# NEAR REAL-TIME ATROCITY EVENT CODING

Sai Vivek Kanaparthy
Shiva Ganga Reddy Chennu
Satya Aravind Chowdary Obellaneni
Sushma Eati
Manisha Galla

# INTRODUCTION:

- Atrocities, terrorism and political unrest proliferated human sufferings to a great extent in recent times. In this modern era, where technology is growing at a rapid speed, dreaming of a tool that predicts the occurrence of an Atrocity event would be of no surprise. This can be achieved with the help of Machine Learning Algorithms and Natural Language Processing techniques. To predict the occurrence of an Atrocity event we require structured data that contains meta data with fields and values.

- So in this paper we develop a near real time machine that detects the occurrence of an Atrocity event. We design a Spark Based Distribution Framework along with web scraper to generate and gather news. Raw text processing is performed more accurately using CoreNLP and the performance can be increased to a great extent with the help of SPARK.

# PROBLEM STATEMENT

- Atrocities, terrorism and political unrest proliferated human sufferings to a great extent in recent times. These activities adversely affect the life of common and innocent people leaving a negative impact on their livelihood, families etc. In this modern era, where technology is growing at a rapid speed, dreaming of a tool that predicts the occurrence of an Atrocity event would be of no surprise. This can be achieved with the help of Machine Learning Algorithms and Natural Language Processing techniques. To predict the occurrence of an Atrocity event we require structured data that contains meta data with fields and values. This process of prediction with utmost precision can be done with the help of common sense and encoding events from news, but this process can be slow and uncertain. So in this paper we develop a near real time machine that detects the occurrence of an Atrocity event. We design a Spark Based Distribution Framework along with web scraper to generate and gather news. Raw text processing is performed more accurately using CoreNLP and the performance can be increased to a great extent with the help of SPARK.

# MOTIVATION & CHALLENGES

- Atrocities cause a lot of human suffering. Prediction of atrocities can save the lives of thousands of people. It is possible with the help of advanced technologies that uses machine learning and natural language processing techniques. Near real-time atrocity event coding is a tool that can predict and avoid such atrocities.

- Building a model is troublesome when the training data is very less
- Extracting real time data is difficult
- Trying to predict more from lesser training data is a tedious task

# Political Instability Task Force Worldwide Atrocities Dataset ANALYSIS:

| Victim Class | Victim Type | No of Documents | Percentage |
|---|---|---|---|
| Victimrand | Random | 1604 | 48.34 |
| Victimreli | Religious | 502 | 15.12 |
| Victimpoly | Political | 490 | 14.76 |
| Victimsoci | Social | 332 | 10.00 |
| Victimethn | Ethnic | 334 | 10.06 |
| Victimcomb | Combat | 56 | 1.68 |

# VICTIM CATEGORY DISTRIBUTION

| Category | Type | No of Documents | % |
|---|---|---|---|
| victimpoli | political | 360 | 19.4 |
| victimbomb | combat | 21 | 1.13 |
| victimreli | reigious | 215 | 11.60 |
| victimsoci | social | 195 | 10.51 |
| victimrand | random | 908 | 48.94 |
| victimethn | ethnic | 156 | 8.4 |
| | TOTAL | 1855 | 100.00 |

# TECHNICAL APPROACH

**Algorithm:**

**Procedure** : real-time news feeds classification

**Input :** human labelled data set

**Begin:**

data ← semafor ( inputdata )

Trainingdata ← add.Victimlabels(data)

Model ← logisticregression(trainingdata)

Realtimenewsfeed ← scrapper()

Testdata ← atrocityfilter(realtimenewsfeed)

Test ← semafor(testdata)

