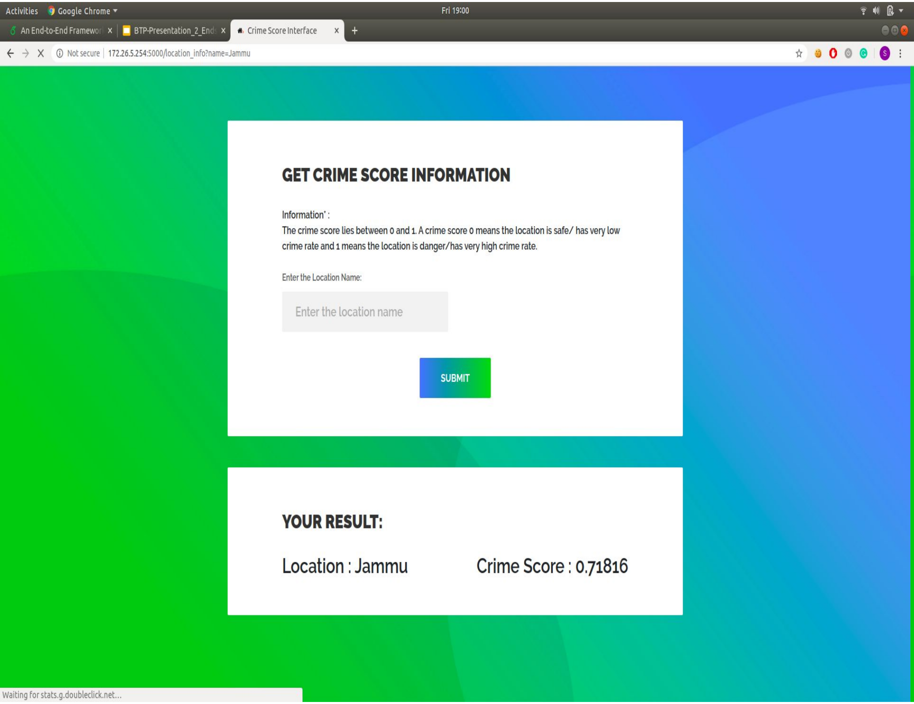
Total Data

- Total Articles : 345870
- Non-crime: 266624
- Crime Articles: 79246
- Crime Duplicate: 12096
- Crime Locations : 3311

Interface

- Crime score review web interface
- Using Python and Flask
    - Input - Location
    - Output - Crime Score
- Check out this: http://172.26.5.254:5000

Interface



GET CRIME SCORE INFORMATION

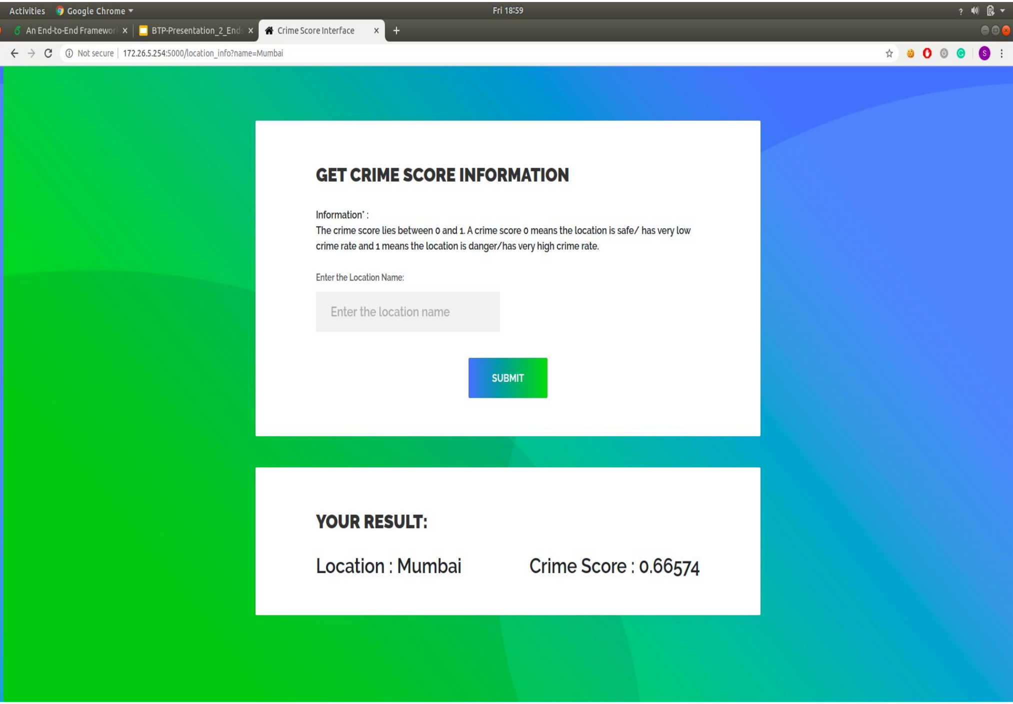
Information* :
The crime score lies between 0 and 1. A crime score 0 means the location is safe/ has very low crime rate and 1 means the location is danger/has very high crime rate.

