A Study on Rumour Detection on Online Social Networks

by Cheng Gibson

ROADMAP

ROADMAP

1. Problem Statement
2. Methodology
3. Dataset Characteristics
4. Hypotheses & Results
5. Problems Encountered
6. Conclusion & Future Work

1.

PROBLEM STATEMENT

FAKE NEWS?

IS FAKE NEWS EVEN A PROBLEM?

- Internet as a reputable news source
- Major news streams leveraging Internet
- Internet information affects public perception
- Internet information may or may not be *legitimate*
- Authorities are taking notice of this trend

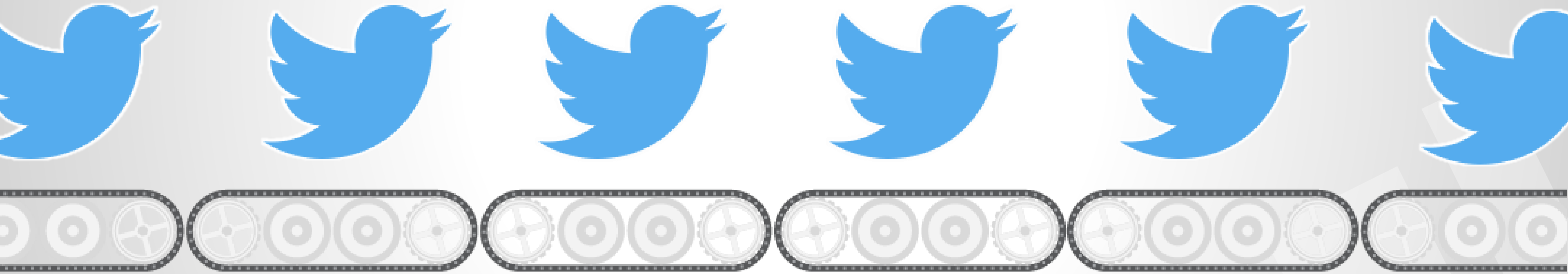
YES, FAKE NEWS IS A PROBLEM!

Top 5 Fake Election Stories by Facebook Engagement

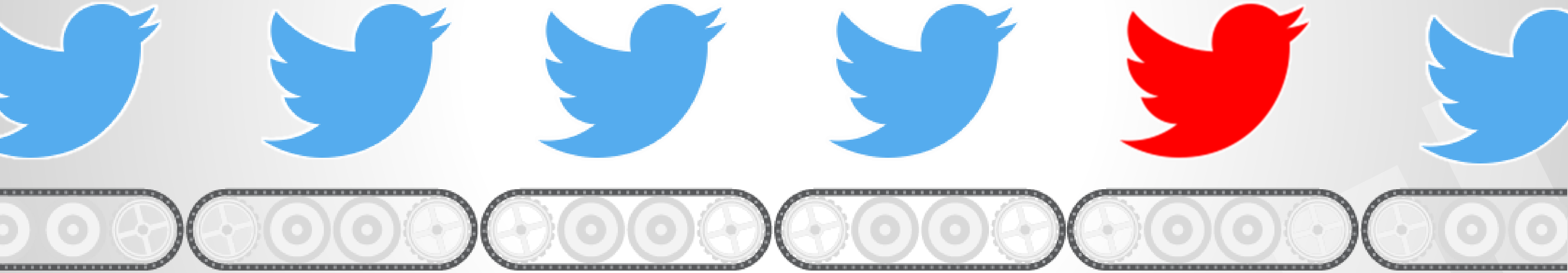
(three months before election)

1. "Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement" *(960,000, Ending the Fed)*
2. "WikiLeaks CONFIRMS Hillary Sold Weapons to ISIS... Then Drops Another BOMBSHELL! Breaking News" *(789,000, The Political Insider)*
3. "IT'S OVER: Hillary's ISIS Email Just Leaked & It's Worse Than Anyone Could Have Imagined" *(754,000, Ending the Fed)*
4. "Just Read the Law: Hillary Is Disqualified From Holding Any Federal Office" *(701,000, Ending the Fed)*
5. "FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide" *(567,000, Denver Guardian)*

ENGAGEMENT REFERS TO THE TOTAL NUMBER OF SHARES, REACTIONS, AND COMMENTS FOR A PIECE OF CONTENT ON FACEBOOK SOURCE: FACEBOOK DATA VIA BUZZSUMO



IN THE NEWS MILL...



IN THE NEWS MILL...

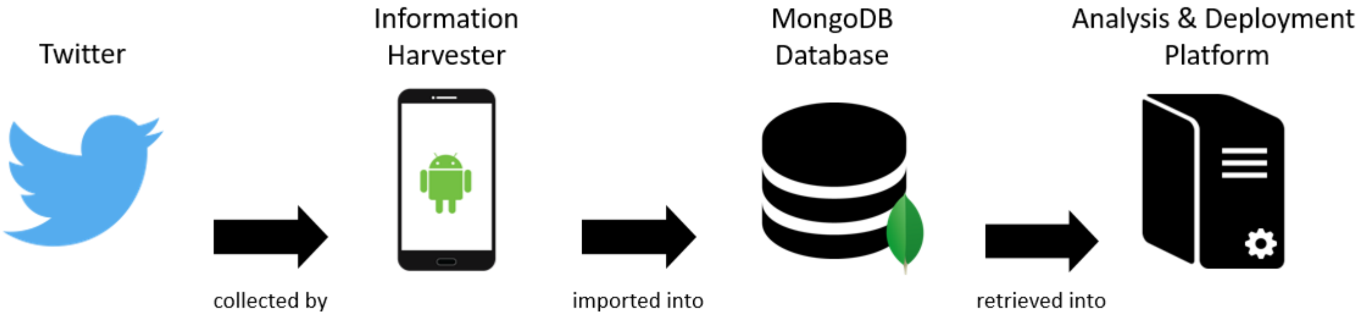
HOW DO WE IDENTIFY RUMORS FROM NEWS?

WHAT ARE THE KEY TRAITS OF RUMORS ON ONLINE SOCIAL NETWORKS?

2.

