# Term Project Proposal Presentation

## CS533 Information Retrieval Systems

## 1/9

# Detecting Fake Tweets and Misinformation

Mestan Fırat Çeliktuğ,
firat.celiktug@bilkent.edu.tr

## 1. DESCRIPTION OF THE PROBLEM/OPPORTUNITIES

- **Ease of dissemination of an idea(w.r.t. previous times)**
- **Online Social Networks(OSN): Medium for propagation of almost any type of information**

## 1. DESCRIPTION OF THE PROBLEM / MAIN PROBLEMS&FOCUS

### PROBLEMS

- Information validity
- Identification of right/false, real/fake, information/misinformation tweets

### FOCUS

- Detection of «fake/not-fake» tweets

## 2. MOTIVATIONS

- Little help for preventing time loss
- Little help for preventing wrong perception management

- There could be more beneficial motivations not written here

- Twitter is widely used medium to achieve those goals
- Relatively high importance for public&personal benefit

## 3. METHODOLOGY/TWO MAIN STEPS

# STEPS

- **1.** Collection of set of tweets(related to selected set of events)

  &

  Labelling a part of the whole tweet set (event basis)

- **2.** Classification of the tweets as "fake or not", "misinformation or not", "rumor or not"

NOTE: Specified problem in the scope of "fake tweet detection" would be determined w.r.t. decided data set and complexity of the problem.

## 3. METHODOLOGY/SECOND STEP(BINARY CLASSIFICATION)

- Second step is simply an example of "Binary Classification" problem.
  ( Information Retrieval & Especially Machine Learning Job )

- POSSIBLE MACHINE LEARNING MODELS TO BE USED

. Artificial Neural Network
. Support Vector Machine
. Decision Trees ( Especially, J48 Decision Tree is said to give (very) high accuracy in one "Tweet Credibility Evaluation" Task )

- POSSIBLE TOOL

. Weka ( Java Based Open Source Machine Learning Tool )

## 4. EXPECTED RESULTS

- Statistically significant accuracy of tweet classification (as "fake or not", "misinformation or not", "rumor or not")
- Succeeding in at least average literature results (Average success in prior work validation)

## 5. REFERENCES

[1] Benevenuto, Fabricio, et al. "Detecting spammers on twitter." Collaboration, electronic messaging, anti-abuse andspam conference (CEAS). Vol. 6. 2010.
[2] Rajdev, Meet. "Fake and Spam Messages: Detecting Misinformation During Natural Disasters on Social Media."(2015).
[3] Mohanraj, V. "A Survey on Spam Detection in Twitter." International Journal of Computer Science and BusinessInformatics 14.1 (2014).
[4] Martinez-Romo, Juan, and Lourdes Araujo. "Detecting malicious tweets in trending topics using a statisticalanalysis of language." Expert Systems with Applications 40.8 (2013): 2992-3000.
[5] Kumar, Arun, and Sandeep Kumar. "Twitter Spamming: Techniques And Defence Approaches." InternationalJournal of Applied Engineering Research 7.11 (2012): 9-13.