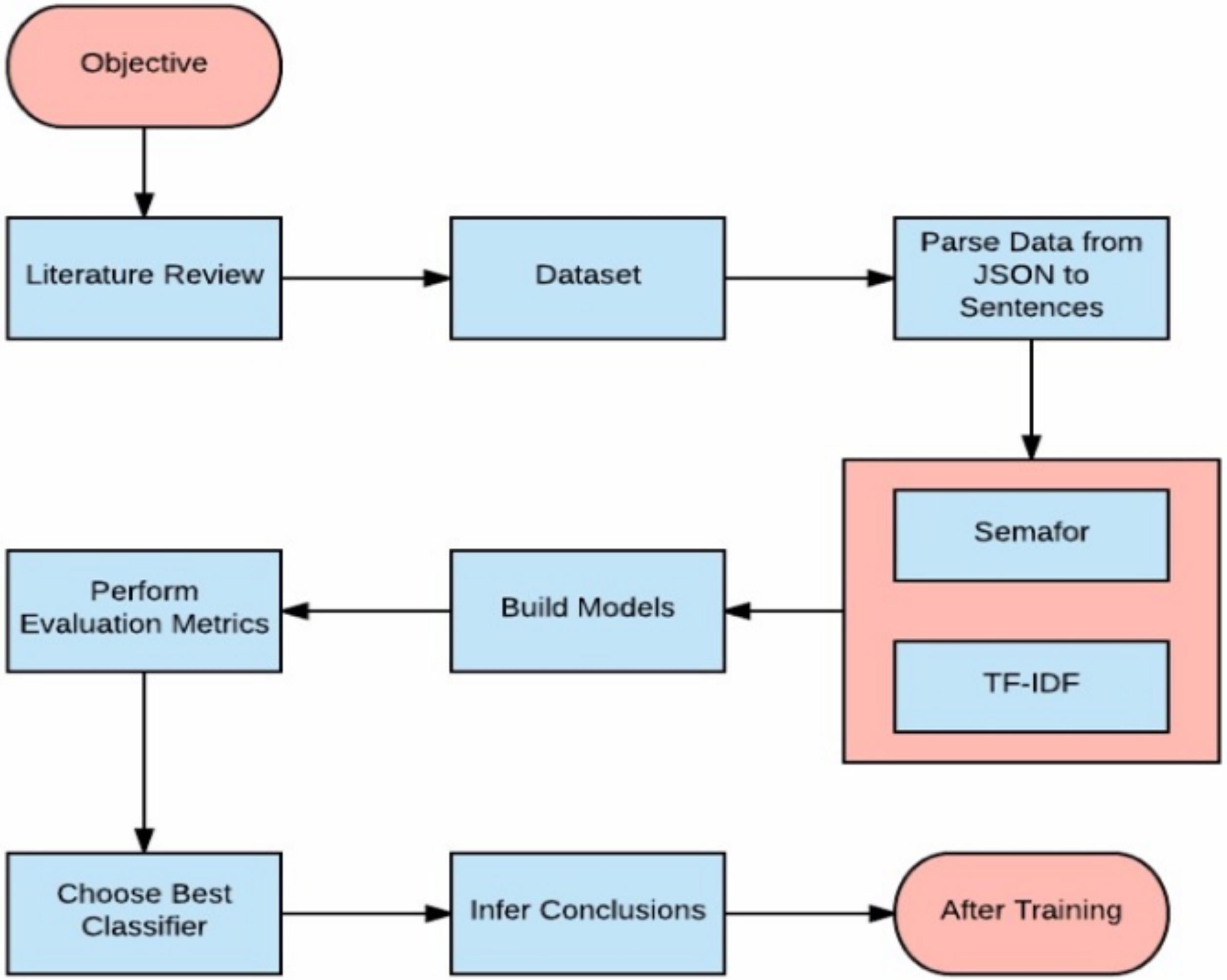
Prediction ← model.Predict(test)

Label ← addlabel(test)

**End:**

**End procedure**

# FLOWCHART FOR TRAINING



# SEMAFOR

- The stories from the Parus Atrocity Coding Dataset are loaded as JSON objects.
- The JSON objects are converted into sentence format which serves as the input for the SEMAFOR.
- SEMAFOR is a frame-semantic parser that parses English sentences automatically with respect to the semantic analysis in FrameNET.
- Frame semantics is a linguistic theory that has been instantiated for English in the FrameNet lexicon.
- This tool attempts to find which words in text evoke which semantic frames, and to find and label each frame's arguments.

# SEMAFOR WORK-FLOW:

- Preprocessing: The sentences are lemmatized, part-of-speech tagged, and syntactically parsed.
- Target identification Frame-evoking words and phrases ("targets") are heuristically identified in each sentence.
- Frame identification A log-linear model, trained on FrameNet 1.5 data with full-text frame annotations, produces for each target a probability distribution over frames in the FrameNet lexicon.
- The target is then labeled with the highest-scoring frame.
- Output An XML or JSON file is produced containing the text of the input sentences, augmented with the frame-semantic information