METHODOLOGY

GENERAL DATA WORKFLOW



TWITTER

2nd

Largest Social Networking Site

1,300,000,000

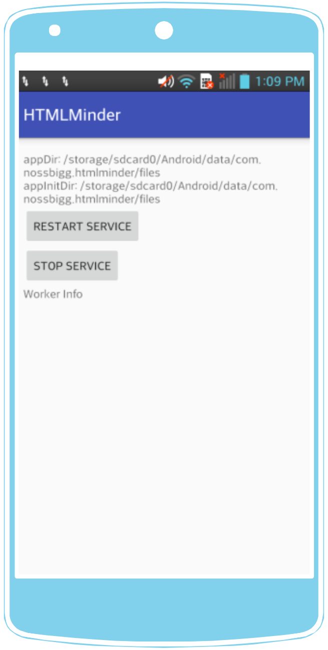
Twitter Accounts

5,000,000

Tweets per Day

INFORMATION HARVESTER

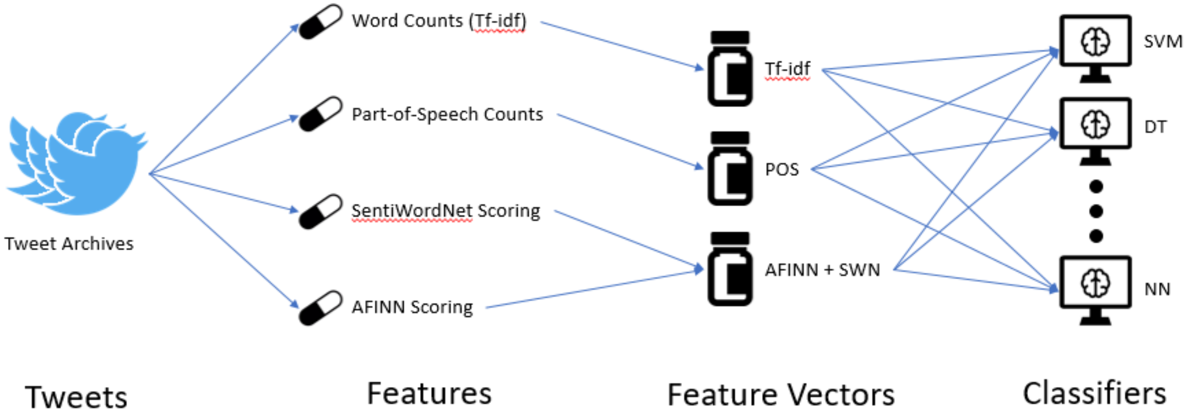
- Automated 24/7 tweet collection
- Network optimizations
- Duplicate tweet reduction
- Gzipped archives for 90% space savings



DATA PREPROCESSING

1. Decompress archives
2. Remove tweet duplicates
3. Label tweets with tweet types
4. Generate tweet relationship data
5. Generate tweet sentiments

FEATURE ENGINEERING: OVERVIEW

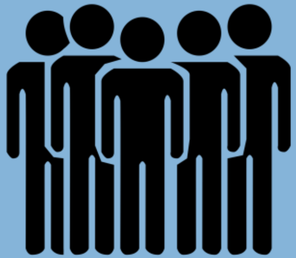


Tt

CONTENT-BASED

PROPAGATION-BASED

USER-BASED

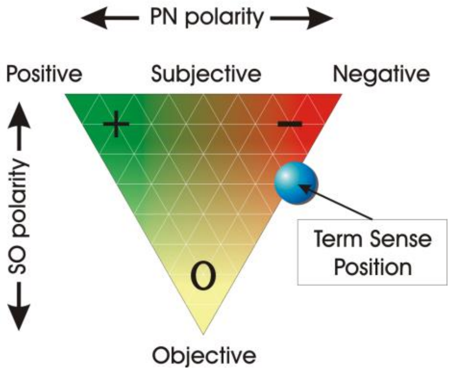


OTHER-BASED

Zhang et al, "Automatic Detection of Rumor on Social Network," *Natural Language Processing and Chinese Computing Lecture Notes in Computer Science*, pp. 113-122, 2015.

SENTIMENT ANALYSIS LIBRARIES

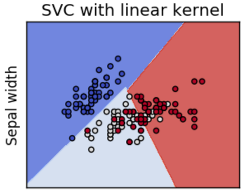
- SentiWordNet Scoring
- AFINN Scoring
- NLTK Part-of-Speech Tagger
  - NPS Corpus



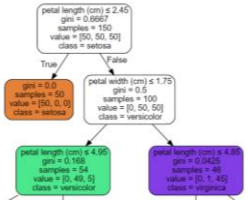
MANUAL LABELLING: SCHEME

Label	Description
Support (S)	Supports a rumour explicitly or implicitly (eg. using rumour to augment another opinion)
Deny (D)	Denies or disproves a rumour
Neutral (N)	Neither supports nor rejects a rumour. Such rumours may also occasionally express doubt about the rumour
Unrelated (U)	Unrelated to the rumour

