

“Project Title: BestBuddy”

Third Increment Report:

Team Members:

Sidrah Junaid.

Yeruva, Vijaya Kumari

Saria Goudarzvand

Project Overview:

The purpose of the project is to create an application that predict the user as happy or unhappy based on his/her Twitter activities. For completing the task Twitter API is used. In our project we implemented Deep Learning algorithm using Tensorflow .

Significance:

The significance of the application is that the user can identify his/her lifestyle whether it is lies in happy/unhappy and on the basis of the category suggested by the application he/she can change his/her lifestyle. It also helps user to find the people having comparable interest and choices.

Existing Services/REST API:

Twitter APIs is used to collect tweets to implement increment3

Implementation:

For increment 3, we trained our simple CNN model and CNN with GLOVE using Tensor flow on the sufficient data we gathered for training to compare the accuracy of the mentioned models and applied categorization as happy and unhappy. We also implemented the Web UI through which the application can fetch the tweets of user and we saved it in TwitterStream Database in MongoDB in JSON format. The data is classified as happy/unhappy by using trained model on twitter data.

Source Code: <https://github.com/VIJAYAYERUVA/BestBuddy/tree/master/Sourcecode/BestBuddy>

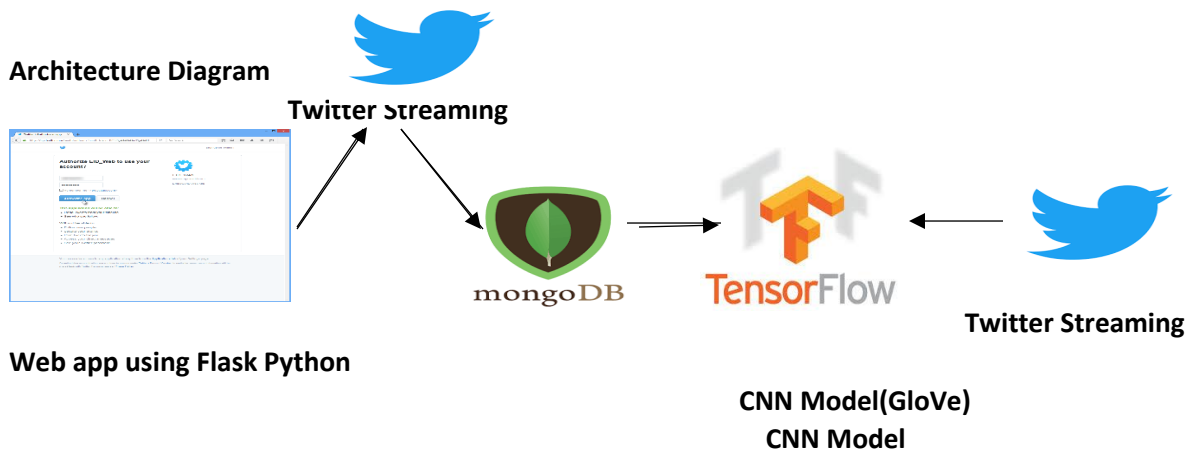
Project Management:

Work completed:

1. We have created the training data using Twitter tweets using CNN and CNN with GLOVE using Tensorflow
2. Implemented user interface of our application using Python and Flask
3. Save the Tweets in Mongo DB

"Project Title: BestBuddy"