# SEMAFOR OUTPUT:

{"frames":[{"target":{"name":"Simple_name","spans":[{"start":4,"end":5,"text":"words"}]},"annotationSets":[{"rank":0,"score":10.350776176959009,"frameElements":[]}]},{"target":{"name":"Calendric_unit","spans":[{"start":6,"end":7,"text":"July"}]},"annotationSets":[{"rank":0,"score":22.97676138943529,"frameElements":[]}]},{"target":{"name":"Death","spans":[{"start":29,"end":30,"text":"death"}]},"annotationSets":[{"rank":0,"score":48.46822683190909,"frameElements":[]}]},{"target":{"name":"Make_noise","spans":[{"start":30,"end":31,"text":"toll"}]},"annotationSets":[{"rank":0,"score":68.23712131851352,"frameElements":[{"name":"Sound_source","spans":[{"start":31,"end":35,"text":"from Thursday 's killings"}]}]}]},{"target":{"name":"Calendric_unit","spans":[{"start":32,"end":33,"text":"Thursday"}]},"annotationSets":[{"rank":0,"score":27.943265603808264,"frameElements":[{"name":"Unit","spans":[{"start":32,"end":33,"text":"Thursday"}]}]}]},{"target":{"name":"Killing","spans":[{"start":34,"end":35,"text":"killings"}]},"annotationSets":[{"rank":0,"score":96.27641808457774,"frameElements":[{"name":"Victim","spans":[{"start":32,"end":34,"text":"Thursday 's"}]}]}]},{"target":{"name":"Natural_features","spans":[{"start":36,"end":37,"text":"Plateau"}]},"annotationSets":[{"rank":0,"score":37.31587596902473,"frameElements":[{"name":"Locale","spans":[{"start":36,"end":37,"text":"Plateau"}]}]}]},{"target":{"name":"Political_locales","spans":[{"start":37,"end":38,"text":"State"}]},"annotationSets":[{"rank":0,"score":34.77858747628415,"frameElements":[{"name":"Locale","spans":[{"start":37,"end":38,"text":"State"}]},{"name":"Name","spans":[{"start":36,"end":37,"text":"Plateau"}]}]}]},{"target":{"name":"Arriving","spans":[{"start":39,"end":40,"text":"reached"}]},"annotationSets":[{"rank":0,"score":90.13406103987668,"frameElements":[{"name":"Goal","spans":[{"start":40,"end":41,"text":"40"}]},{"name":"Theme","spans":[{"start":28,"end":38,"text":"The death toll from Thursday 's killings in Plateau State"}]}]}]},{"target":{"name":"Cardinal_numbers","spans":[{"start":40,"end":41,"text":"40"}]},"annotationSets":[{"rank":0,"score":26.131747637102997,"frameElements":[{"name":"Number","spans":[{"start":40,"end":41,"text":"40"}]}]}]}],"tokens":["by","Achor","Abimaje","327","words","1","July","2013","04",":","23","All","Africa","AFNWS","English","Jul","01",",","2013","-LRB-","Leadership/All","Africa","Global","Media","via","COMTEX","-RRB-","--","The","death","toll","from","Thursday","'s","killings","in","Plateau","State","has","reached","40"]}

# PARSED SEMAFOR OUTPUT:

[{'Victim': [u'12 people']}, {'Victim': [u'of her son Nzomo Mengi and his wife']}, {'Victim': [u'on the early hours of Thursday morning']}, {'Victim': [u'the other five members of his terror squad']}, {'Victim': [u'bows']}, {'Victim': [u'by the gangsters']}, {'Victim': [u'of such magnitude']}, {'Victim': [u'Mayor']}, {'Victim': [u'everybody including children during raids']}, {'Victim': [u'brother-in-law Musyoki , his wife and workers']}, {'Victim': [u'after their victims recognised them']}, {'Victim': [u'his son']}, {'Victim': [u'his']}, {'Victim': [u'him and his wife']}, {'Victim': [u'some']}, {'Victim': [u'him']}, {'Victim': [u'him']}, {'Victim': [u'me']}]