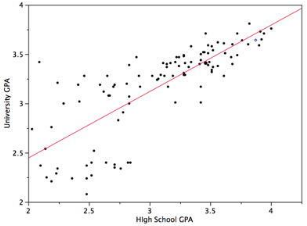
MACHINE LEARNING TECHNIQUES



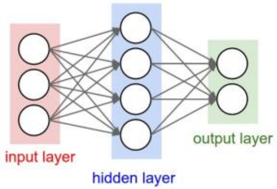
Support Vector Machine



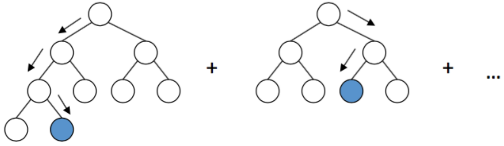
Decision Trees



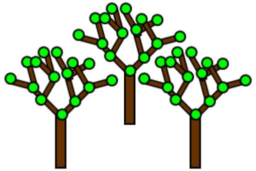
Linear Regression



Neural Networks



Random Forest Regressor



Gradient Boosted Trees

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

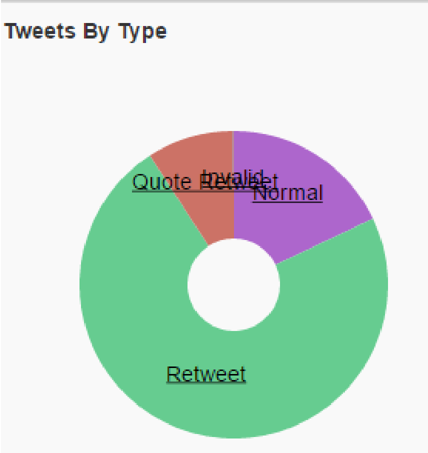
Naïve Bayes

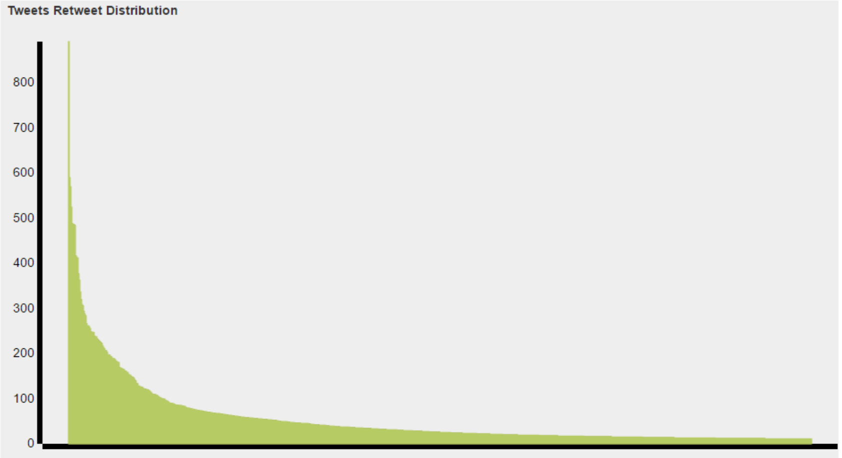
FULL SYSTEM INTEGRATION

Web backend with
Python + Django

Visualization with D3.js

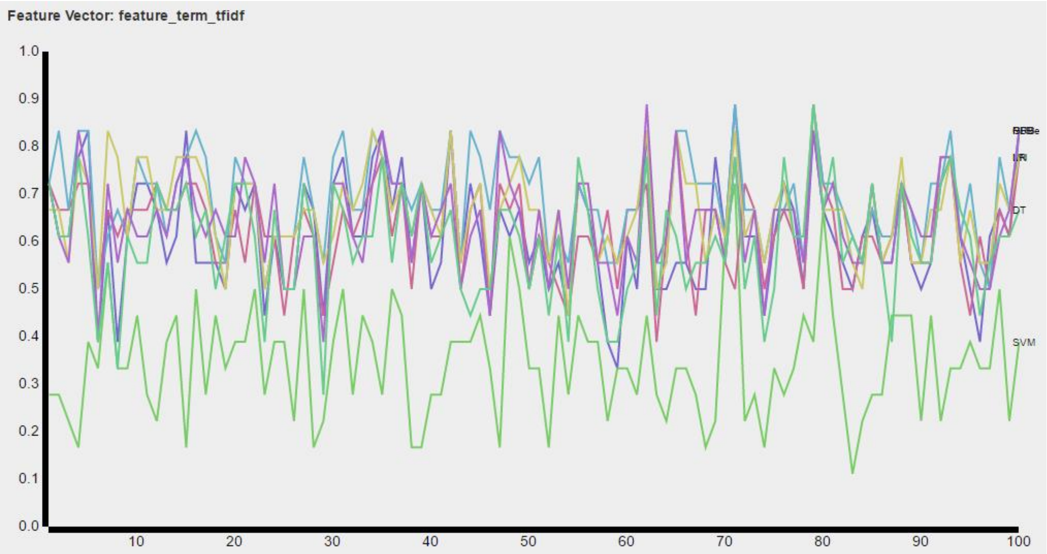
VISUALIZATIONS





for Datasets

VISUALIZATIONS



for Experiment Results

VISUALIZATIONS

A LIVE DEMONSTRATION! ☺

< insert live demonstration here >

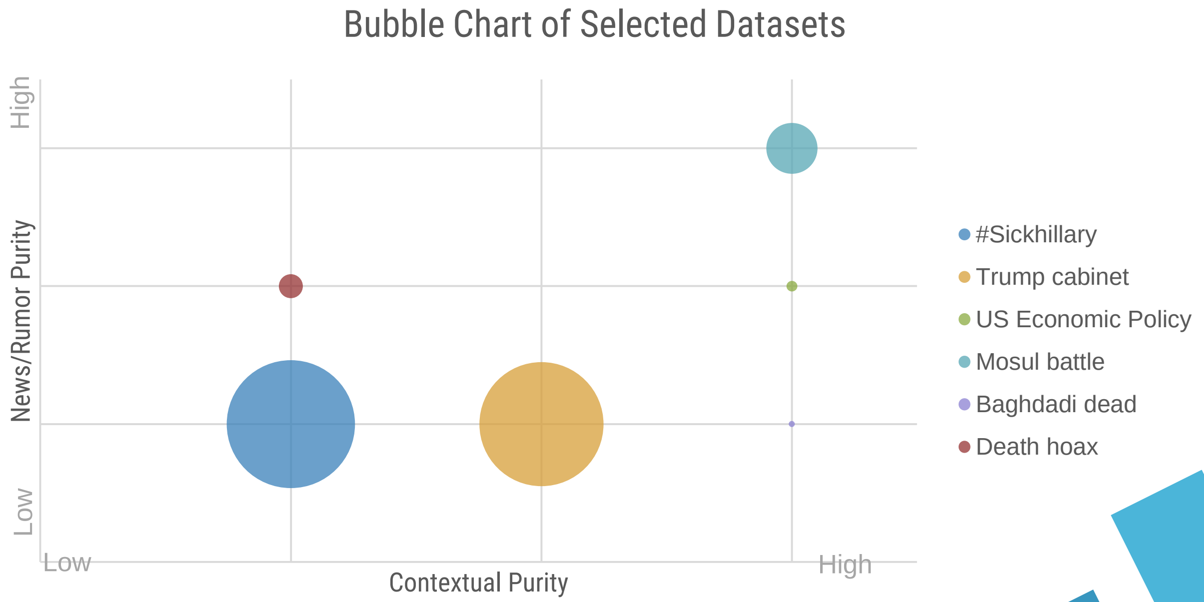
3.

DATASET CHARACTERISTICS

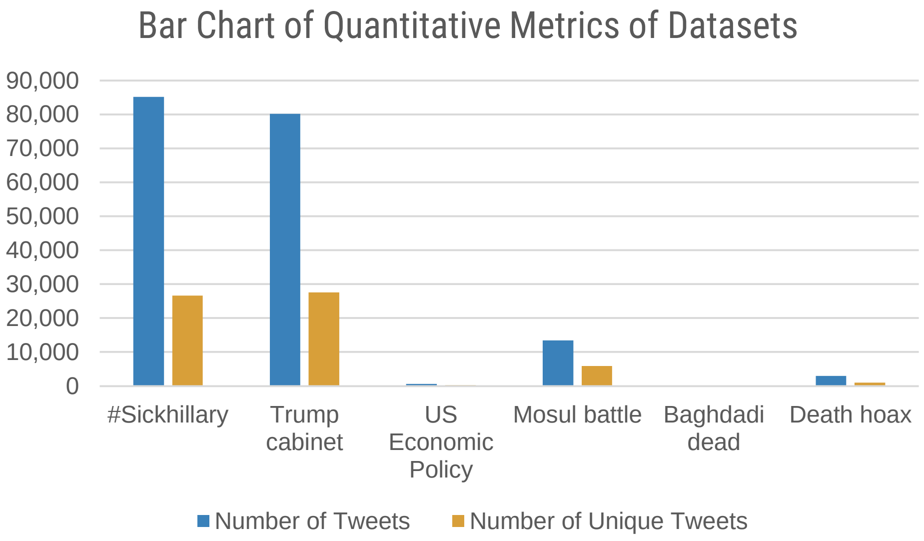
DATASET TYPES

Rumour-centric Datasets		
#Sickhillary	Baghdadi dead	Death hoax
News-centric Datasets		
Mosul Battle	US Economic Policy	Trump cabinet