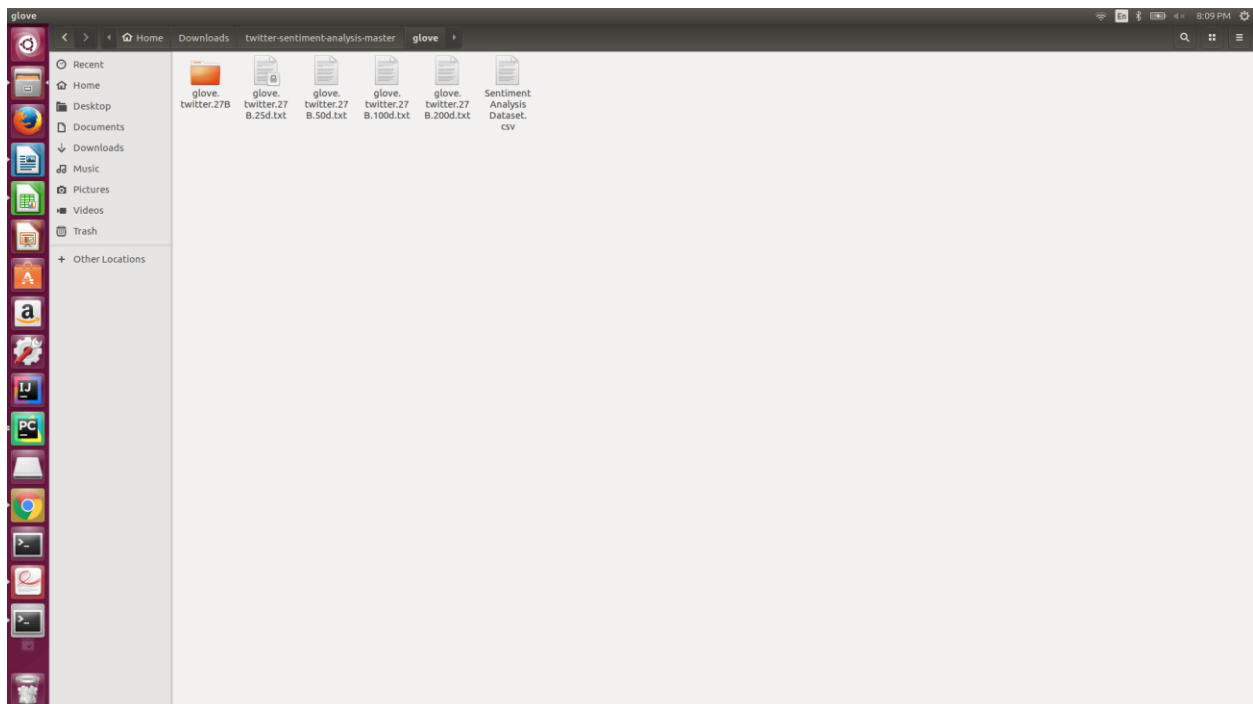
Architecture Diagram



Details of Increment 3:

CNN Model with GLOVE using Tensorflow

CNN model is used for performing visual recognition tasks and GloVe is an unsupervised learning algorithm for obtaining vector representation of words.



The accuracy of above mentioned model is about 53% using 30000 of twitter data, by providing more than 30000 of tweets its accuracy is 79%

“Project Title: BestBuddy”

CNN Model using Tensorflow

Here we provide two categories of tweets in negative and positive category and trained our CNN model.

The screenshot shows the PyCharm IDE interface. The top pane displays the project structure for 'cnn-text-classification-tf-master', with the 'rt-polaritydata' folder expanded to show 'rt-polarity.pos'. The bottom pane shows the 'train.py' script being executed. The script's parameters are listed below the execution command:

```
Parameters:
ALLOW_SOFT_PLACEMENT=True
BATCH_SIZE=64
CHECKPOINT_EVERY=100
DEV_SAMPLE_PERCENTAGE=0.1
DROPOUT_KEEP_PROB=0.5
EMBEDDING_DIM=128
EVALUATE_EVERY=100
FILTER_SIZES=3,4,5
L2_REG_LAMBDA=0.0
LOG_DEVICE_PLACEMENT=False
NEGATIVE_DATA_FILE=./data/rt-polaritydata/rt-polarity.neg
NUM_CHECKPOINTS=5
NUM_EPOCHS=200
NUM_FILTERS=128
POSITIVE_DATA_FILE=./data/rt-polaritydata/rt-polarity.pos
```