Enter the Location Name:

Enter the location name

YOUR RESULT:

Location : Jammu Crime Score : 0.71816



GET CRIME SCORE INFORMATION

Information* :
The crime score lies between 0 and 1. A crime score 0 means the location is safe/ has very low crime rate and 1 means the location is danger/has very high crime rate.

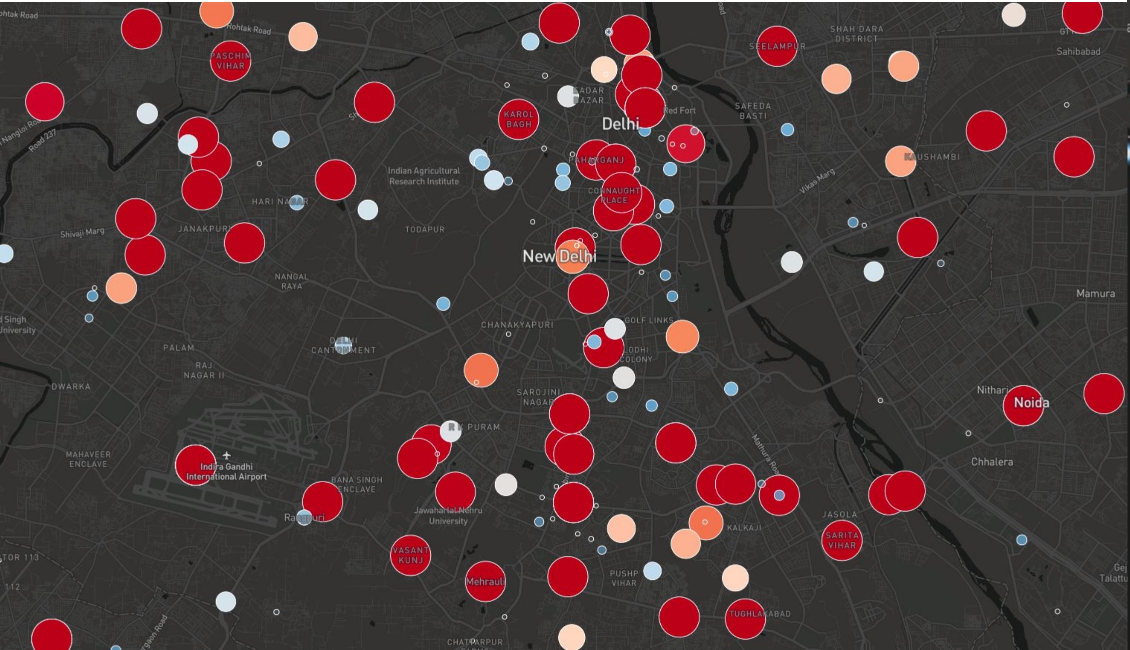
Enter the Location Name:

Enter the location name

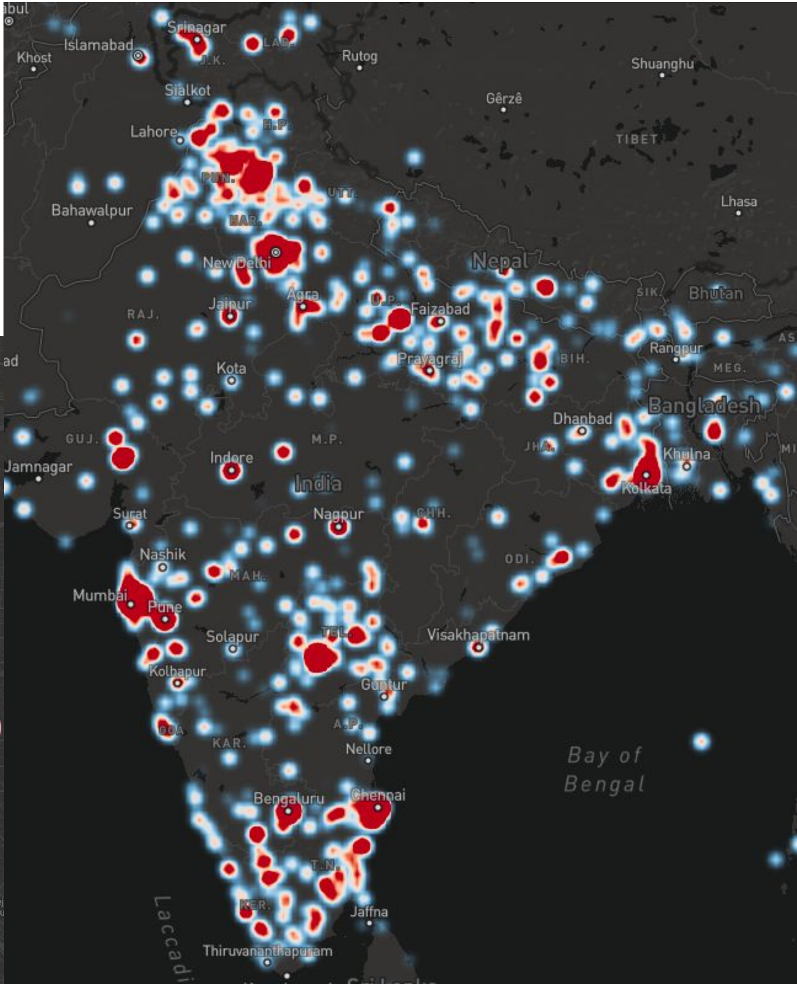
YOUR RESULT:

Location : Mumbai Crime Score : 0.66574

Heat Map



PASCHIM VIHAR
KAROL BAGH
Delhi
SADAR BAZAR
Red Fort
SEELAMPUR
SHAH DARA DISTRICT
Sahibabad
SAFEDA BASTI
PAHARGANJ
CONNAUGHT PLACE
Indian Agricultural Research Institute
HARI NAGAR
JANAKPURI
TODAPUR
KAUSHAMBI
Vikas Marg
New Delhi
NANGAL RAYA
CHANAKYAPURI
GOLF LINKS
LODHI COLONY
Mamura
PALAM
DELHI CANTONMENT
RAJ NAGAR II
DWARKA
SAROJINI NAGAR
R K PURAM
Nithari
Noida
MAHAVEER ENCLAVE
Indira Gandhi International Airport
BASANT ENCLAVE
Jawaharlal Nehru University
Chhalera
Mathura Road
Rajpuri
VASANT KUNJ
KALKAJI
JASOLA
SARITA VIHAR
Mehrauli
PUSHP VIHAR
TUGHLAKABAD
CHATTARPUR
Shivaji Marg
Rohtak Road