QUALITATIVE METRICS



QUANTITATIVE METRICS



QUANTITATIVE METRICS

Dataset	Number of Tweets	Number of Unique Tweets
#Sickhillary	85,194	26,642
Trump cabinet	80,145	27,621
US Economic Policy	587	216
Mosul battle	13,483	5,886
Baghdadi dead	192	89
Death hoax	2,934	1,035

MANUAL LABELLING EXAMPLES

Dataset	Label	Tweet Text
Sickhillary	Support	RT @ThePatriot143: Kimmel: #SickHillary Conspiracy Theories Would Be Harder to Believe If They Didn't Actually Come True #ThanksObama https…
Death hoax	Deny	Hoax Exposed: Muslim Man Blamed Trump For Iraqi Mother's Death » Alex Jones' Infowars: There's a war on for your... https://t.co/5bOrJ8daCD
Baghdadi dead	Neutral	Isis commanders killed in Iraqi air strikes targeting Baghdadi, but leader's fate remains unknown https://t.co/oqlyJks1e5
Death hoax	Unrelated	We have a new record coming. We would love for you to pre-order with us at @Bandcamp . Thank you https://t.co/2M9cdJ54xC

4.

HYPOTHESES & RESULTS

TESTING METHODOLOGY

- **100** Iterations per Experiment
- Statistical Analysis of Classifier Performance
- Scoring Equation
  - Highest Minimum, Highest Median, Lowest Standard Deviation

$$score(c)_i = w_{min} rank\_min(c)_i + \mathrm{rank\_median}(c)_i w_{median} + \mathrm{rank\_stddev}(c)_i w_{stddev}$$

HYPOTHESES

1. What are the general sentiment profiles of the datasets?
2. How well can rumours and non-rumours be separated in rumour-centric datasets?
3. How well can rumours and non-rumours be separated using all datasets?

i.

WHAT ARE THE GENERAL SENTIMENT PROFILES OF THE DATASETS?

1. What are the general sentiment profiles of the datasets?

- » Discovering sentiment differences within dataset
- » Comparing tweets only within dataset
- » Logistic Regression (LR) used only
- » Interpreting LR coefficients
- » Using AFINN + SWN and POS features only

1. What are the general sentiment profiles of the datasets?

- ‘trump cabinet’ - Good
- ‘sickhillary’ - Slightly Good
- ‘us economic policy’ - Slightly Good
- ‘baghdadi dead’ - Poor

1. What are the general sentiment profiles of the datasets?

Dataset: ‘trump cabinet’ (News)

Classifier Performance: Good

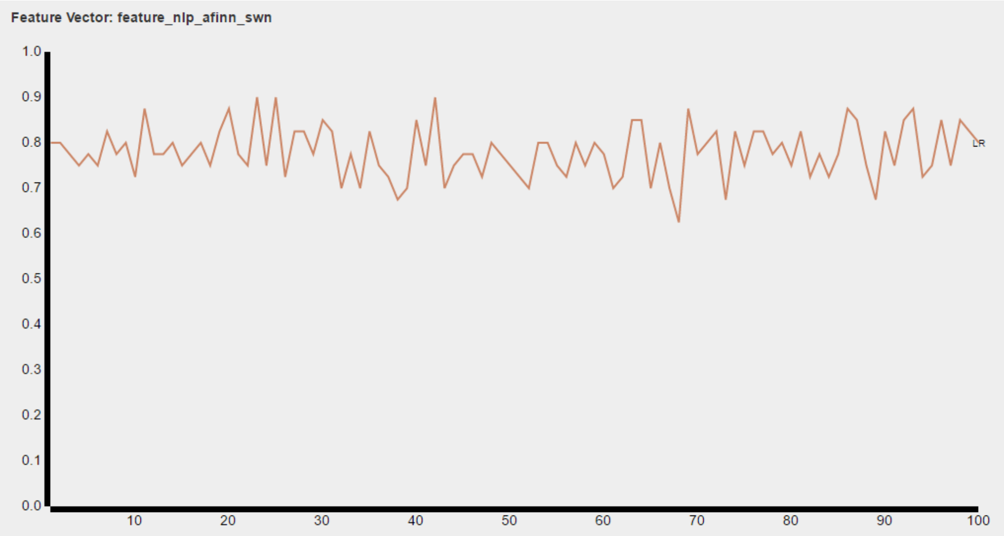
Collection: trump_cabinet

Best feature vector: feature_nlp_afinn_swn

Feature	Best Classifier	Min	Max	Mean	Median	Std Dev	25th Percentile	75th Percentile	Overall Score
feature_nlp_afinn_swn	LR	0.62500	0.90000	0.77975	0.77500	0.05609	0.75000	0.82500	0.35157
feature_nlp_pos	LR	0.62500	0.87500	0.75525	0.75000	0.06065	0.72500	0.80000	0.34559

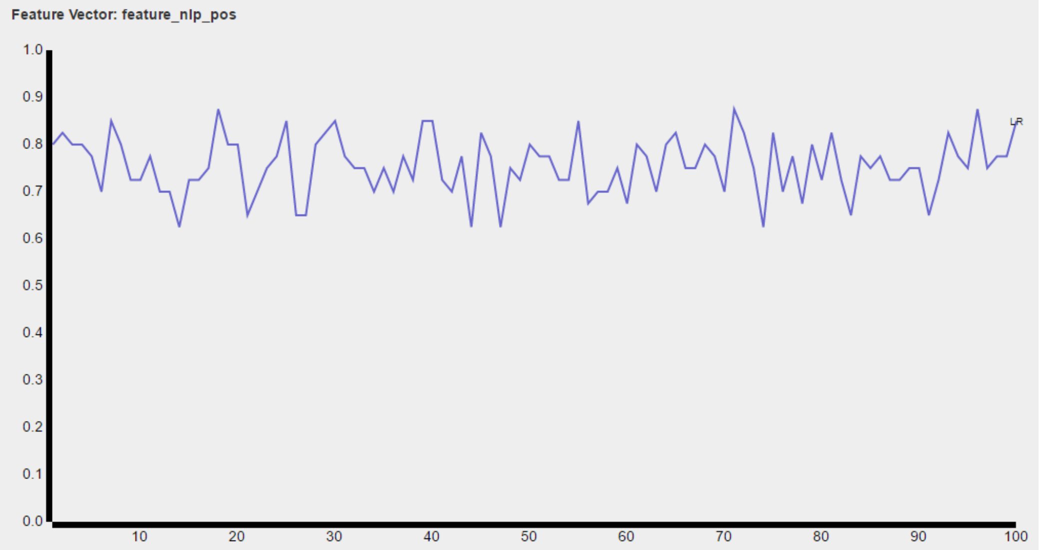
1. What are the general sentiment profiles of the datasets?