The output of the script shows the loading of data and a warning message from TensorFlow:

```
Loading data...
2017-04-24 20:08:16.927859: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use SSE4.1 instructions, but these are available on your machine
and could speed up CPU computations.
2017-04-24 20:08:16.927874: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use SSE4.2 instructions, but these are available on your machine
and could speed up CPU computations.
```

The screenshot shows the PyCharm IDE interface. The top pane displays the project structure for 'cnn-text-classification-tf-master', with the 'rt-polaritydata' folder expanded to show 'rt-polarity.neg'. The bottom pane shows the 'train.py' script being executed. The script's parameters are listed below the execution command:

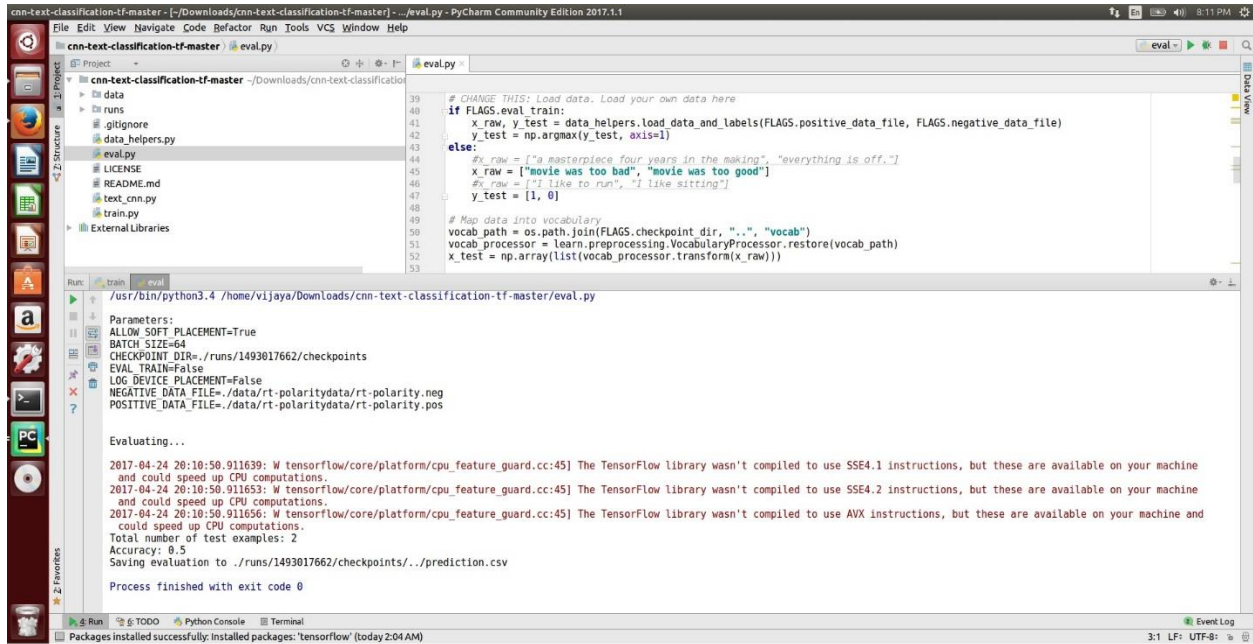
```
Parameters:
ALLOW_SOFT_PLACEMENT=True
BATCH_SIZE=64
CHECKPOINT_EVERY=100
DEV_SAMPLE_PERCENTAGE=0.1
DROPOUT_KEEP_PROB=0.5
EMBEDDING_DIM=128
EVALUATE_EVERY=100
FILTER_SIZES=3,4,5
L2_REG_LAMBDA=0.0
LOG_DEVICE_PLACEMENT=False
NEGATIVE_DATA_FILE=./data/rt-polaritydata/rt-polarity.neg
NUM_CHECKPOINTS=5
NUM_EPOCHS=200
NUM_FILTERS=128
POSITIVE_DATA_FILE=./data/rt-polaritydata/rt-polarity.pos
```

