



---

# Twitter Hate-Speech Detection

Muhammad Saleh Ahsan Sakir  
Muhammad Uzair Umar

---

# Overview

- Introduction
  - Data collection and description
  - Data preprocessing
  - Feature Engineering
  - Baseline method and Evaluation
  - Main methods and evaluation
  - Challenges
-

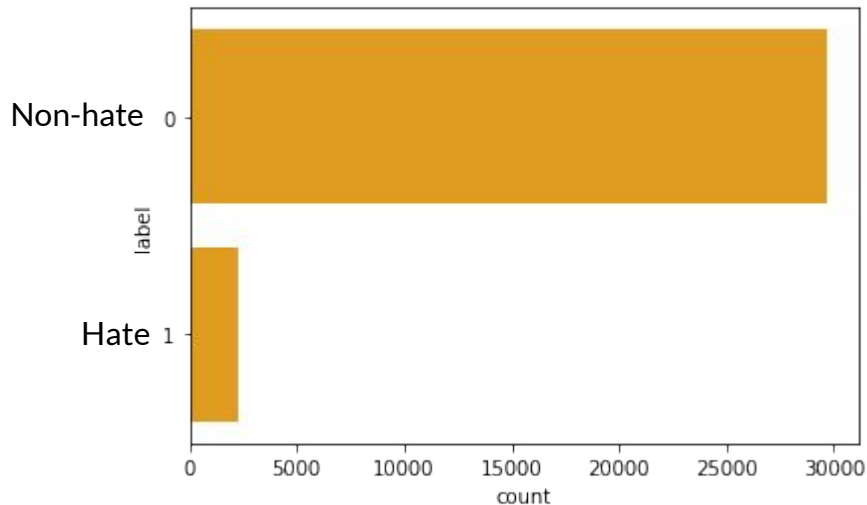
---

# Data Collection and description

## Collection

- 32000 Tweets (Source:Kaggle)
- 25569 Train Tweets (80%)
- 6362 Test Tweets (20%)

## Description



---

# Data pre-processing

## Tweet-preprocessor

- URLs
- User mentions (@)
- Emojis
- Numbers
- Reserved Words (RT, FAVs)

- Stop Word (nltk)
- Hash (#) Symbol

tweet	cleaned
@user when a father is dysfunctional and is s...	when a father is dysfunctional and is so selfi...
@user @user thanks for #lyft credit i can't us...	thanks for lyft credit i can't use cause they ...
bihday your majesty	bihday your majesty
#model i love u take with u all the time in ...	model i love u take with u all the time in ur!!

---

# Feature Engineering

---

# Feature Engineering


- CountVectorizer
- TfidfVectorizer

Missing:  
Cosine  
similarity

label	tweet	cleaned	word_count	unique_words_count	stopwords_count	punctuation_count	polarity	keywords	POS_tagged
0	@user when a father is dysfunctional and is s...	father dysfunctional selfish drags kids dysfun...	17	15	10	1	-0.500000	father dysfunctional selfish drags kids dysfun...	RBR JJ JJ NNS NNS NN . VB
0	@user @user thanks for #lyft credit i can't us...	thanks lyft credit can't use cause offer wheel...	17	17	5	2	0.200000	use cause offer wheelchair vans pdx thanks lyf...	NNS VBP NN MD RB VB NN NN NN NNS VBP . VBN VBD
0	bihday your majesty	bihday majesty	3	3	1	0	0.000000	bihday majesty	NN NN
0	#model i love u take with u all the time in ...	model love u take u time ur!!!	12	11	5	3	0.976562	model love u take u time ur !!	NN NN JJ VB JJ NN JJ . . .

# Feature Engineering

## Meta Features



label	tweet	cleaned	word_count	unique_words_count	stopwords_count	punctuation_count	polarity	keywords	POS_tagged
0	@user when a father is dysfunctional and is s...	father dysfunctional selfish drags kids dysfun...	17	15	10	1	-0.500000	father dysfunctional selfish drags kids dysfun...	RBR JJ JJ NNS NNS NN . VB
0	@user @user thanks for #lyft credit i can't us...	thanks lyft credit can't use cause offer wheel...	17	17	5	2	0.200000	use cause offer wheelchair vans pdx thanks lyf...	NNS VBP NN MD RB VB NN NN NN NNS VBP . VBN VBD
0	bihday your majesty	bihday majesty	3	3	1	0	0.000000	bihday majesty	NN NN
0	#model i love u take with u all the time in ...	model love u take u time ur!!!	12	11	5	3	0.976562	model love u take u time ur !!	NN NN JJ VB JJ NN JJ . . .

# Feature Engineering

TextBlob

label	tweet	cleaned	word_count	unique_words_count	stopwords_count	punctuation_count	polarity	keywords	POS_tagged
0	@user when a father is dysfunctional and is s...	father dysfunctional selfish drags kids dysfun...	17	15	10	1	-0.500000	father dysfunctional selfish drags kids dysfun...	RBR JJ JJ NNS NNS NN . VB
0	@user @user thanks for #lyft credit i can't us...	thanks lyft credit can't use cause offer wheel...	17	17	5	2	0.200000	use cause offer wheelchair vans pdx thanks lyf...	NNS VBP NN MD RB VB NN NN NN NNS VBP . VBN VBD
0	bihday your majesty	bihday majesty	3	3	1	0	0.000000	bihday majesty	NN NN
0	#model i love u take with u all the time in ...	model love u take u time ur!!!	12	11	5	3	0.976562	model love u take u time ur !!	NN NN JJ VB JJ NN JJ . . .



