STAATLICH
ANERKANNTE
HOCHSCHULE

Fake news detection using topic and author agnostic approach supplemented by Twitter data

Case Study 1

Presented by:
Manmeet Kumar Chaudhuri (11011780)
Rishabh Garg (11011875)
Rohit Bewoor (11011831)
Sanika Medankar (11011861)
Shekhar Singh (11011694)
Siddarth Venkateswaran (11011753)

HOCHSCHULE HEIDELBERG
Intelligence in Learning

Content

STAATLICH
ANERKANNTE
HOCHSCHULE

- Introduction
- Topic and Author agnostic features
- Related works
- Methodology
  - Stage 1 : Data Collection
  - Stage 2 : Feature Extraction
  - Stage 3 : Classification Models
- Result
- Conclusion
- Future Scope
- Demo

Introduction

- People now consume news via online websites and social media more than ever before
- As per a report from Pew Research Centre (October 2019): 55% of US adults get their news from social media either “often” or “sometimes”, and that’s a 9% jump over last year
- Social media is like a double-edged sword, it can be used by authorities for low cost, easy and rapid dissemination of information or by malicious entities to rapidly spread rumours and fake news

- **We designed and implemented a method that is topic and author agnostic.**

Image Credit :
https://en.wikipedia.org/wiki/Fake_news

Topic and Author agnostic features

STAATLICH ANERKANNTE HOCHSCHULE

- **Morphological features:** Can be used to generalize the text.
  - Replacing words by their parts of speech ➡ allows identifying constructs (e.g. preposition, noun) instead of combinations of specific words.

pronoun	verb	adjective	noun	conjunction	pronoun	verb	pronoun
She	likes	big	snakes	but	I	hate	them

- **Psychological features:** They capture the percentage of total semantic words in texts.
  - Explore the links between ***word usage*** and basic ***social and personality characteristics***.

> **Donald Trump has been in public eye for over 30 years and he was never once accused of being racist by anyone until he decided to run against the Democrats.**

Analytical	Clout	Authenticity	Emotional Tone
83.9	74.8	65.1	25.8

Topic and Author agnostic features

- **Readability features:** Help text writers to adjust their content as per the target audience's level.
  - Obtained through readability scores and counting of character, words, and sentences usage (number of capitalised word in the text, Flesch reading rate, Gunning fog index, etc)

The attorney came up with several far-fetched arguments in a vain attempt to buttress his weak case.

Flesch reading ease	Gunning fog index
54.2	13.86

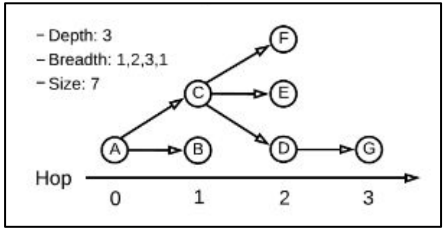
- **Author features:** Refer to the identity and credibility of an author.
  - E.g. presence of author name in a news article, location and professional affiliation of an author, Number of followers of an author, etc.

Topic and Author agnostic features

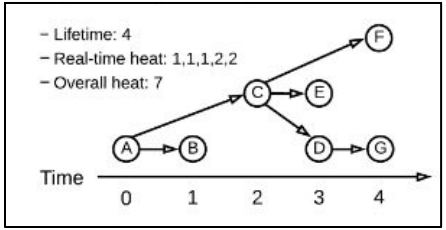
STAATLICH
ANERKANNTE
HOCHSCHULE

- Network propagation and social features ➡ ***Twitter Features***
- Extracted Tweets containing the URLs as proxy for this group
- Our features are broadly classified into four subcategories:
  - *User based*: Average number of tweets per unique user
  - *Timestamp based*: Span (time difference between the earliest and the latest tweet), Average Time between Tweets, Average Time of next Tweet, Binning based on Span of the tweets (Bin size = 6 hours, #Bins = 5), Counts of tweets by the day of the week
  - Count of total tweets for the URL
  - *User Action based*: actions taken by other Twitter users after seeing the tweets captured: Aggregates of the number of Likes, Retweets and Replies

Hop based



Time based



Related Works

STAATLICH
ANERKANNTE
HOCHSCHULE

According to The Guardian's David Shariatmadari:

- Fake news mimics reliable reporting – and people can't always tell the difference. Researchers have been trying to work out, what are the linguistic characteristics of fake news.