Srinagar
Islamabad
Khost
J.K.
LAD.
Rutog
Shuanghu
Sialkot
Gêrzê
Lahore
H.P.
TIBET
PUN.
UTT.
Lhasa
Bahawalpur
HAR.
New Delhi
Nepal
Agra
Jaipur
U.P.
Faizabad
SIK.
Bhutan
RAJ.
Prayagraj
BIH.
Rangpur
Kota
MEG.
Bangladesh
Dhanbad
GUJ.
M.P.
JHA.
Jamnagar
Indore
Khulna
India
Kolkata
Surat
Nagpur
CHH.
ODI.
Nashik
MAH.
Mumbai
Pune
Visakhapatnam
Solapur
TEL.
Kolhapur
Guntur
GOA
A.P.
KAR.
Nellore
Bay of Bengal
Bengaluru
Chennai
T.N.
KER.
Jaffna
Thiruvananthapuram
Laccadive

Crime Score of Unknown Location

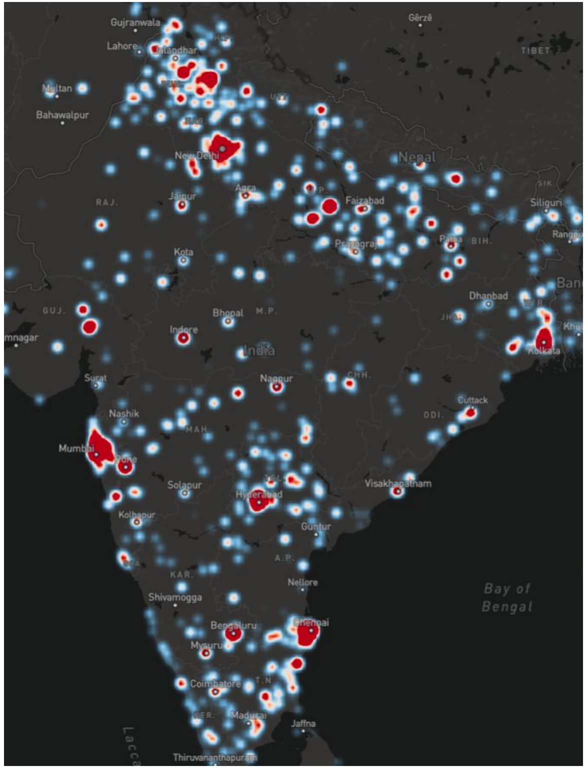
- Assuming Gaussian distribution of crime score
- Using the neighbour crime
- Calculate crime score for unknown location

Continuous Heat Map

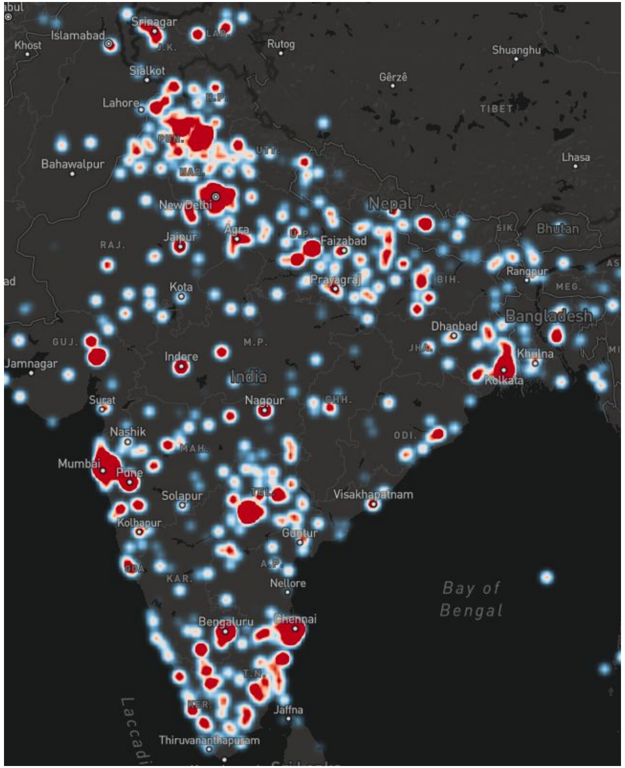
- We don’t have crime score of every location
- Finding the crime score of unknown location
- To fill the gaps in the heat map
- Took the geoCoordinate inside India
- With granularity of 0.1 degree

Heat Map Density Difference

BEFORE



AFTER



Questions and Answers

Thanks!