The output of the script shows the loading of data and a warning message from TensorFlow:

```
Loading data...
2017-04-24 20:08:16.927859: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use SSE4.1 instructions, but these are available on your machine
and could speed up CPU computations.
2017-04-24 20:08:16.927874: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use SSE4.2 instructions, but these are available on your machine
and could speed up CPU computations.
```

“Project Title: BestBuddy”

The accuracy of the CNN model on tweets data is 50%



```
cnntext-classification-tf-master - [~/Downloads/cnntext-classification-tf-master] - ./eval.py - PyCharm Community Edition 2017.1.1
File Edit View Navigate Code Refactor Run Tools VCS Window Help
cnntext-classification-tf-master - [~/Downloads/cnntext-classification-tf-master] - eval.py
Project: cnntext-classification-tf-master
  data
  runs
  .gitignore
  data_helpers.py
  eval.py
  LICENSE
  README.md
  text_cnn.py
  train.py
  External Libraries

Run: /usr/bin/python3.4 /home/vijaya/Downloads/cnntext-classification-tf-master/eval.py

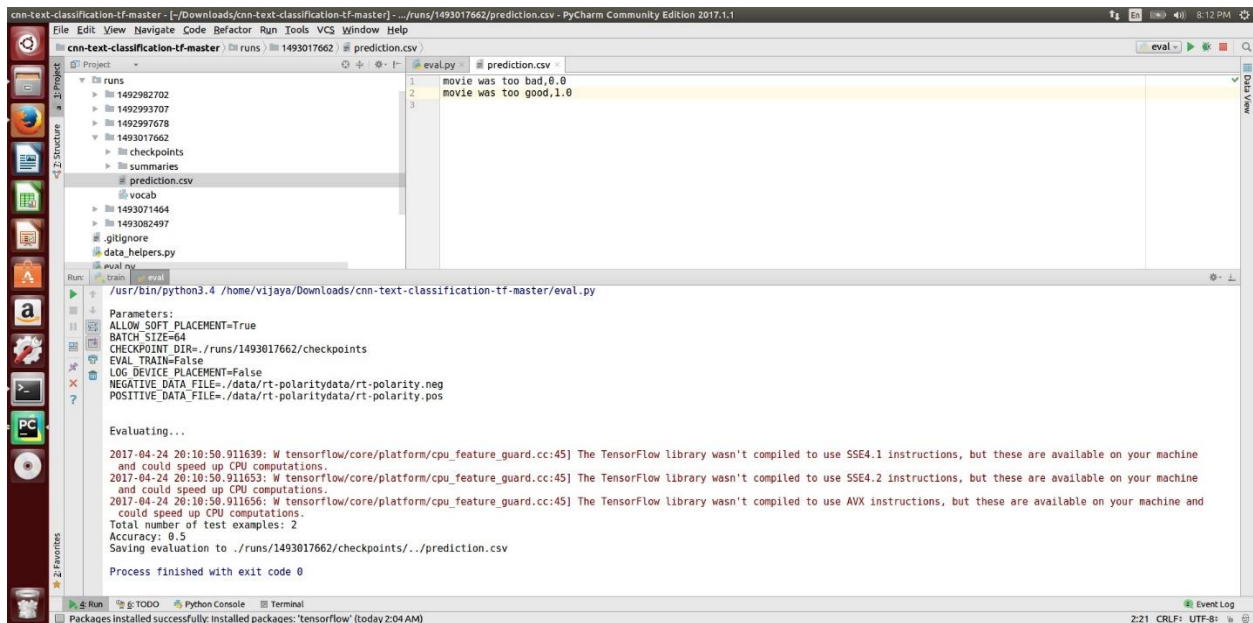
Parameters:
ALLOW_SOFT_PLACEMENT=True
BATCH_SIZE=64
CHECKPOINT_DIR=./runs/1493017662/checkpoints
EVAL_TRAIN=False
LOG_DEVICE_PLACEMENT=False
NEGATIVE_DATA_FILE=./data/rt-polaritydata/rt-polarity.neg
POSITIVE_DATA_FILE=./data/rt-polaritydata/rt-polarity.pos

Evaluating...
2017-04-24 20:10:50.911639: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use SSE4.1 instructions, but these are available on your machine and could speed up CPU computations.
2017-04-24 20:10:50.911653: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use SSE4.2 instructions, but these are available on your machine and could speed up CPU computations.
2017-04-24 20:10:50.911656: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use AVX instructions, but these are available on your machine and could speed up CPU computations.
Total number of test examples: 2
Accuracy: 0.5
Saving evaluation to ./runs/1493017662/checkpoints/./prediction.csv

Process finished with exit code 0

Packages installed successfully: installed packages: 'tensorflow' (today 2:04 AM)
```