- One project, led by Fatemeh Torabi Asr, at Simon Fraser University in Canada, recently found that "on average, fake news articles use more words related to sex, death and anxiety". "Overly emotional" language is often deployed. In contrast, "Genuine news contains a larger proportion of words related to work (business) and money (economy)." - *Psychological features*

- Another group of researchers analyzed the relationship of various grammatical categories to fake news. They concluded that words which can be used to exaggerate are all found more often in deliberately misleading sources. These included superlatives, like "most" and "worst", and so-called subjective, like "brilliant" and "terrible". - *Morphological features*

Related Works

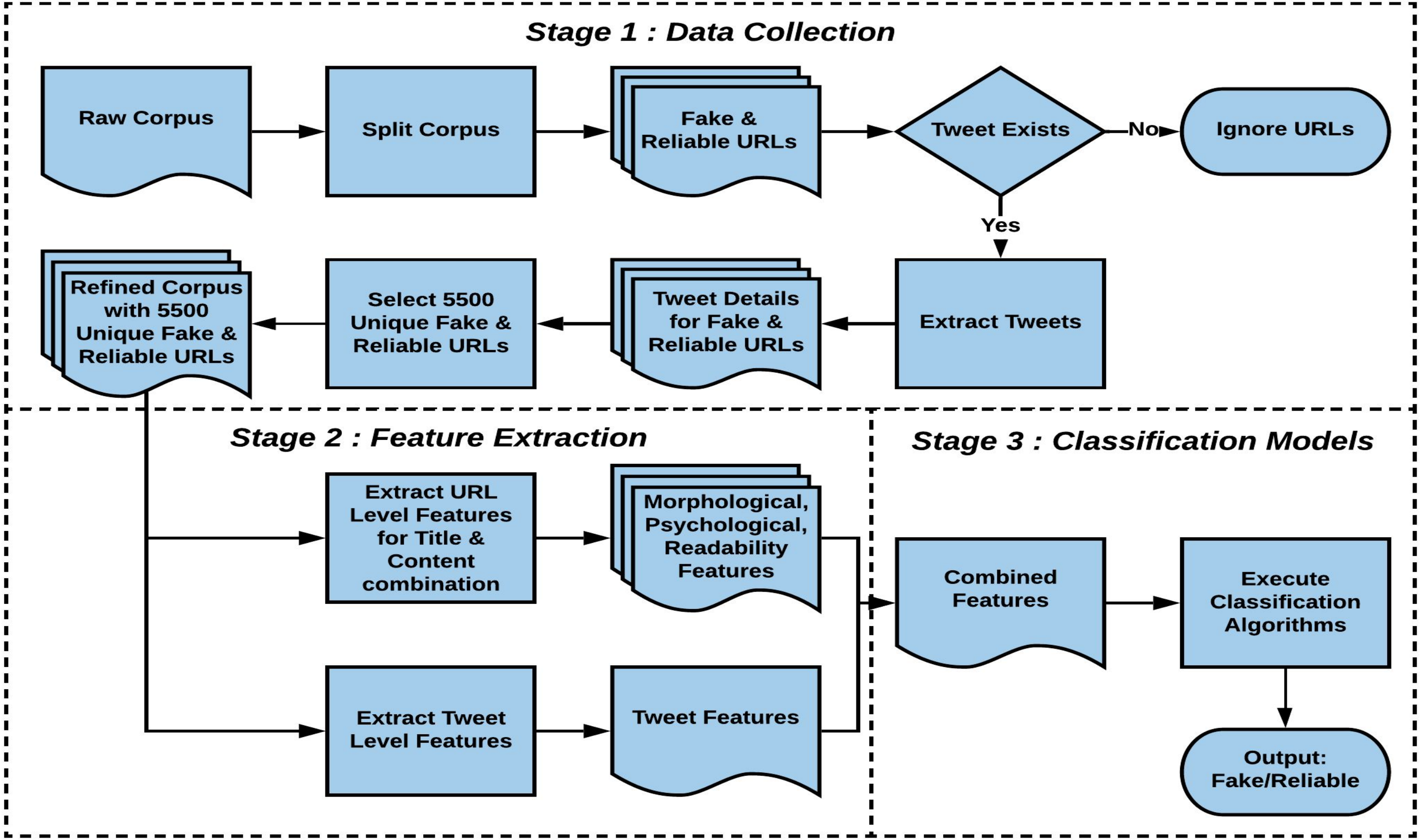
STAATLICH
ANERKANNTE
HOCHSCHULE

- Jack Grieve's team [University of Birmingham] compared 40 retracted and 41 non-retracted articles by Jason Blair, who resigned from the New York Times in disgrace in 2003. They found that there were more emphatic words, like "really" and "most", in Blair's retracted articles. Jason used shorter words and his language was less "informationally dense". The present tense cropped up more often and he relied on the third person pronouns "he" and "she" rather than full names – something that's typical of fiction. - *Readability and Morphological features.*

