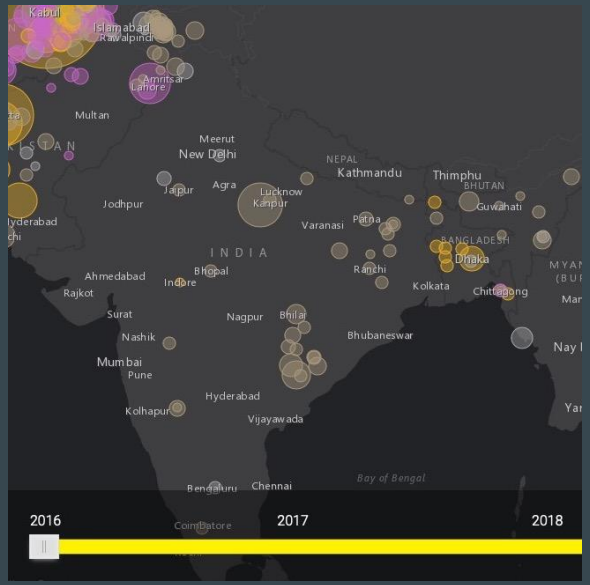
Counter Terrorism using Density Based Clustering and Naive Bayes Classification



•••

Gagandeep Singh (2016037)

Problem Statement

- Terrorism leads to reduced tourism and affects the economy of a country.
- Causes fear and anxiety among citizens.
- Leads to lots of causality every year.
- Terrorist attacks can be combat and controlled unlike other issues like climate change.
- Understand the reason behind terrorism and what triggers these groups.
- Understand the patterns in frequent attacks - preventive measures can be taken.

Dataset Used

- Kaggle dataset was used. Almost 35 years of data was available.
- The dataset was used as the dataset had 135 features ranging from geospatial coordinate to attacks, type of attacks, the type of people attacked etc.
- It had specificity feature as well which gave a clear idea as to how correct is the dataset were. With proper analysis suitable inference can be done.
- Approach : Data Mining Technique was used as there were 11000 records for India out of 2 lakh records. They were cleaned according to our interest.
- Suitable clustering technique used for group analysis..

d	iyear	imonth	iday	provstate	city	latitude	longitude	specificity	success	suicide	attacktype	targtype1	targsubty	gname	weaptype
201001070003.00	2010	1	7	Jharkhand	Salpatra	23.7014	86.58667	1	1	0	Hostage T	Violent Pc	Party Offi	Communi	Firearms
201001070004.00	2010	1	7	West Ben	Murabani	22.56916	86.97367	1	1	0	Hostage T	Terrorists/	Non-State	Communi	Firearms
201001150002.00	2010	1	15	Jharkhand	Gumla	23.04151	84.54449	1	1	0	Bombing/	Police	Police Sec	Communi	Explosive
201001150005.00	2010	1	15	Jharkhand	Gumla dis	23.04151	84.54449	3	1	0	Armed As	Police	Police Sec	Communi	Firearms
201001170002.00	2010	1	17	Bihar	Gaya	24.78147	84.98641	1	1	0	Bombing/	Telecomm	Telephon	Communi	Explosive
201001170003.00	2010	1	17	Bihar	Gaya	24.78147	84.98641	1	1	0	Facility/In	Private Ci	Vehicles/	Communi	Incendiar
201001170014.00	2010	1	17	Andhra pr	Khamman	17.24705	80.1493	1	1	0	Armed As	Terrorists	Terrorist	Unknown	Melee
201001180024.00	2010	1	18	Bihar	Nadaul	25.29663	85.01738	1	1	0	Bombing/	Transport	Train/Trai	Communi	Explosive
201001180027.00	2010	1	18	Punjab	Nabha	30.37367	76.14519	1	0	0	Bombing/	Utilities	Oil	Khalistan	Explosive
201001190001.00	2010	1	19	Jharkhand	East Singh	22.48668	86.49966	3	1	0	Armed As	Private Ci	Student	Communi	Firearms
201001190016.00	2010	1	19	Manipur	Imphal	24.79835	93.94043	1	1	0	Bombing/	Police	Police Sec	Unknown	Explosive
201001200003.00	2010	1	20	Manipur	Imphal	24.79835	93.94043	1	0	0	Armed As	Private Ci	Unnamed	Unknown	Firearms
201001200012.00	2010	1	20	Orissa	Koraput d	18.81524	82.7121	3	1	0	Bombing/	Telecomm	Telephon	Communi	Explosive
201001210004.00	2010	1	21	Manipur	Imphal	24.79835	93.94043	1	0	0	Bombing/	Private Ci	House/Ap	Unknown	Explosive
201001210005.00	2010	1	21	Manipur	Hiyangtha	24.72843	93.89672	1	0	0	Bombing/	Governme	Governme	United Ku	Explosive
201001210008.00	2010	1	21	Jharkhand	Palamu di	24.12861	84.18571	3	1	0	Facility/In	Telecomm	Telephon	Communi	Incendiar
201001210009.00	2010	1	21	Manipur	Bishenpur	24.55934	93.81465	2	1	0	Bombing/	Business	Mining	Unknown	Explosive
201001210011.00	2010	1	21	Manipur	Bishnupur	24.62872	93.76133	1	1	0	Bombing/	Governme	Governme	Unknown	Explosive
201001210012.00	2010	1	21	Manipur	Imphal	24.79835	93.94043	1	0	0	Bombing/	Private Ci	House/Ap	Kuki Liber	Explosive
201001220012.00	2010	1	22	Orissa	Koraput	18.81524	82.7121	1	1	0	Facility/In	Governme	Governme	Communi	Incendiar
201001230004.00	2010	1	23	Orissa	Sundargar	22.124	84.04318	3	1	0	Hostage T	Violent Pc	Party Offi	Communi	Unknown
201001230006.00	2010	1	23	Assam	Chirang di	26.63755	90.64405	3	1	0	Bombing/	Transport	Subway	Unknown	Explosive
201001230011.00	2010	1	23	Orissa	Koraput	18.81524	82.7121	1	1	0	Bombing/	Police	Police Sec	Communi	Explosive
201001240001.00	2010	1	24	Tripura	Bhagirath	23.69563	91.91281	1	0	0	Armed As	Governme	Politician	National L	Firearms