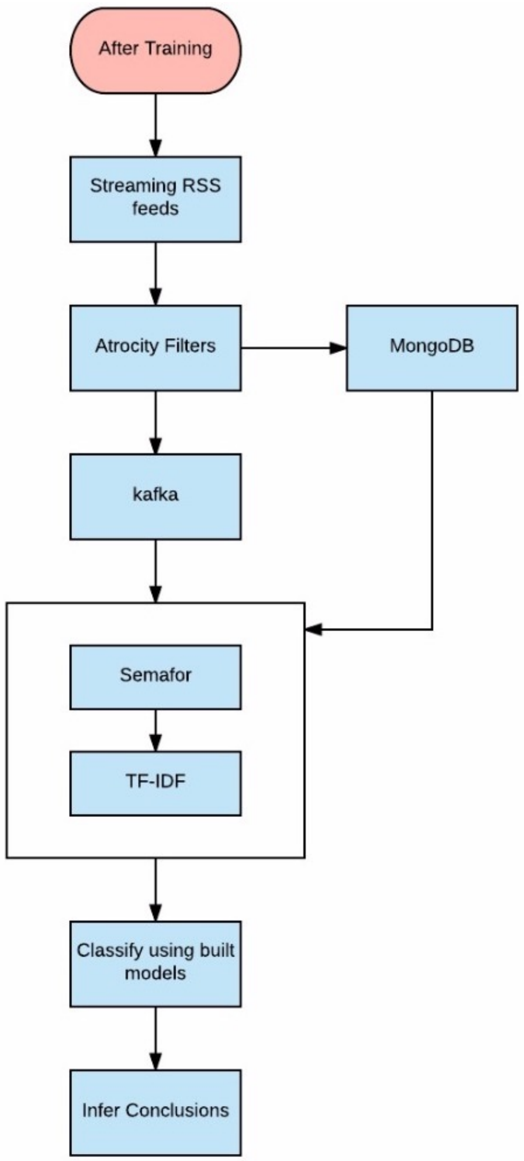
# TRAINING DATA:

```
AFNWS00020130704e974000xs::At least 10 people retaliation At least 10 people ::victimrand::5::people/civilians
AFNWS00020130706e976000e0::14 people 12 of its members ::victimpoli::0::pro-Morsi protestors
AFNWS00020130706e976000gp::at least 29 students and a teacher at a Nigerian boarding school some of the students schools ::victimrand::5::students
AFNWS00020130710e97a00119::12 people of her son Nzomo Mengi and his wife on the early hours of Thursday morning the other five members of his terror squad bows by the gangsters of such magnitude Mayor everybody including children during raids brother-in-law Musyoki , his wife and workers after their victims recognised them his son his him and his wife some him him me ::victimrand::5::people/civilians
AFNWS00020130719e97j000s0::Six people four people ::victimrand::5::villagers
AFNWS00020130807e9870004e::at Siney junction of Wardhigley district pistols of Tako `` heinous and terror act within two months ::victimpoli::0::deputy commissioner
AFNWS00020130813e98d000fk::at least 52 villagers on Sunday 40 people About 45 people nobody weapons chemical The leader of the Boko Haram sect weapons children the adults and young that refuse to accept Boko Haram. '' He swore that the group will continue to terrorise the country not only in the north or Borno State us ::victimrand::5::people
AFNWS00020130823e98n000l3::Thursday while five others sustained gunshot robbery the market Other persons at five ::victimrand::5::persons
AFNWS00020130825e98p00010::a village in Borno the village 35 people ::victimrand::5::revenge attack?
AFNWS00020130826e98q0002w::At least six people of a man near water borehole situating some 8 KMs north of Oog ::victimethn::1::fraternal clans in Og
AFNWS00020130827e98r000ur::about seven people more than 18 people ::victimrand::5::"persons"
AFNWS00020130831e98v00071::A pregnant woman and five men as the suspected armed robbers opened fire on their vehicle along Foron Road around 7.30 weapons ::victimrand::5::men and women
AFNWS00020130902e9920006m::-- At least seven people including five civilians during the battle ::victimrand::5::civilians
AFNWS00020130911e99b00075::No fewer than six persons ::victimrand::5::people
AFNWS00020130911e99b000l1::All the five villages in Effiat ::victimrand::5::people
AFNWS00020130912e99c00030::Six persons whose member Gambiri village Our men my men ::victimrand::5::persons
AFNWS00020130912e99c000hr::Six persons whose member Gambiri village Our men my men dupe suspecting job seekers of their money ::victimrand::5::persons
AFNWS00020130916e99g000hc::scores of security personnel of yesterday ::victimethn::1::Eggon and Alago clash
AFNWS00020130923e99n000vf::31 people 300 houses ::victimethn::1::Eggon and Alago clash
AFNWS00020130929e99t000gy::on Saturday the village ::victimrand::5::persons/civilians
AFNWS00020130930e99u0006f::they travellers one group guns around guns the attack , scores of travellers at least 20 persons to death of 38 students of the College of Agriculture , Gujba , YobeState others ::victimsoci::3::students
AFNWS00020131110e9ba00043::yesterday at Ikpele and Okpopolo communities in the state on our own side the number of deaths seven villages Agatu ::victimsoci::3::farmers
AFNWS00020131111e9bb000kh::-- At least 10 persons Six of them Four others Five Agatu villages 36 people ::victimsoci::3::farmers
AFNWS00020131120e9bk0006b::At least 20 people on the policemen 10 Al Shabaab attackers ::victimrand::5::civilians
AFNWS00020131122e9bm0003i::by arsonists who set his palace on fire to death the house ::victimpoli::0::traditional ruler
```