- Sonia Castelo and team in May 2019 presented an approach to detect fake news websites which uses topic-agnostic features using Web Markup, Morphological, Readability and Psychological features.

Methodology

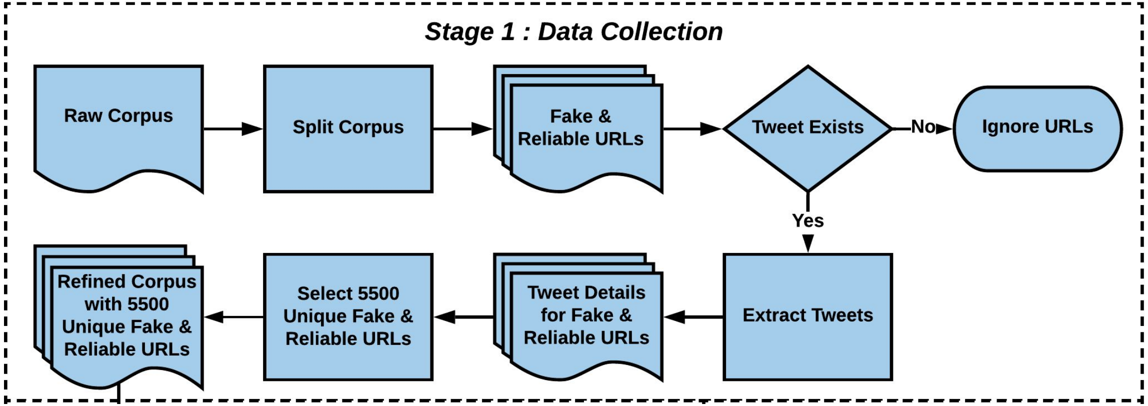
STAATLICH
ANERKANNTE
HOCHSCHULE



Stage 1 : Data Collection

STAATLICH
ANERKANNTE
HOCHSCHULE

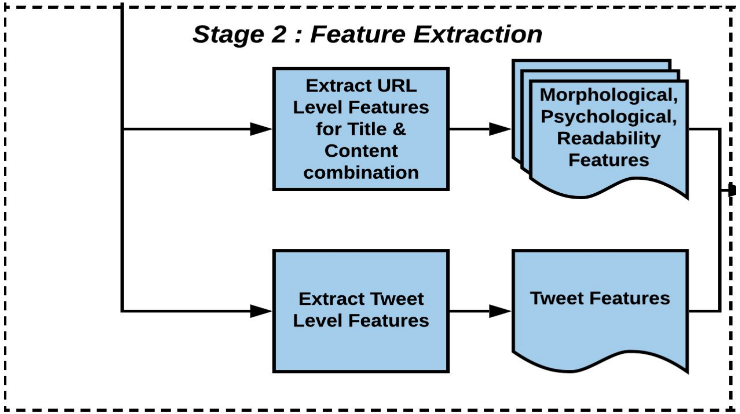
- The Raw Corpus from a GitHub repository with URLs tagged under various categories such as Fake, Conspiracy, Neutral, Reliable, etc. ***We selected only Fake and Reliable.***
- Tweet details for selected URLs extracted into CSV files.
- Refined Corpus: Final selection of 11000 URLs with equal split of Fake (25 web domains) vs Reliable (35 web domains)



Stage 2 : Feature Extraction

STAATLICH ANERKANNTE HOCHSCHULE

- Linguistic and Twitter features extracted for all 11,000 URLs into flat files.
- Linguistic Features:
  - Full Text used = Title + Content extracted for each URL.
  - **Morphological:** Extracted via spaCy library's part-of-speech tagger. Total 50 features.
  - **Psychological:** Used Linguistic Inquiry and Word Count software (LIWC, Version 1.3.1 2015). Total 93 features.
  - **Readability:** Via TextSTAT and NLTK libraries. Total 18 features.