Dataset: ‘trump cabinet’, Feature: AFINN + SWN



1. What are the general sentiment profiles of the datasets?

Dataset: ‘trump cabinet’, Feature: POS



1. What are the general sentiment profiles of the datasets?

Dataset: ‘trump cabinet’, LR Coefficients

Feature (Ft)	AFINN	Pos	Neg	Obj
Weight (Wt)	-0.103	0.059	0.065	0.430

Ft	ADJ	ADP	ADV	CONJ	DET	NOUN	NUM	PRT	PRON	VERB	.	X
Wt	0.032	-0.025	0.061	0.142	0.063	0.104	0.066	0.048	-0.003	0.082	-0.531	0.066

1. What are the general sentiment profiles of the datasets?

Dataset: ‘sickhillary’ (Rumour)

Classifier Performance: Slightly Good

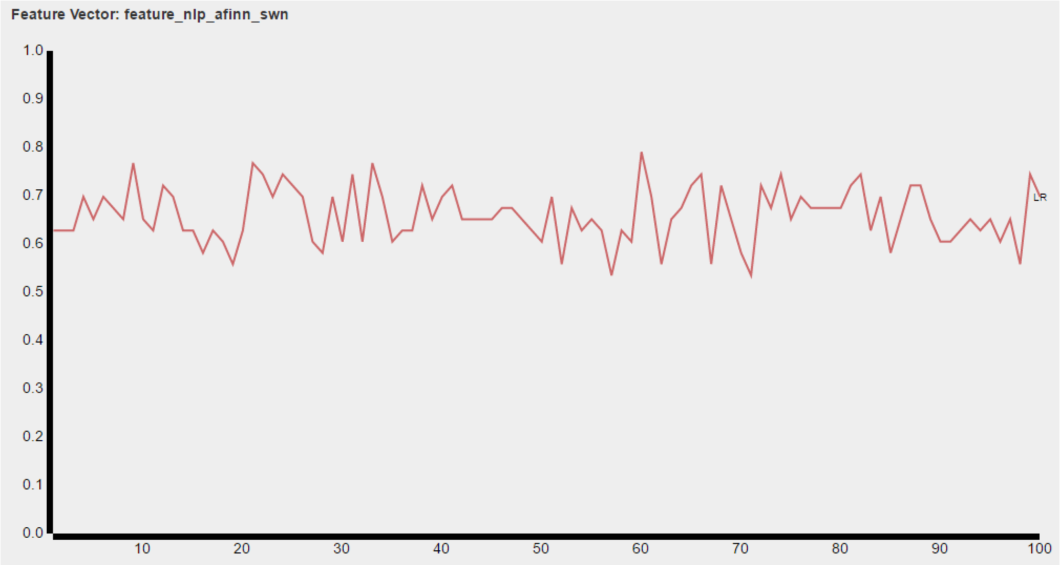
Collection: sickhillary

Best feature vector: feature_nlp_afinn_swn

Feature	Best Classifier	Min	Max	Mean	Median	Std Dev	25th Percentile	75th Percentile	Overall Score
feature_nlp_afinn_swn	LR	0.53488	0.79070	0.65953	0.65116	0.05625	0.62791	0.69767	0.29809
feature_nlp_pos	LR	0.41860	0.86047	0.64465	0.62791	0.07486	0.60465	0.69767	0.26443

1. What are the general sentiment profiles of the datasets?

Dataset: ‘sickhillary’, Feature: AFINN + SWN



1. What are the general sentiment profiles of the datasets?

Dataset: ‘sickhillary’, LR Coefficients

Feature (Ft)	AFINN	Pos	Neg	Obj
Weight (Wt)	-0.026	0.060	0.054	0.258

1. What are the general sentiment profiles of the datasets?

Dataset: ‘us economic policy’ (News)

Classifier Performance: Slightly Good

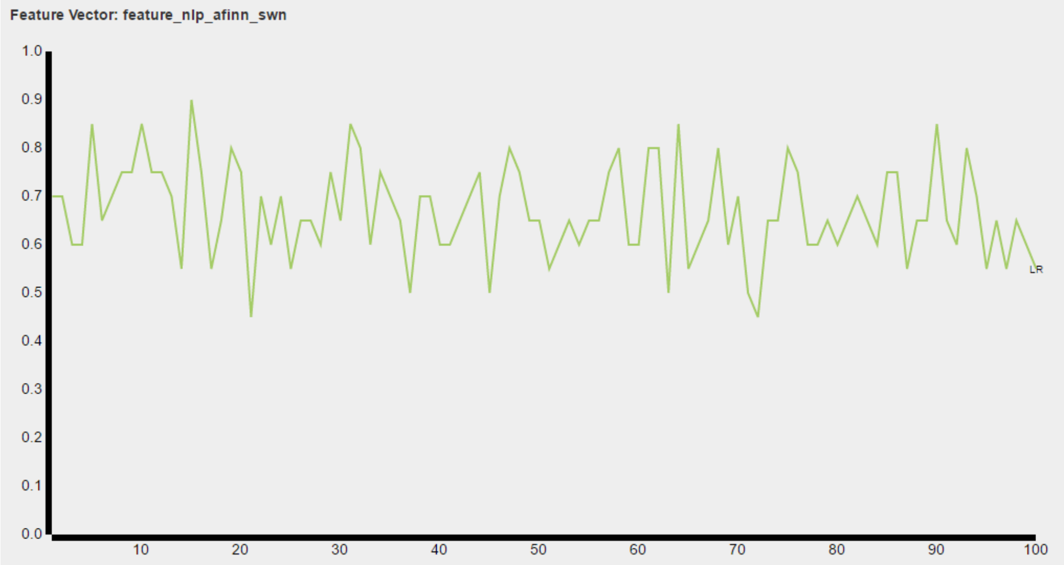
Collection: sickhillary

Best feature vector: feature_nlp_afinn_swn

Feature	Best Classifier	Min	Max	Mean	Median	Std Dev	25th Percentile	75th Percentile	Overall Score
feature_nlp_afinn_swn	LR	0.53488	0.79070	0.65953	0.65116	0.05625	0.62791	0.69767	0.29809
feature_nlp_pos	LR	0.41860	0.86047	0.64465	0.62791	0.07486	0.60465	0.69767	0.26443

1. What are the general sentiment profiles of the datasets?

Dataset: ‘us economic policy’, Feature: AFINN + SWN



1. What are the general sentiment profiles of the datasets?

Dataset: ‘us economic policy’, LR Coefficients

Feature (Ft)	AFINN	Pos	Neg	Obj
Weight (Wt)	-0.004	-0.024	-0.017	0.224

1. What are the general sentiment profiles of the datasets?

Dataset: ‘baghdadi dead’ (Rumour)