# TF- IDF TRAINING INSTANCES:

```
5.0 45:0.596766477591 62:1.05249395615 75:0.288613505481 92:0.289390873189 122:0.229185267099 135:0.161548927756 152:0.208265328377 165:0.44493173192 182:0.94225619319 225:0.779982760985 285:0.520865191863
0.0 32:0.338885510794 62:0.175415659359 75:0.288613505481 92:0.289390873189 122:0.158666723376 135:0.107699285171 152:0.208265328377 165:0.889863463841 182:0.134608027599 225:0.194995690246 272:0.568923812307 285:0.297637252493
5.0 2:1.67219081676 15:1.22953039475 32:0.677771021588 45:1.79029943277 62:1.57874093423 75:1.29876077467 92:0.868172619568 122:0.299703810821 135:0.376947498098 152:1.24959197026 165:0.22246586596 182:0.403824082796 225:0.584987070739 242:1.26193141913 255:1.03819026877 272:0.568923812307 285:0.595274504986
5.0 2:5.01657245028 15:4.50827811409 32:3.04996959715 45:3.43140724615 62:2.1049879123 75:3.1747485603 92:6.65599008335 105:1.37045185372 122:1.11066706363 135:1.18469213688 152:3.74877591078 165:2.89205625748 182:0.673040137993 195:2.63914765882 225:5.65487501714 242:7.25610565998 255:4.15276107508 272:5.68923812307 285:2.75314458556
5.0 2:0.41804770419 45:0.149191619398 92:0.289390873189 122:0.141037087445 135:0.161548927756 182:0.269216055197 225:0.194995690246 255:0.519095134385 272:0.568923812307 285:0.297637252493
0.0 2:1.25414311257 15:2.04921732459 45:1.04434133578 62:2.1049879123 75:0.721533763703 92:1.44695436595 122:0.317333446752 135:0.484646783269 152:1.45785729864 165:0.22246586596 182:0.269216055197 195:2.63914765882 212:1.57799997346 225:1.75496121222 242:1.57741427391 255:1.03819026877 272:1.13784762461 285:0.44645587874
5.0 2:2.50828622514 15:5.73780850885 32:1.01665653238 45:4.02817372374 62:4.20997582461 75:1.44306752741 92:4.34086309784 105:2.74090370744 122:1.18118560735 135:1.45394034981 152:3.54051058241 165:0.66739759788 182:1.74990435878 195:2.63914765882 212:6.31199989383 225:1.55996552197 242:3.78579425738 255:2.59547567192 272:1.13784762461 285:2.45550733307
5.0 2:0.836095408381 15:0.819686929835 32:1.35554204318 45:1.04434133578 62:1.22790961551 75:1.01014726918 92:2.02573611232 122:0.246814903029 135:0.215398570342 152:0.624795985131 165:0.22246586596 182:0.134608027599 195:1.31957382941 225:0.779982760985 242:2.20837998347 255:0.519095134385 272:1.13784762461 285:0.89291175748
5.0 32:0.677771021588 45:0.447574858193 62:0.175415659359 92:0.289390873189 105:1.37045185372 122:0.176296359307 135:0.215398570342 152:0.416530656754 182:0.673040137993 212:1.57799997346 225:0.584987070739 242:0.315482854782 255:1.03819026877 285:0.372046565617
1.0 2:0.41804770419 15:0.819686929835 32:0.338885510794 45:1.19353295518 62:1.22790961551 75:0.865840516444 92:1.15756349276 122:0.334963082682 135:0.430797140683 152:0.833061313508 165:0.44493173192 182:0.403824082796 225:0.779982760985 242:0.630965709564 255:1.03819026877 272:1.13784762461 285:0.595274504986
5.0 32:0.677771021588 45:0.447574858193 62:0.526246978076 75:0.144306752741 92:0.289390873189 122:0.193925995237 135:0.215398570342 152:0.416530656754 165:0.44493173192 182:0.269216055197 242:0.315482854782 285:0.520865191863
5.0 2:2.50828622514 15:1.63937385967 32:1.35554204318 45:1.64110781337 62:0.701662637434 75:0.865840516444 92:2.31512698551 122:0.458370534197 135:0.59234606844 152:2.49918394052 165:0.66739759788 182:0.403824082796 212:1.57799997346 225:0.974978451231 242:0.946448564346 255:1.03819026877 272:1.70677143692 285:1.11613969685
5.0 2:0.836095408381 15:0.819686929835 32:1.35554204318 45:0.745958096988 62:0.877078296793 75:0.432920258222 92:0.289390873189 122:0.229185267099 135:0.0538496425854 152:1.04132664188 182:0.673040137993 225:1.94995690246 242:0.315482854782 255:1.03819026877 272:0.568923812307 285:0.595274504986
5.0 2:0.41804770419 15:0.409843464918 32:0.338885510794 45:0.149191619398 62:0.175415659359 75:0.432920258222 92:0.578781746378 122:0.105777815584 135:0.107699285171 152:0.416530656754 225:0.194995690246 242:0.315482854782 272:0.568923812307 285:0.22322793937
5.0 32:0.677771021588 45:0.298383238795 62:0.350831318717 75:0.144306752741 122:0.105777815584 152:0.208265328377 165:0.22246586596 182:0.538432110394 225:0.974978451231 242:0.315482854782 255:0.519095134385 272:1.70677143692 285:0.22322793937
5.0 2:0.41804770419 15:0.819686929835 32:1.01665653238 45:0.447574858193 75:0.432920258222 92:1.15756349276 122:0.193925995237 135:0.161548927756 152:0.624795985131 165:1.33479519576 182:0.269216055197 225:0.779982760985 242:0.315482854782 255:1.03819026877 285:0.595274504986
```

# FLOWCHART FOR TESTING



# RSS FEEDS SAMPLE:


```
{
    "_id" : ObjectId("590e91aec5231b5e08af4443"),
    "content" : "The families of victims of San Bernardino terror shootout which took place in December 2015, have sued social media giants Facet
    "source" : "pakistan_thenews",
    "date" : "Fri, 05 May 2017 08:46:51 +0000",
    "language" : "english",
    "title" : "Facebook, Twitter, Google sued by San Bernardino victims' families",
    "url" : "http://www.newspakistan.pk/2017/05/05/facebook-twitter-google-sued-san-bernardino-victims-families/",
    "date_added" : ISODate("2017-05-07T03:17:02.933Z"),
    "stanford" : 0
}
```