- **Twitter** Features:
  - Extracted Tweets containing the URLs into flat files.
  - Processed flat files to create total 22 features

Stage 3 : Classification Models

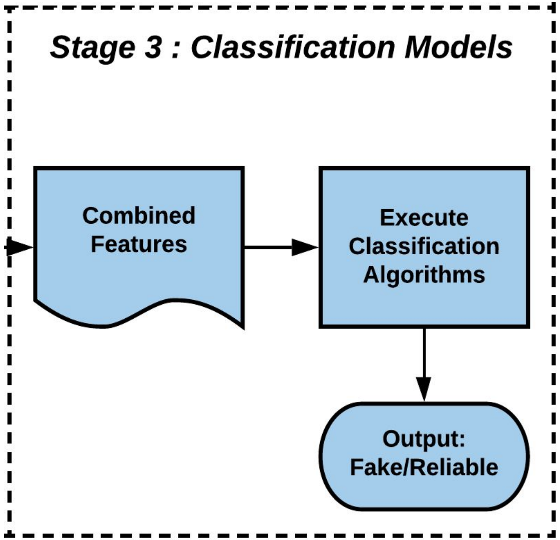
STAATLICH
ANERKANNTE
HOCHSCHULE

- All features combined at URL level in a flat-file. Total 183 features.
- Feature selection using scikit-learn Random Forest: **reduced to 64**.

	Count of features	Count of effective features
Morphological Features	50	24
Psychological Features	93	24
Readability Features	18	7
Tweet Features	22	9
Total	**183**	**64**

- Classifiers used:
  Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), Decision Trees and XGBoost.

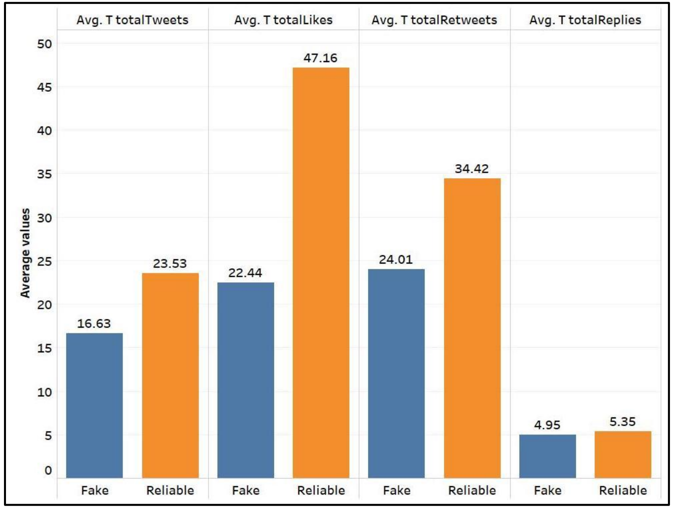
- These were implemented on a combination of all as well as on the selective set of features to compare the model accuracy.