Classifier Performance: Poor

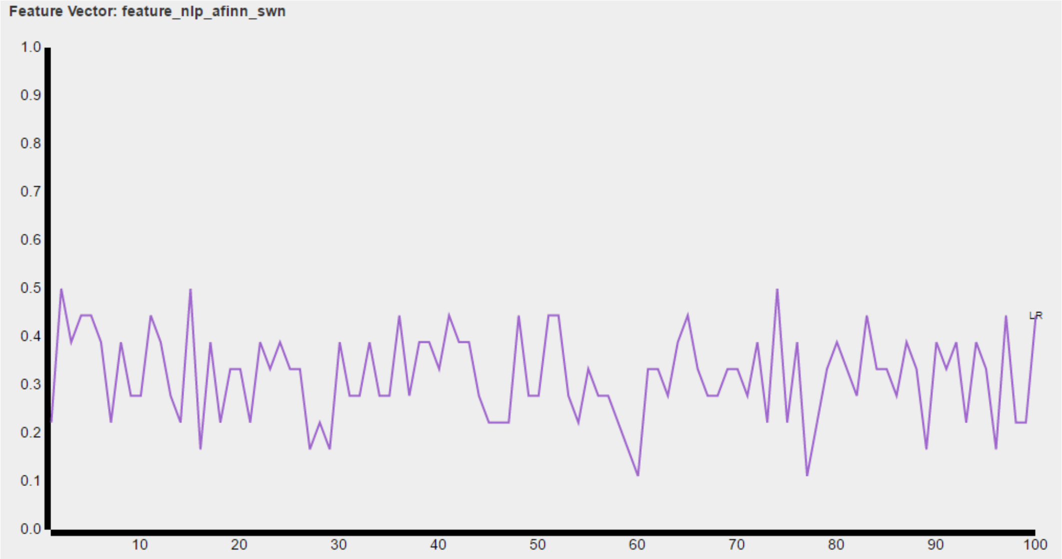
Collection: baghdadi_dead

Best feature vector: feature_nlp_afinn_swn

Feature	Best Classifier	Min	Max	Mean	Median	Std Dev	25th Percentile	75th Percentile	Overall Score
feature_nlp_afinn_swn	LR	0.11111	0.50000	0.31889	0.33333	0.09014	0.26389	0.38889	0.11517
feature_nlp_pos	LR	0.05556	0.50000	0.32667	0.33333	0.09663	0.27778	0.38889	0.10189

1. What are the general sentiment profiles of the datasets?

Dataset: ‘baghdadi dead’, Feature: AFINN + SWN



ii.

HOW WELL CAN RUMOURS AND NON-RUMOURS BE SEPARATED IN RUMOUR-CENTRIC DATASETS?

2. How well can rumours and non-rumours be separated in rumour-centric datasets?

- Comparing tweets only within dataset
- All machine learning models used
- All feature vectors used

2. How well can rumours and non-rumours be separated in rumour-centric datasets?

Dataset: ‘sickhillary’ (Rumour)

Classifier Performance: Slightly Good

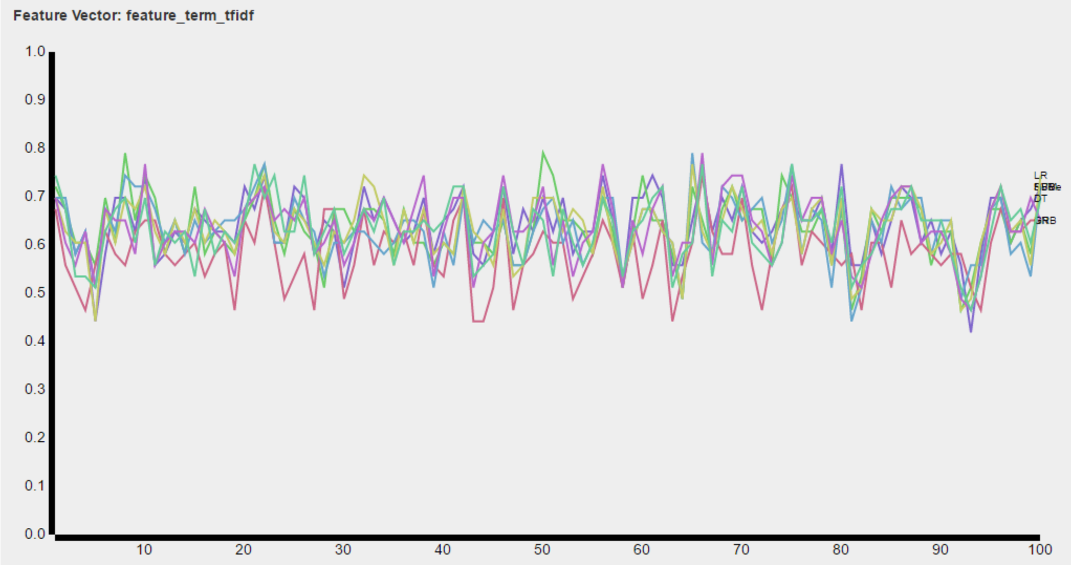
Collection: sickhillary

Best feature vector: feature_nlp_pos

Feature	Best Classifier	Min	Max	Mean	Median	Std Dev	25th Percentile	75th Percentile	Overall Score
feature_nlp_afinn_swn	NN	0.44186	0.79070	0.62581	0.62791	0.07100	0.58140	0.67442	0.26996
feature_nlp_pos	SVM	0.46512	0.79070	0.64419	0.65116	0.06903	0.58140	0.69767	0.28145
feature_term_tfidf	SVM	0.46512	0.79070	0.63628	0.63953	0.06607	0.58140	0.67442	0.27835

2. How well can rumours and non-rumours be separated in rumour-centric datasets?

Dataset: ‘sickhillary’, Feature: Tf-idf



2. How well can rumours and non-rumours be separated in rumour-centric datasets?

Dataset: ‘baghdadi dead’ (Rumour)