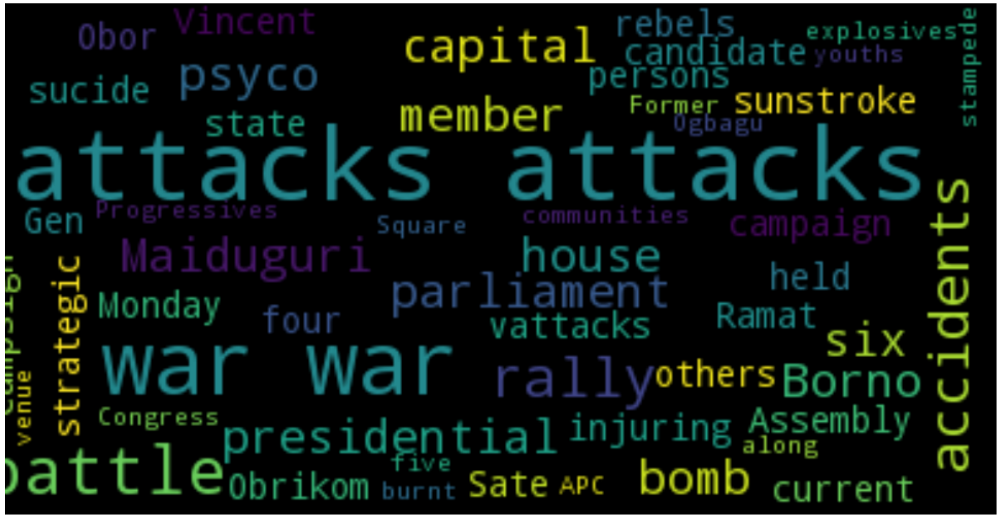
# PREDICTIONS FOR TEST DATA

```
>>> predictions.collect()
[Row(label=0.0, features=SparseVector(285, {}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 2.0902
, 14: 1.2295, 31: 0.6778, 44: 1.9395, 61: 0.3508, 74: 0.5772, 91: 1.447, 104: 1.3705, 121: 0.2292, 134: 0.3231, 151: 1.2496, 164: 0.4449, 181: 0.1346, 194: 2.6391, 211: 3.156, 224:
1.17, 241: 0.631, 254: 0.5191, 271: 1.7068, 284: 0.2232}), prediction=3.0), Row(label=0.0, features=SparseVector(285, {1: 3.7624, 14: 4.9181, 31: 3.05, 44: 3.2822, 61: 2.8067, 74: 2
.7418, 91: 2.3151, 104: 1.3705, 121: 0.8286, 134: 0.6462, 151: 2.9157, 164: 1.5573, 181: 0.9423, 194: 1.3196, 211: 3.156, 224: 5.6549, 241: 2.8393, 254: 4.6719, 271: 4.5514, 284: 1.
3394}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 0.8361, 14: 2.0492, 31: 1.3555, 44: 0.4476, 61: 0.8771, 74: 1.0101, 91: 1.1576, 104: 1.3705, 121: 0.3173, 134:
 0.2692, 151: 0.4165, 164: 1.1123, 181: 0.5384, 194: 1.3196, 224: 1.17, 241: 1.5774, 254: 0.5191, 284: 0.5209}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 1.672
2, 14: 1.2295, 31: 0.3389, 44: 1.7903, 61: 1.0525, 74: 1.1545, 91: 1.447, 104: 2.7409, 121: 0.5994, 134: 0.5923, 151: 1.2496, 164: 1.7797, 181: 1.6153, 224: 2.145, 241: 1.8929, 254:
 0.5191, 271: 0.5689, 284: 0.9673}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 0.418, 14: 0.4098, 31: 0.3389, 44: 0.5968, 61: 0.3508, 74: 0.1443, 91: 0.8682, 10
4: 1.3705, 121: 0.2116, 134: 0.1077, 151: 0.4165, 181: 0.673, 211: 3.156, 224: 0.975, 241: 0.3155, 271: 2.2757, 284: 0.5209}), prediction=5.0), Row(label=0.0, features=SparseVector(
285, {1: 1.6722, 14: 2.4591, 31: 0.3389, 44: 2.8346, 61: 1.0525, 74: 1.1545, 91: 2.0257, 121: 0.4055, 134: 0.3769, 151: 2.7074, 164: 0.6674, 181: 0.8076, 194: 1.3196, 224: 2.145, 24
1: 0.3155, 254: 1.5573, 271: 0.5689, 284: 0.8185}), prediction=0.0), Row(label=0.0, features=SparseVector(285, {1: 2.0902, 14: 0.8197, 31: 1.0167, 44: 2.0887, 61: 0.1754, 74: 0.5772
, 91: 0.5788, 121: 0.2292, 134: 0.3231, 151: 0.6248, 164: 0.6674, 181: 0.4038, 224: 0.975, 241: 0.9464, 254: 0.5191, 271: 1.1378, 284: 0.2232}), prediction=5.0), Row(label=0.0, feat
ures=SparseVector(285, {1: 0.418, 14: 0.4098, 31: 0.3389, 44: 0.746, 61: 0.7017, 74: 0.4329, 91: 0.8682, 121: 0.1939, 134: 0.1615, 151: 0.6248, 164: 0.2225, 181: 0.4038, 224: 0.39,
241: 0.3155, 254: 0.5191, 284: 0.4465}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 1.6722, 14: 1.6394, 31: 1.3555, 44: 0.4476, 61: 0.5262, 74: 1.0101, 91: 1.447