The results will be saved in csv file.



```
cnntext-classification-tf-master - [~/Downloads/cnntext-classification-tf-master] - ./runs/1493017662/prediction.csv - PyCharm Community Edition 2017.1.1
File Edit View Navigate Code Refactor Run Tools VCS Window Help
cnntext-classification-tf-master - [~/Downloads/cnntext-classification-tf-master] - runs - 1493017662 - prediction.csv
Project: cnntext-classification-tf-master
  runs
  1492992702
  1492993707
  1492997678
  1493017662
  checkpoints
  summaries
  prediction.csv
  vocab
  1493071464
  1493082497
  .gitignore
  data_helpers.py
  word.py

Run: /usr/bin/python3.4 /home/vijaya/Downloads/cnntext-classification-tf-master/eval.py

Parameters:
ALLOW_SOFT_PLACEMENT=True
BATCH_SIZE=64
CHECKPOINT_DIR=./runs/1493017662/checkpoints
EVAL_TRAIN=False
LOG_DEVICE_PLACEMENT=False
NEGATIVE_DATA_FILE=./data/rt-polaritydata/rt-polarity.neg
POSITIVE_DATA_FILE=./data/rt-polaritydata/rt-polarity.pos

Evaluating...
2017-04-24 20:10:50.911639: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use SSE4.1 instructions, but these are available on your machine and could speed up CPU computations.
2017-04-24 20:10:50.911653: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use SSE4.2 instructions, but these are available on your machine and could speed up CPU computations.
2017-04-24 20:10:50.911656: W tensorflow/core/platform/cpu_feature_guard.cc:45] The TensorFlow library wasn't compiled to use AVX instructions, but these are available on your machine and could speed up CPU computations.
Total number of test examples: 2
Accuracy: 0.5
Saving evaluation to ./runs/1493017662/checkpoints/./prediction.csv

Process finished with exit code 0

Packages installed successfully: installed packages: 'tensorflow' (today 2:04 AM)
```

“Project Title: BestBuddy”

Web Application:

The UI of application is created using Flask a light weight Python Framework for building Webapplications.



Flask Login App

Sign in with twitter.



Database:

The database is created using MongoDB (noSQL) database.

```
> use TwitterStream
switched to db TwitterStream
> show dbs
TwitterStream  0.000GB
admin          0.000GB
local          0.000GB
tweet_db       0.000GB
```

Tweet collected in **tweets** collection

```
> show collections
tweets
> show TwitterStream
2017-04-24T18:43:10.060-0500 E QUERY [thread1] Error: don't know how to show [TwitterStream] :
shellHelper.show@src/mongo/shell/utils.js:906:11
shellHelper@src/mongo/shell/utils.js:659:15
@shellHelper2:1:1
```

