

California State University San Bernardino  
School of Computer Science and Engineering

**CSE 595 Independent Study Presentation**

**Date**

June 8, 2016

**Time**

11:00 AM – 11:30 AM

**Place**

JB-359

**Title**

Analyzing 2016 Presidential Primary Twitter Data Using Hadoop

**Student**

Mark Baltes

**Advisor**

Kay Zemoudeh

**Abstract**

Twitter has become an increasingly important tool for those seeking public office in the United States. Never before have candidates been given such direct access to the public to make statements or break news. Similarly, never before has the public had direct access to a candidate to tell him/her what they feel about them or subjects they are addressing. Being able to analyze this new communication dynamic is crucial for the candidates or anyone interested in finding out where the subset of the population that uses Twitter stands on the candidates and the issues they discuss. This project accomplishes the analysis by utilizing the Twitter API to build a dataset of tweets relevant to the 2016 Presidential Primaries, storing that dataset in the cloud using Amazon Web Services and using Hadoop to run analysis on the large dataset.