, 121: 0.2997, 134: 0.4308, 151: 0.6248, 164: 0.6674, 181: 0.5384, 224: 0.975, 241: 0.631, 271: 1.7068, 284: 1.0417}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1:
 0.418, 14: 0.4098, 31: 0.3389, 44: 0.5968, 61: 0.3508, 74: 0.1443, 91: 0.2894, 121: 0.1939, 134: 0.2154, 151: 1.0413, 164: 0.4449, 181: 0.2692, 224: 0.39, 271: 0.5689, 284: 0.4465}
), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 2.5083, 14: 0.8197, 31: 0.3389, 44: 1.0443, 61: 1.0525, 74: 0.2886, 91: 1.447, 121: 0.3173, 134: 0.2692, 151: 0.624
8, 164: 0.6674, 194: 1.3196, 224: 0.39, 241: 1.2619, 254: 0.5191, 271: 1.1378, 284: 1.0417}), prediction=0.0), Row(label=0.0, features=SparseVector(285, {1: 0.418, 14: 0.8197, 31: 0
.6778, 44: 1.1935, 61: 0.3508, 74: 0.4329, 91: 1.1576, 121: 0.141, 134: 0.0538, 151: 0.4165, 164: 0.4449, 211: 3.156, 224: 0.78, 241: 0.631, 254: 1.0382, 284: 0.1488}), prediction=5
.0), Row(label=0.0, features=SparseVector(285, {1: 0.418, 14: 1.2295, 31: 1.0167, 44: 0.5968, 61: 0.7017, 74: 0.4329, 91: 0.2894, 121: 0.1939, 134: 0.2692, 151: 0.8331, 164: 1.3348,
 211: 1.578, 224: 0.975, 241: 0.9464, 271: 1.1378, 284: 0.372}), prediction=2.0), Row(label=0.0, features=SparseVector(285, {1: 2.0902, 14: 0.8197, 31: 0.6778, 44: 1.7903, 74: 0.288
6, 91: 1.447, 121: 0.2821, 134: 0.1077, 151: 0.6248, 164: 0.4449, 181: 0.9423, 224: 1.365, 241: 1.5774, 254: 1.0382, 284: 0.6697}), prediction=5.0), Row(label=0.0, features=SparseVe
ctor(285, {1: 0.418, 14: 0.8197, 44: 0.5968, 61: 0.3508, 74: 0.7215, 91: 1.447, 121: 0.1763, 134: 0.2692, 151: 0.4165, 181: 0.5384, 224: 1.17, 241: 0.631, 271: 0.5689, 284: 0.2232})
, prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 0.418, 14: 0.4098, 44: 0.746, 61: 0.3508, 74: 0.1443, 91: 0.5788, 121: 0.141, 134: 0.2154, 151: 0.8331, 181: 0.1346,
 254: 0.5191, 284: 0.4465}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 0.418, 14: 0.4098, 44: 0.4476, 61: 0.5262, 74: 0.2886, 121: 0.1234, 134: 0.0538, 181: 0.2
692, 224: 0.585, 241: 0.3155, 254: 0.5191, 284: 0.372}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 0.418, 14: 0.4098, 44: 0.4476, 61: 0.5262, 91: 0.5788, 121: 0
.1058, 134: 0.1077, 151: 0.6248, 164: 0.2225, 181: 0.1346, 224: 0.78, 241: 0.631, 271: 0.5689, 284: 0.2232}), prediction=5.0), Row(label=0.0, features=SparseVector(285, {1: 0.418, 4
4: 0.1492, 61: 0.3508, 74: 0.1443, 91: 0.5788, 121: 0.0529, 134: 0.0538, 151: 0.4165, 164: 0.2225, 211: 1.578, 224: 0.195, 241: 0.3155, 284: 0.1488}), prediction=5.0), Row(label=0.0
```