Datatin tweets collection in JSON format

“Project Title: BestBuddy”

```
> db.tweets.find()
{ "_id" : ObjectId("58fe8b20b75cbb2a6418ff96"), "username" : "stephanie_dress", "language" : "en", "created" : ISODate("2017-04-24T23:32:48Z"), "text" : "Parents Confis
cated his iPad Then He kills Himself Because Of It! https://t.co/Egok449fcv", "hashtags" : [ ], "followers" : 1723, "id" : "856652174301904896",
{ "_id" : ObjectId("58fe8b20b75cbb2a6418ff97"), "username" : "kaitlynjoy06", "language" : "en", "created" : ISODate("2017-04-24T23:32:48Z"), "text" : "RT @ricklondon: F
luffy Cohen #Attorney @ Claw by @LTCartoons #cats #mice #lawyers #humor #cartoon #comic #funny #LTCartoons https://t.co/dy", "hashtags" : [ { "indices" : [ 29, 38 ], "text" : "Attorney" }, { "indices" : [ 61, 66 ], "text" : "cats" }, { "indices" : [ 67, 72 ], "text" : "mice" }, { "indices" : [ 73, 81 ], "text" : "lawyers" }, { "indices" : [ 82, 88 ], "text" : "humor" }, { "indices" : [ 89, 97 ], "text" : "cartoon" }, { "indices" : [ 98, 104 ], "text" : "comic" }, { "indices" : [ 105, 111 ], "text" : "funny" }, { "indices" : [ 112, 122 ], "text" : "LTCartoons" }, { "followers" : 31, "id" : "856652173291261952" } },
{ "_id" : ObjectId("58fe8b20b75cbb2a6418ff98"), "username" : "johnston_am", "language" : "en", "created" : ISODate("2017-04-24T23:32:49Z"), "text" : "RT @DrTonyWagner:
Powerful essay on developing kids' \"muscle\" of resilience and ensuring that all kids feel as though they matter. https://", "hashtags" : [ ], "followers" : 159, "id" : "856652175312900096" },
{ "_id" : ObjectId("58fe8b21b75cbb2a6418ff99"), "username" : "fahma311", "language" : "en", "created" : ISODate("2017-04-24T23:32:49Z"), "text" : "@JordanYorkMusic A we
ek but can't take my iPad away. #0000000", "hashtags" : [ ], "followers" : 23217, "id" : "856652178395627520" },
{ "_id" : ObjectId("58fe8b23b75cbb2a6418ff9a"), "username" : "juntomazine", "language" : "en", "created" : ISODate("2017-04-24T23:32:52Z"), "text" : "With these #ears
no one will #suspect a thing. #funny #disguise #cat #rabbit #photography https://t.co/tgicnzVasZ", "hashtags" : [ { "indices" : [ 11, 16 ], "text" : "ears" }, { "indices" : [ 29, 37 ], "text" : "suspect" }, { "indices" : [ 47, 53 ], "text" : "funny" }, { "indices" : [ 54, 63 ], "text" : "disguise" }, { "indices" : [ 64, 68 ], "text" : "cat" }, { "indices" : [ 69, 76 ], "text" : "rabbit" }, { "indices" : [ 77, 89 ], "text" : "photography" }, { "followers" : 665, "id" : "856652188441071616" } },
{ "_id" : ObjectId("58fe8b2db75cbb2a6418ff9b"), "username" : "missbumptious", "language" : "en", "created" : ISODate("2017-04-24T23:33:02Z"), "text" : "RT @nationallamp
oon: Don't put your #relationship in a box. Today's #cartoon by @CrowdenSatz https://t.co/6ad5gqSU0p #dating #funny #art ht", "hashtags" : [ { "indices" : [ 36, 49 ], "text" : "relationship" }, { "indices" : [ 68, 76 ], "text" : "cartoon" }, { "indices" : [ 117, 124 ], "text" : "dating" }, { "indices" : [ 125, 131 ], "text" : "funny" }, { "indices" : [ 132, 136 ], "text" : "art" }, { "followers" : 2281, "id" : "856652230342082560" } },