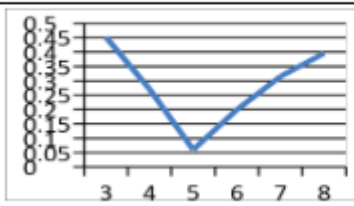
K-Means Clustering

- Using Apache Spark MLlib.
- Cluster Distance Performance operator of RapidMiner used to find out Davies-Bouldin index for value of K.
- Inference - 5 Spherical clusters obtained.

Disadvantages

- The outliers were also stored within the clusters.
- Spherical clusters were too huge and not accurate.
- Difficult for analysis.

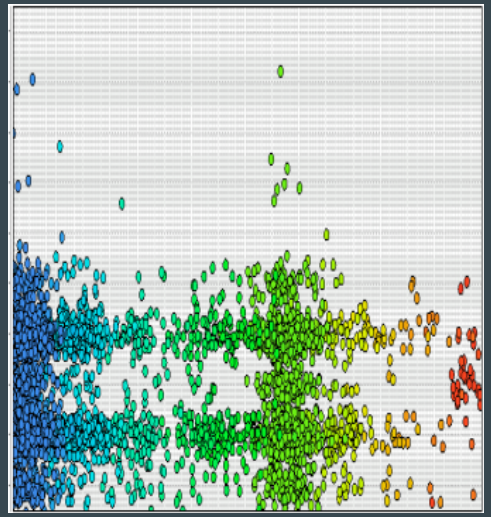
Graphical Representation to identify the value of K



Davies-Bouldin index index for K=3, 4,5,6,7,8

k	3	4	5	6	7	8
Davies Bouldin:	0.452	0.271	0.06	0.199	0.315	0.396

```
prediction,count
1,3048
3,3166
4,4108
2,7883
0,46895
```



Naive Bayes Classification

- Uses Bayes' Theorem to predict class probabilities and apache spark mllib.
- Features are considered independent of each other

Advantages

- Categorical and continuous values are supported.
- Only the variance of the variables for each label to be determined due to independent features.
- Intensity of computational memory and processing is low.
- The training time is also less compared to other methods.

- Outcome - Attack on Feb 10,2018 in Sunjuwan, away from Sopore.
    - Prediction - Attack on Feb 19, 2018 in Sopore, Jammu.
- Outcome - Attack on May 4, 2018 in Tinsukia, Assam.
    - Prediction - Attack in Ramnagar, 600 km away from Tinsukia.
- Outcome - Attack on Oct 20, 2018 in Imphal, Manipur.
    - Prediction - Attack in Wangoi, 17.2 km from Imphal.

Map-Reduce

Map reduce was use to find the countries and the number of attacks they have had.

```
('Mexico', '\t', 332.0)
('Moldova', '\t', 15.0)
('Morocco', '\t', 23.0)
('Mozambique', '\t', 219.0)
('Myanmar', '\t', 188.0)
('Namibia', '\t', 123.0)
('Nepal', '\t', 32.0)
('Netherlands', '\t', 105.0)
('New Caledonia', '\t', 31.0)
('New Hebrides', '\t', 1.0)
('New Zealand', '\t', 9.0)
('Nicaragua', '\t', 1952.0)
('Niger', '\t', 35.0)
('Nigeria', '\t', 49.0)
('North Korea', '\t', 1.0)
('North Yemen', '\t', 6.0)
('Norway', '\t', 11.0)
('Pakistan', '\t', 1593.0)
('Panama', '\t', 119.0)
('Papua New Guinea', '\t', 78.0)
('Paraguay', '\t', 27.0)
("People's Republic of the Congo", '\t', 4.0)
('Peru', '\t', 5986.0)
('Philippines', '\t', 1908.0)
('Poland', '\t', 30.0)
('Portugal', '\t', 138.0)
('Qatar', '\t', 3.0)
('Republic of the Congo', '\t', 11.0)
('Rhodesia', '\t', 83.0)
```

Result and Observation

- It has been observed from the clusters that the frequently targeted months are budget session months of the Parliament.
- Constitution Day (Samvidhan Divas) on 26 November is celebrated to commemorate the adoption of Constitution of India. Many clusters showed that there are frequent attacks on date 26 of months or around 27 ,28 of different months.
- There were several attacks in 2003 after Nov 23 when India accepted Pakistan’s offer of a ceasefire in Kashmir.
- The targets were mostly defense or police personnel.
- The attacks were repeated on specific dates like 11/23 and 11/25. The eastern region of Kashmir like the town of Sopore is affected several times because of the Pakistan border, as it is easy for the terrorists to enter.
- A pattern has also been recognized that the Eastern coast of India is attacked with terrorists finding their way to enter through Arabian Sea.

Thank You