Classifier Performance: Slightly Good

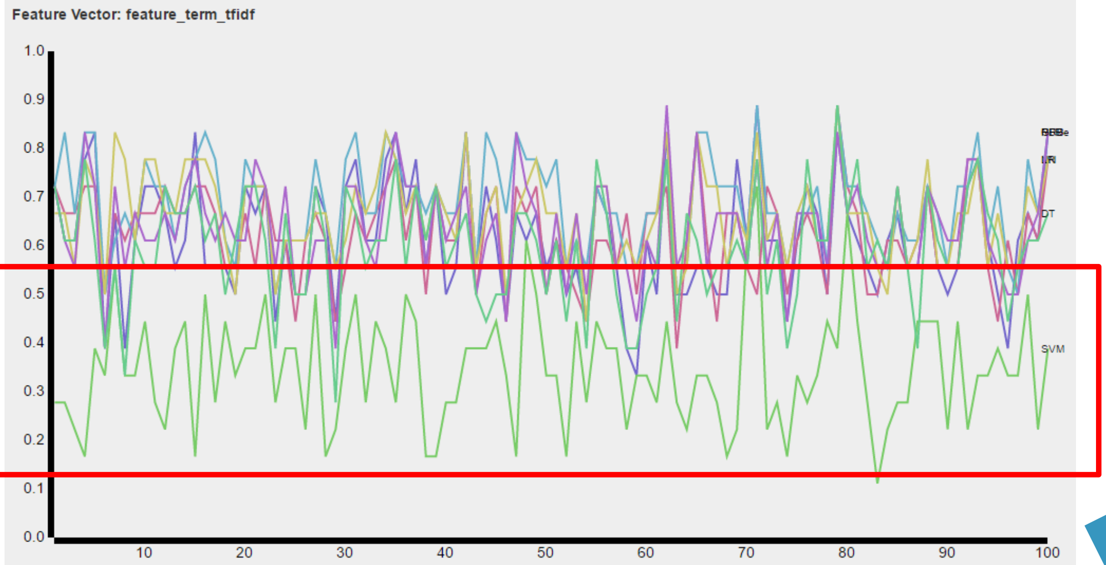
Collection: baghdadi_dead

Best feature vector: feature_term_tfidf

Feature	Best Classifier	Min	Max	Mean	Median	Std Dev	25th Percentile	75th Percentile	Overall Score
feature_nlp_afinn_swn	RFR	0.33333	0.72222	0.53278	0.55556	0.09464	0.44444	0.61111	0.22670
feature_nlp_pos	GRB	0.27778	0.77778	0.52556	0.55556	0.11848	0.44444	0.61111	0.21535
feature_term_tfidf	NBBernoulli	0.50000	0.88889	0.69111	0.66667	0.09601	0.61111	0.77778	0.29628

2. How well can rumours and non-rumours be separated in rumour-centric datasets?

Dataset: 'baghdadi dead', Feature: Tf-idf



2. How well can rumours and non-rumours be separated in rumour-centric datasets?

Dataset: ‘death hoax’ (Rumour)

Classifier Performance: Average

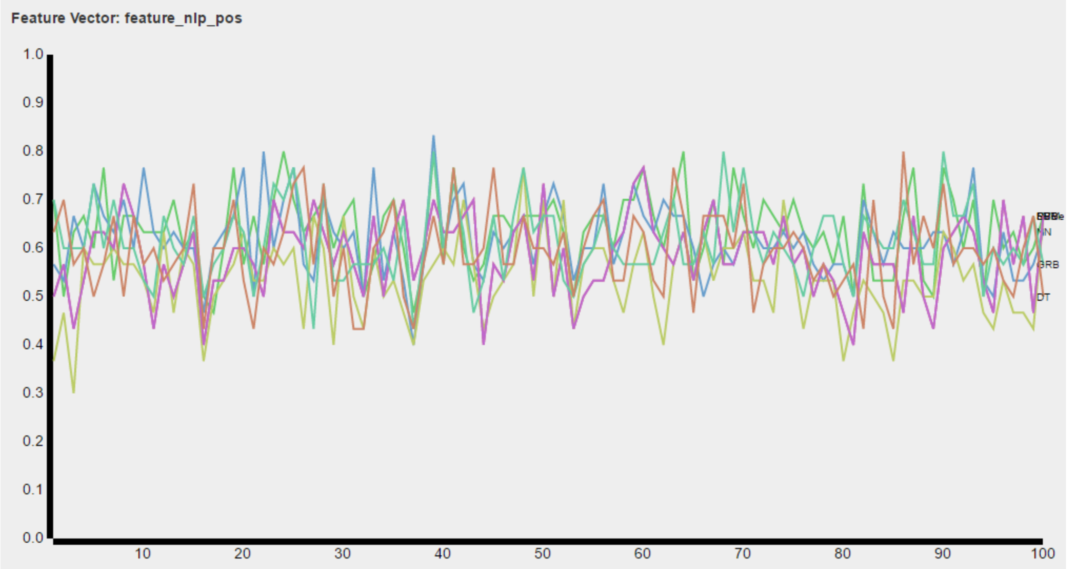
Collection: death_hoax

Best feature vector: feature_nlp_pos

Feature	Best Classifier	Min	Max	Mean	Median	Std Dev	25th Percentile	75th Percentile	Overall Score
feature_nlp_afinn_swn	NN	0.43333	0.80000	0.61500	0.63333	0.07369	0.56667	0.66667	0.26938
feature_nlp_pos	RFR	0.46667	0.80000	0.63633	0.63333	0.07875	0.59167	0.70000	0.27810
feature_term_tfidf	NN	0.46667	0.80000	0.62000	0.63333	0.07394	0.56667	0.66667	0.27773

2. How well can rumours and non-rumours be separated in rumour-centric datasets?

Dataset: ‘death hoax’, Feature: POS



2. How well can rumours and non-rumours be separated in rumour-centric datasets?

- ‘sickhillary’ - Slightly Good
- ‘baghdadi dead’ - Slightly Good
- ‘death hoax’ - Average

iii.

HOW WELL CAN RUMOURS AND NON-RUMOURS BE SEPARATED USING ALL DATASETS?

3. How well can rumours and non-rumours be separated using all datasets?

- Aggregating all datasets together
- All machine learning models used
- All feature vectors used

3. How well can rumours and non-rumours be separated using all datasets?

Classifier Performance: Good – Very Good

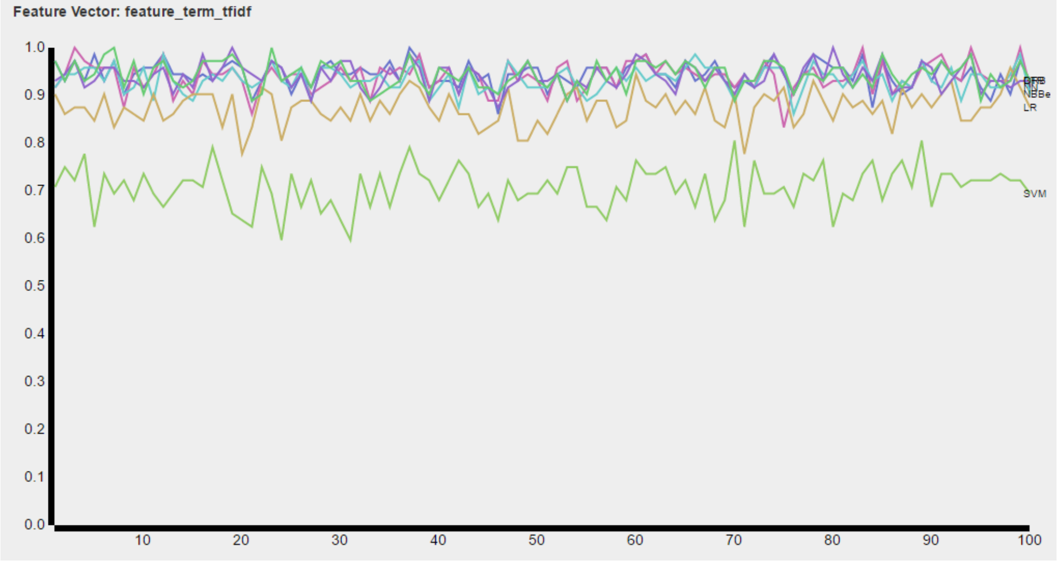
Collection: experiment_3

Best feature vector: feature_term_tfidf

Feature	Best Classifier	Min	Max	Mean	Median	Std Dev	25th Percentile	75th Percentile	Overall Score
feature_nlp_afinn_swn	RFR	0.65278	0.90278	0.77958	0.78472	0.04870	0.75000	0.81944	0.36056
feature_nlp_pos	RFR	0.63889	0.90278	0.76681	0.76389	0.04816	0.73611	0.79167	0.35185
feature_term_tfidf	DT	0.88889	1.00000	0.94306	0.94444	0.02732	0.92708	0.97222	0.45871