# Feature Engineering

Rake algorithm  
from nltk

label	tweet	cleaned	word_count	unique_words_count	stopwords_count	punctuation_count	polarity	keywords	POS_tagged
0	@user when a father is dysfunctional and is s...	father dysfunctional selfish drags kids dysfun...	17	15	10	1	-0.500000	father dysfunctional selfish drags kids dysfun...	RBR JJ JJ NNS NNS NN . VB
0	@user @user thanks for #lyft credit i can't us...	thanks lyft credit can't use cause offer wheel...	17	17	5	2	0.200000	use cause offer wheelchair vans pdx thanks lyf...	NNS VBP NN MD RB VB NN NN NN NNS VBP . VBN VBD
0	bihday your majesty	bihday majesty	3	3	1	0	0.000000	bihday majesty	NN NN
0	#model i love u take with u all the time in ...	model love u take u time ur!!!	12	11	5	3	0.976562	model love u take u time ur !!	NN NN JJ VB JJ NN JJ . . .

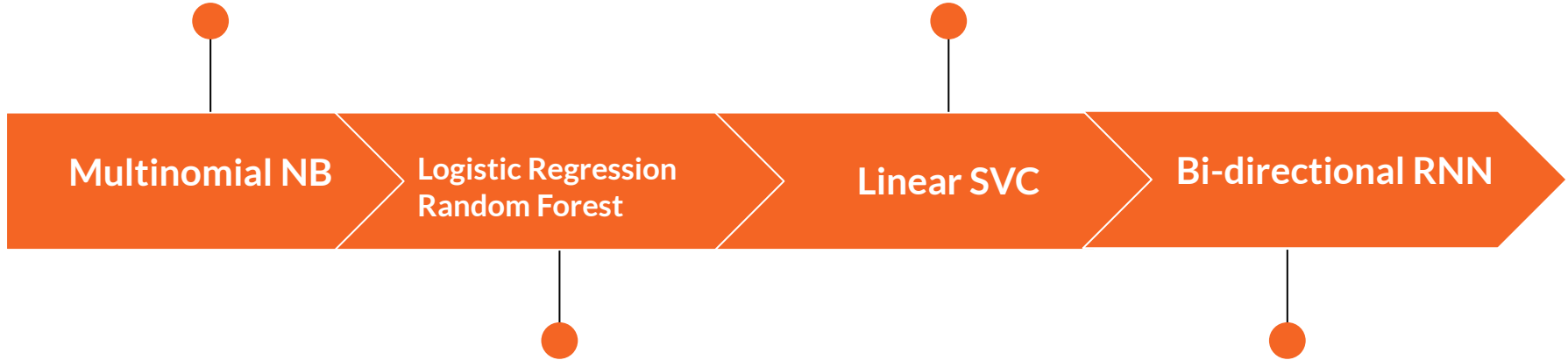
# Feature Engineering

Pos\_tag\_sents  
from nltk

label	tweet	cleaned	word_count	unique_words_count	stopwords_count	punctuation_count	polarity	keywords	POS_tagged
0	@user when a father is dysfunctional and is s...	father dysfunctional selfish drags kids dysfun...	17	15	10	1	-0.500000	father dysfunctional selfish drags kids dysfun...	RBR JJ JJ NNS NNS NN . VB
0	@user @user thanks for #lyft credit i can't us...	thanks lyft credit can't use cause offer wheel...	17	17	5	2	0.200000	use cause offer wheelchair vans pdx thanks lyf...	NNS VBP NN MD RB VB NN NN NN NNS VBP . VBN VBD
0	bihday your majesty	bihday majesty	3	3	1	0	0.000000	bihday majesty	NN NN
0	#model i love u take with u all the time in ...	model love u take u time ur!!!	12	11	5	3	0.976562	model love u take u time ur !!	NN NN JJ VB JJ NN JJ . . .

**Baseline Method**

**1st Approach**



**Failed Attempts**

**2nd Approach**

---

# Baseline Method

## Multinomial NB

- Score difference  
(CountVectorizer vs  
TfidfVectorizer)

```
Accuracy score: 0.9501016737056155  
Precision score: 0.9913793103448276  
Recall score: 0.26558891454965355  
F1 score: 0.41894353369763204
```

---

# Failed Attempts

- Random Forest
  - Logistic Regression
-

---

# 1st Approach

## LinearSVC

- Loss: squared\_hinge

```
Accuracy score: 0.9641795714062256  
Precision score: 0.8333333333333334  
Recall score: 0.5889145496535797  
F1 score: 0.6901217861975644
```

---

---

# 2nd Approach

## Bi-directional RNN

- Optimizer: Adam
- Loss function: binary cross entropy
- Activation: Sigmoid
- Epochs = 5

```
Test accuracy: 94.85%  
Test precision: 94.87%  
Test recall: 94.81%
```

# Challenges

1. Biased Dataset
  2. Memory allocation error for abundant features
  3. Limited Access to sufficient RAM/memory
  4. Implementation of POS\_tag & Cosine similarity
-



---

# Questions!

---