

FINAL

PROJECT PROPOSALS

ELECTION DAY TWEETS – SENTIMENT ANALYSIS AND TRENDS

THE PROBLEM – HOW DID PEOPLE ON TWITTER REACT TO THE ELECTION?

- ▶ What does sentiment analysis of tweets referencing the election (or a candidate) from election day reveal?
 - ▶ did this change over time?
 - ▶ Was one candidate referenced in an overwhelmingly negative or positive way?
 - ▶ Can any state-specific analysis be made?
- ▶ What demographic data can be gleaned? - Predict male or female from tweets?
- ▶ Most common phrases for referring to each candidate?

THE DATA – TWITTER DATA FROM ELECTION DAY

- ▶ Tweets containing: text, handle, favorite count, retweet count, date, time, original author, geo or lat/long
- ▶ Data set available from Kaggle datasets
- ▶ Clinton and Trump tweets (for comparison)
 - ▶ <https://www.kaggle.com/benhamner/clinton-trump-tweets>
- ▶ Sample of all tweets from Election day
 - ▶ <https://www.kaggle.com/kinguistics/election-day-tweets>

THE HYPOTHESIS – SENTIMENTS VARY THROUGH TIME, BY DEMOGRAPHICS

- ▶ Sentiment analysis will vary over time on election day
- ▶ Sentiment analysis and common phrases (text characteristics) can be used to identify subject of tweet (Hillary Clinton or Donald Trump)
- ▶ Sentiment analysis and common phrases (text characteristics) can be used to identify origin of tweet by either Donald or Hillary

STUDENT SUCCESS IN HIGHER ED

THE PROBLEM – WHICH COLLEGE STUDENTS NEED ASSISTANCE?

- ▶ Many higher education students will struggle to complete courses and fail to get a degree
- ▶ They could be helped, but schools do not have infinite resources to devote to these students, so they must be identified
- ▶ Schools often have demographic data and past academic performance data that could be used to identify students who would benefit from assistance

THE DATA – DEMOGRAPHIC AND ACADEMIC CHARACTERISTICS

- ▶ Ideally, data with both demographic and past academic performance information
- ▶ Examples:
 - ▶ first generation college student or not
 - ▶ high school GPA
 - ▶ past class grades
 - ▶ age
 - ▶ class enrolled in
- ▶ (still looking for an appropriate dataset)

THE HYPOTHESIS – STUDENT FACTORS CAN PREDICT SUCCESS

- ▶ Student demographic and prior academic information can be used to predict if students are more or less likely to need assistance in a class
- ▶ (or alternatively, we can predict if students are likely to drop or fail a class, or fail to graduate, based on academic and demographic data)

PREDICTING EMPLOYEE ACCESS RIGHTS

THE PROBLEM – EMPLOYEE ACCESS RIGHTS

- ▶ Employee access rights are a cumbersome problem for a company's IT department or staff to deal with
- ▶ Companies would like an automated way to determine if an employee should be granted access to a particular resource based on which resource it is and characteristics of the employee

DATA – AMAZON EMPLOYEE ACCESS DATA

- ▶ Features: Resource (to be gained access to or not), employee characteristics - company role, manager id, title, department
- ▶ Response variable: Action (binary) - was employee granted access to the resource or not
- ▶ Available from Kaggle competition that closed 2 years ago
- ▶ <https://www.kaggle.com/c/amazon-employee-access-challenge/data>

THE HYPOTHESIS -

- ▶ Employee information including manager, role, department, and resource information can be used to predict whether or not an employee should be granted access to a particular resource.