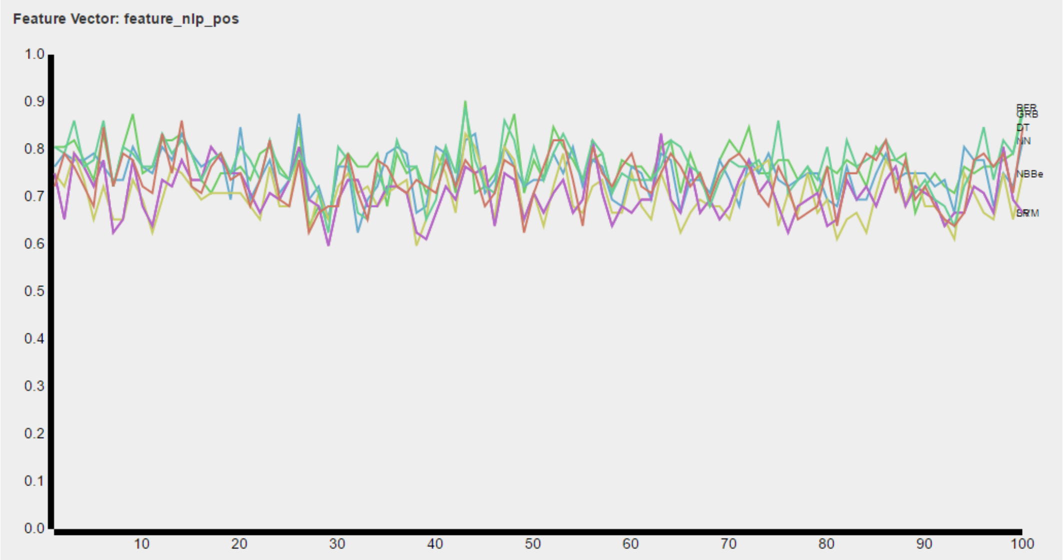
3. How well can rumours and non-rumours be separated using all datasets?

Feature: Tf-idf



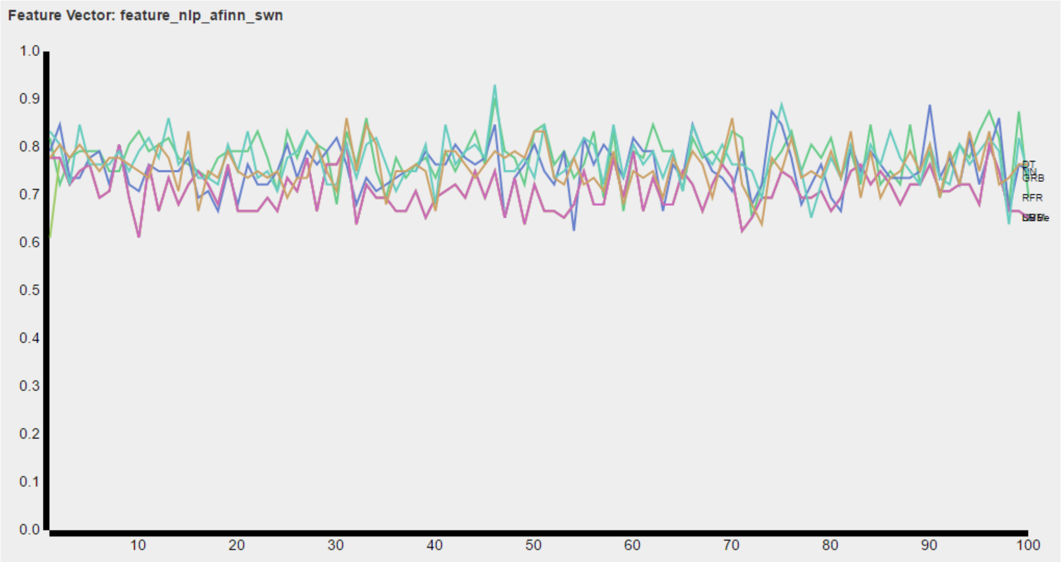
3. How well can rumours and non-rumours be separated using all datasets?

Feature: POS



3. How well can rumours and non-rumours be separated using all datasets?

Feature: AFINN + SWN



3. How well can rumours and non-rumours be separated using all datasets?

LR Coefficients

Feature (Ft)	AFINN	Pos	Neg	Obj
Weight (Wt)	0.515	-0.137	-0.166	0.827

Ft	ADJ	ADP	ADV	CONJ	DET	NOUN	NUM	PRT	PRON	VERB	.	X
Wt	-0.126	0.001	0.028	-0.24	-0.131	0.307	-0.028	-0.057	-0.090	0.138	1.018	-0.022

3. How well can rumours and non-rumours be separated using all datasets?

- Very Good
- High accuracy demonstrated in all models
- Sentiment Analysis features provided high performance levels for models
- Sentiment Analysis Libraries shows potential

5.

PROBLEMS ENCOUNTERED

PROBLEMS: RESEARCH DIRECTION

1. Uncertainty in knowing which next step to take

PROBLEMS: ANDROID

1. Inconsistent storage implementation across Android versions
2. Tackling Android’s automated task-killing
3. Unreliable MTP file access

PROBLEMS: MANUAL LABELLING

1. Uncertainty in identifying major contexts in datasets
2. Labelling large amounts of data

6.

CONCLUSION & FUTURE WORK

1.

Punctuations and **Objective words** are indicative of rumor/tweet nature of tweet.

II.

Sentiment Analysis libraries show promise

FUTURE WORK: INVESTIGATIVE

- Addressing sampling bias
- Testing against public datasets
- Testing against other types of corpuses
  - Articles, Forums
- Using other sentiment analysis libraries
  - LWIC, SentiStrength, ANEW
- Using NLP-specific classifiers
  - Conditional Random Field

FUTURE WORK: IMPLEMENTATION

- GPU acceleration
- Distributed databases
- Real-time processing & analysis
- Web UI tweet labelling

FIN

Thank you for your ears ☺
Any questions?