{ "_id" : ObjectId("58fe8b2db75cbb2a6418ff9c"), "username" : "Spiryal13Upkat", "language" : "en", "created" : ISODate("2017-04-24T23:33:02Z"), "text" : "Oh I definitely
will... I see the cover with the baby sleeping on the iPad thing. Thanks https://t.co/40GzPUEf4", "hashtags" : [ ], "followers" : 476, "id" : "856652230274957313" },
{ "_id" : ObjectId("58fe8b2fb75cbb2a6418ff9d"), "username" : "komikten", "language" : "en", "created" : ISODate("2017-04-24T23:33:03Z"), "text" : "When your girl pulls
up wit her friends to argue this how they be in the background https://t.co/m873wi0a06 #fun #funny #lolz #lol #comedy", "hashtags" : [ { "indices" : [ 108, 112 ], "text" : "fun" }, { "indices" : [ 113, 119 ], "text" : "funny" }, { "indices" : [ 120, 125 ], "text" : "lolz" }, { "indices" : [ 126, 130 ], "text" : "lol" }, { "indices" : [ 131, 138 ], "text" : "comedy" }, { "followers" : 1285, "id" : "856652237489266689" } },
{ "_id" : ObjectId("58fe8b2fb75cbb2a6418ff9e"), "username" : "DavidWmag", "language" : "en", "created" : ISODate("2017-04-24T23:33:04Z"), "text" : "RT @nationallampoon
: Don't put your #relationship in a box. today's #cartoon by @CrowdenSatz https://t.co/6ad5gqSU0p #dating #funny #art ht", "hashtags" : [ { "indices" : [ 36, 49 ], "text" : "relationship" }, { "indices" : [ 68, 76 ], "text" : "cartoon" }, { "indices" : [ 117, 124 ], "text" : "dating" }, { "indices" : [ 125, 131 ], "text" : "funny" }, { "indices" : [ 132, 136 ], "text" : "art" }, { "followers" : 44, "id" : "856652237954830336" } },
{ "_id" : ObjectId("58fe8b31b75cbb2a6418ff9f"), "username" : "brinky53", "language" : "en", "created" : ISODate("2017-04-24T23:33:05Z"), "text" : "Apple iPad (2017) rev
iew: the best feature is the price - https://t.co/yUs2ci7oAq https://t.co/p4awgK7nj", "hashtags" : [ ], "followers" : 37, "id" : "856652245638799360" },
{ "_id" : ObjectId("58fe8b35b75cbb2a6418ffa0"), "username" : "vmontielarquero", "language" : "en", "created" : ISODate("2017-04-24T23:33:09Z"), "text" : "RT @L0LB00Kcom
: \"Could You Help Me Out Here For A Second?\" #funny #smart #animals #monkey Get more #videos at - \"https://t.co/6ad5gqSU0p #dating #funny #art ht\", "hashtags" : [ { "indices" : [ 59, 65 ], "text" : "funny" }, { "indices" : [ 66, 72 ], "text" : "smart" }, { "indices" : [ 73, 81 ], "text" : "animals" }, { "indices" : [ 82, 89 ], "text" : "monkey" }, { "indices" : [ 99, 106 ], "text" : "videos" }, { "followers" : 159, "id" : "856652245638799360" } }
```

Project Responsibilities:

- 1.CNN model with GloVe using Tensorflow Saria,Vijaya,Sidrah
- 2.CNN model using Tensorflow Saria,Vijaya,Sidrah
- 3.MongoDB creation Sidrah,Vijaya(Activities),Saria
- 4.Webapplication using Flask Sidrah,Vijaya(Activities),Saria

Bibliography:

<https://www.youtube.com/watch?v=nzkrRQGCeME>

<https://www.android.com/>

<https://dev.twitter.com/rest/collections>

<http://flask.pocoo.org/docs/0.12/>

<https://docs.mongodb.com/getting-started/shell/query/>

<https://nlp.stanford.edu/projects/glove/>

<https://github.com/keon/deep-api>