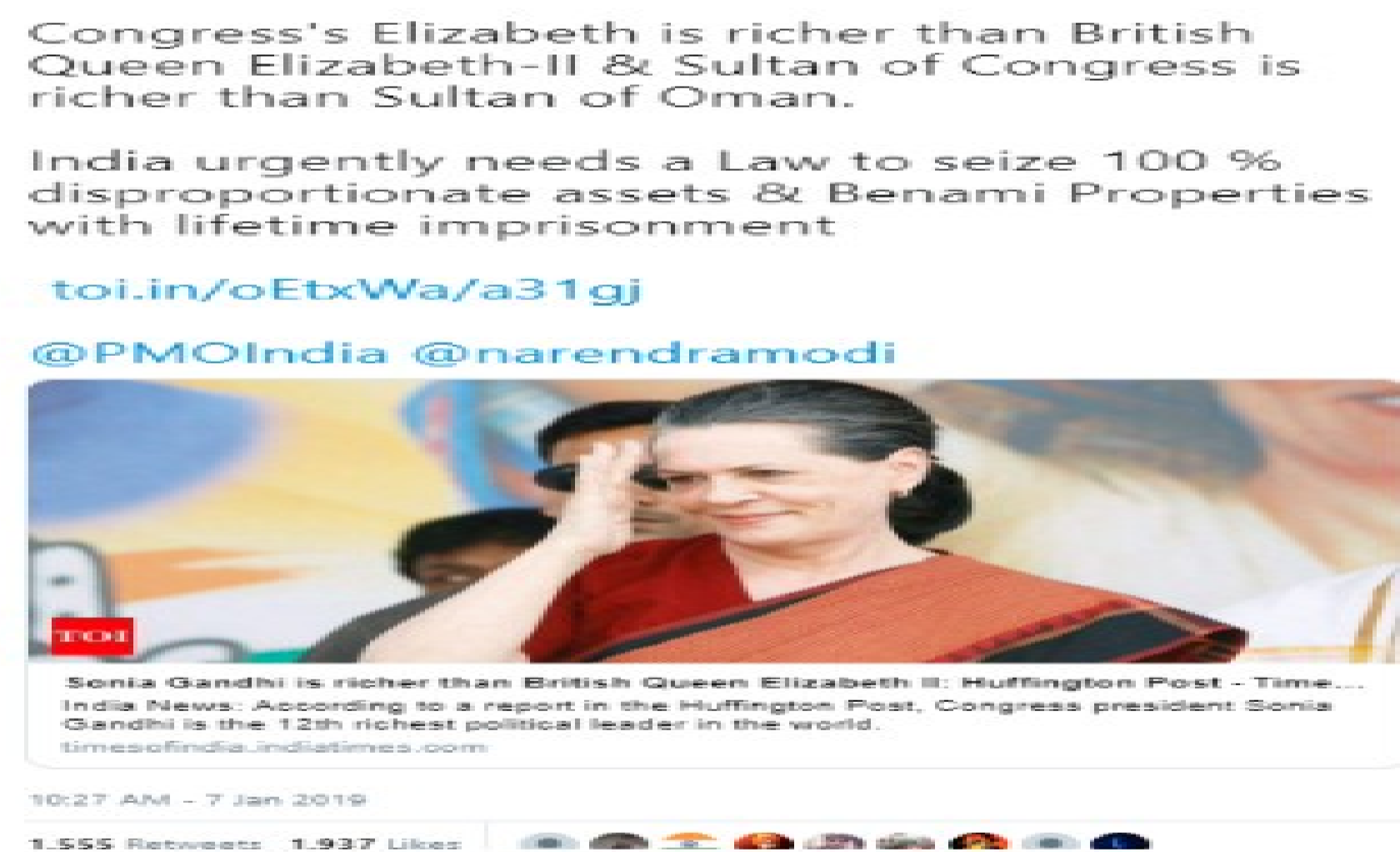# Fake News Detection

INTRODUCTION

- Fake news consist of news, stories, shares, messages, forwards or hoaxes created to deliberately misinform or deceive readers or receivers.
- Major sources of fake news are Web content, Social Media i.e Twitter, Facebook, Mobile Apps like WhatsApp, Instagram.
- Fake Content can be categorized into
  - Clickbait
  - Propaganda
  - Misleading Headings
  - Biased News
  - Inaccurate Journalism

- Fact Check: Viral post claiming Sonia Gandhi richer than Britain's Queen Elizabeth II is false

TASK DESCRIPTION

We participated in QICC (Qatar International Cybersecurity Contest) which attempts to detect fake news.There are 3 sub-tasks related to Fake News in QICC contest. Task 1A aims for systems to predict whether a piece of text is fake or not. It also aims to identify the domain of the same text piece.
The domain included Politics,Business,Sports,Entertainment,Technology,Education.The objective of Task 1B is to identify whether a given Arabic tweet is by a bot or not. Task 2 aims at creating of an annotated fake news dataset where the task is to submit pieces of text with fake or legit labels.

APPROACH

We tried majorly two kinds of approaches for subtasks 1A and 1B.
1. Machine Learning Techniques
   - Voting Classifiers with Logistic Regression,Naive Bayes, Adaboost
2. Neural Network Approaches
   - Fast Text

- For Track 2, we collected English Text mainly in the Indian subcontinent content.
- We collected 0.71K pieces of text from the web, used different semi-automatic labeling functions through SNORKEL to get the expected annotations.

CORPUS DETAILS

Language	Task	Training Samples	Test Samples
English	1A	384	48
Arabic	1B	18384	1615

Language	Task	Legit	Fake
English	2	0.4K	0.3K

RESULTS

Task	Model	Features	F1
1A-Fake	LR+NB+AB	Word 1-2+Char 2-5	0.83
	Fast Text	Word Embeddings	0.79
1A-Domain	LR+NB+AB	Word 1-2+Char 2-5	0.83
	Fast Text	Word Embeddings	0.75
1B	LR+NB+AB	Word 1-2+Char 2-5	0.95
	Fast Text	Word Embeddings	0.94

Authors:Vandan Mujadia, Pruthwik Mishra, Dipti Misra Sharma   Research Center Name:MT&NLP Lab, LTRC