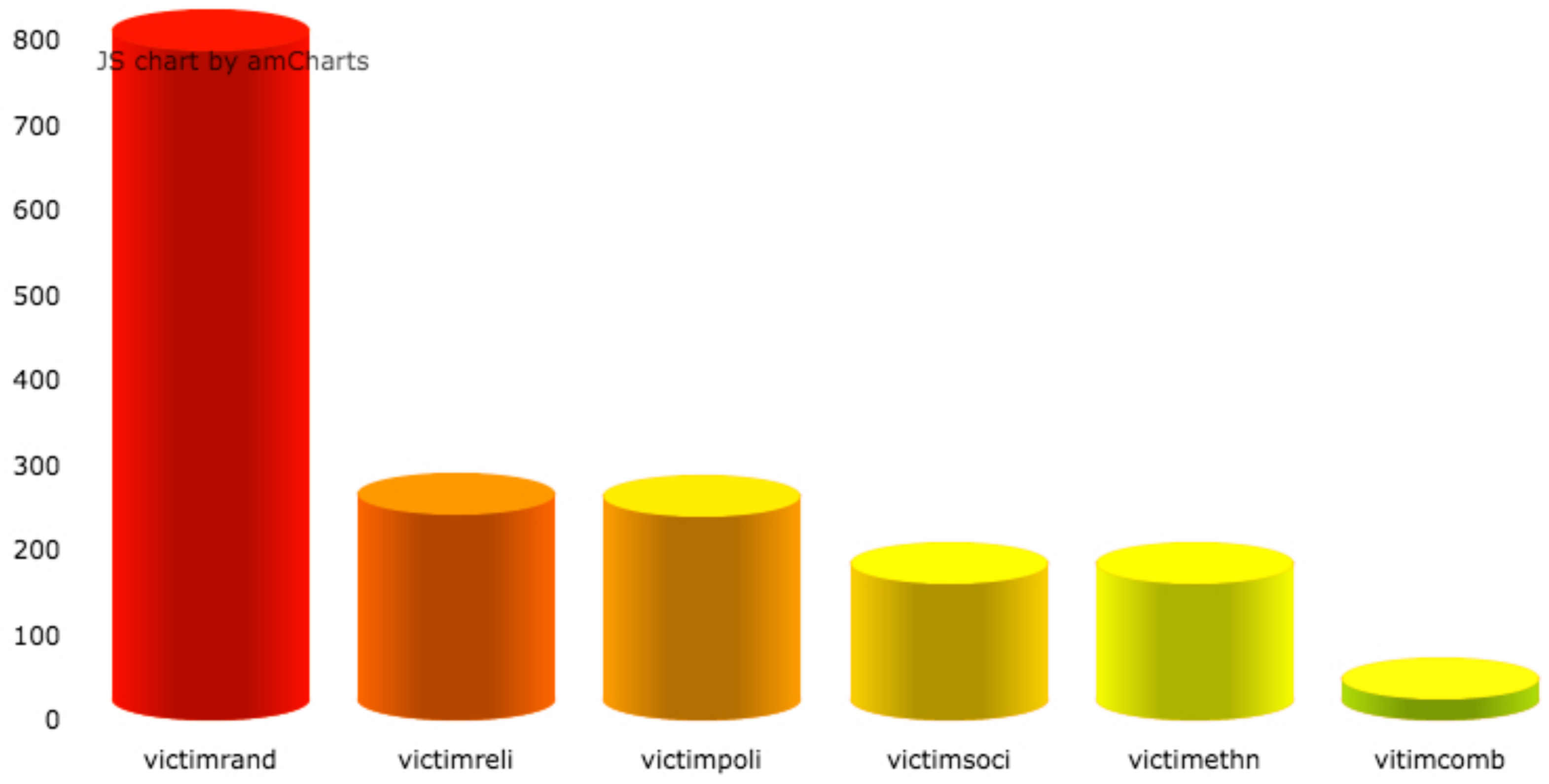
# LABELS PREDICTION

VICTIM POLITICAL


VICTIM COMBAT



# NEWS FEED VS VICTIM CLASS



# METRICS EVALUATION TABLE

| Classifier | Recall | Precision | F1 Score | Accuracy |
|---|---|---|---|---|
| Multinomial Naïve Bayes | 0.527 | 0.517 | 0.511 | 0.528 |
| Logistic Regression | 0.51 | 0.512 | 0.512 | 0.518 |

# OBSERVATION:

In our project, Multinomial Naïve Bayes outperforms Logistic Regression.

# CONCLUSION:

- In the Project we have successfully implemented the paper **Near Real-Time Atrocity Event Coding.** We have introduced a spark-based model and a machine learning classifier to predict the atrocity events in near real time and also predicted the label of the predicted class of the victim.

# FUTURE WORK:

- We can gain more accuracy by implementing deep learning techniques instead of logistic regression, a simple linear based classifier.
- As the data is imbalanced, the performance can be improved by increasing the quality of the. Data by oversampling and under sampling.

# THANK YOU