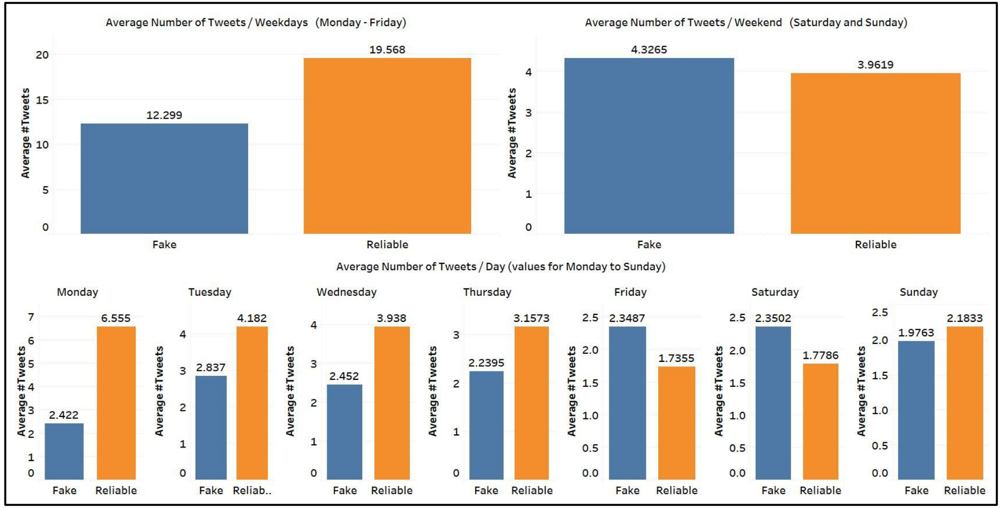


Result

STAATLICH
ANERKANNTE
HOCHSCHULE

- It is seen that the values of Twitter features for Reliable news is generally greater than that of fake news
- The only exception being the Average number of tweets/weekend feature, wherein Fake news has a higher value than that of Reliable news





Result

STAATLICH
ANERKANNTE
HOCHSCHULE

- There is not much difference between the accuracy of classification models run without and with feature selection.
- XGBoost provided the best accuracy for both the scenarios whereas KNN provides the least accuracy.

Without Feature Selection - Classification results (accuracies)

	DT (E)	DT (G)	KNN	RF	SVM	LR	XGBoost
ALL	0.84	0.85	0.71	0.90	0.88	0.88	0.93
L	0.82	0.82	0.70	0.88	0.84	0.83	0.90
M	0.78	0.78	0.77	0.87	0.80	0.79	0.86
R	0.70	0.71	0.65	0.77	0.72	0.73	0.76
T	0.68	0.69	0.68	0.72	0.65	0.63	0.76
MLR	0.84	0.83	0.71	0.91	0.89	0.90	0.92

With Feature Selection - Classification results (accuracies)

	DT (E)	DT (G)	KNN	RF	SVM	LR	XGBoost
ALL	0.85	0.84	0.70	0.90	0.85	0.87	0.93
L	0.82	0.82	0.72	0.87	0.83	0.81	0.89
M	0.76	0.77	0.75	0.83	0.77	0.76	0.84
R	0.69	0.66	0.67	0.75	0.71	0.71	0.72
T	0.66	0.68	0.69	0.72	0.62	0.61	0.74
MLR	0.84	0.83	0.70	0.90	0.86	0.85	0.91

Data
All: All features
L: Psychological
M: Morphological
R: Readability
T: Twitter

Model
DT(G / E): Decision Tree - Entropy / Gini
KNN: K Nearest Neighbors
RF: Random Forest
SVM: Support Vector Machine
LR: Logistic Regression
XG Boost: Extreme Gradient Boost

Conclusion

STAATLICH ANERKANNTE HOCHSCHULE

- In the context of analysis of the content of URLs, one can ignore identification of topics and any author related characteristics. Therefore, a topic and author agnostic approach holds water.

- Identifying fakeness based purely on Psychological group features consistently gives very high accuracy as compared to other groups taken individually viz. Morphological, Readability and Twitter

- Accuracy of models using only Twitter group features used by us is the least. Maybe if more Twitter features are used, the accuracy may have improved

- XGBoost consistently gives the best accuracy

Future Scope

Possible approaches to enhance the accuracy of fake news detection:

- Since the news published on the web page is taken as input, the web markup features can also be added to classification model

- For Twitter features: extract additional features related to Twitter User, media content, etc.

- The fake news detection process can be extended to detection in many other languages other than English

STAATLICH
ANERKANNTE
HOCHSCHULE

Demo

Demo Run

STAATLICH
ANERKANNTE
HOCHSCHULE

1) Using a single Fake URL as input data:
   a) File has 5 columns: Domain, Url, UrlType, Title, Content
   b) Twint used to first extract tweets
   c) Create 4 files: temp_M_features.csv, temp_L_features.csv, temp_T_features.csv, temp_R_features.csv
   d) Use above 4 files to create one file with all features combined.
      i) This could be either for training a model, or
      ii) As data to use for a saved model to predict
2) Random Forest classifier:
   a) Using all 11,000 URLs as input data with all features (183)
   b) Validation data of approx. 1000 data points - used to check accuracy of trained model
   c) Remaining 10k data points split in Train: Test = 80:20 ratio.

STAATLICH
ANERKANNTE
HOCHSCHULE

Thank You!