

SØREN KOLBYE JENSEN
Bsc. Ha(IT)
#43124133123
December 1, 2017

Contents

1 Summary 1

2 Introduction 2

2.1 Case introduction 2

3 Problem Formulation and Reserach questions 3

4 Theoretical Framework 3

5 Methodology 3

5.1 Data Acquisition and Dataset Description 5

1 Summary

Summary section

2 Introduction

Big Data has changed how data is viewed in society, businesses have invested heavily in infrastructure that can improve data collection. It is now possible, to use Big Data to analyze the behaviour of Social Media channels such as Twitter, Facebook and Instagram. However, the vast amount of data available with modern technologies has proven difficult to analyze. According to ??, Big Data can be either structured or unstructured, where the determining parameters are "Velocity, Veracity and Validity". In order to surmount this difficulty, new datamining technologies have been invented, in order to aid the field of Data Science.

These principles are called "Datascience and Datamining". According to (INSERT BOOK), datascience and datamining can be defined as follows:

Data science is the act of using fundamental principles, to guide extraction of knowledge from datasets.

Datamining on the other hand, deals with extracting knowledge from data using technologies, which incorporates the principles of data science.

This research paper, aims to use the fundamental principles of Data science, in conjunction with datamining technologies. To analyze the behaviour of U.S president Donald J. Trump. and how this behaviour affected his political campaign, as well as how his behaviour affected the American financial markets.

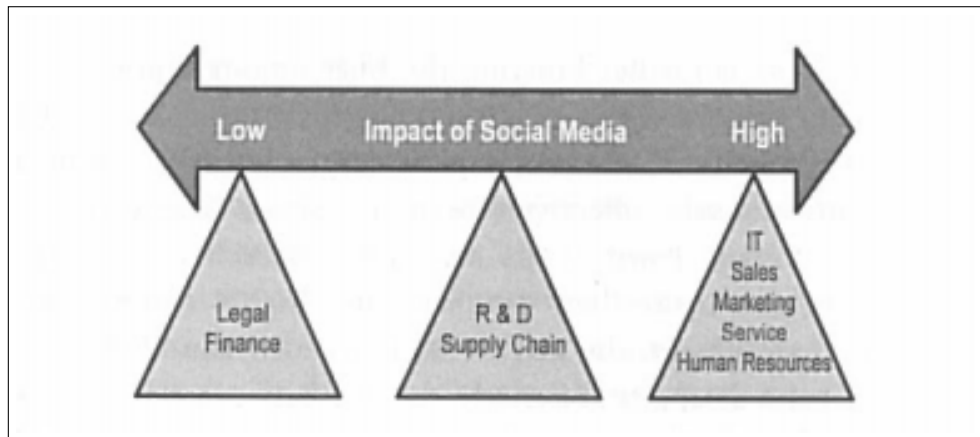


Figure 2: real, local caption for reference

2.1 Case introduction

Donald John Trump, born in 1946 in Queens New York, is the CEO of the Trump Organization and is well known as a real-estate broker in the United States. However, on June 16th, 2015, Trump officially announced his candidacy as the 20th US President". This political campaign was heavily involved with Social Media, as Trump was actively fighting big news channels, such as CNN, NBC etc. labelling them as "fake news". Thus, Trump figured out a way to distribute highly controversial statements and gain political support by his use of Social Media Channels such as Twitter.

3 Problem Formulation and Reserach questions

The goal of this research paper is to analyze trumps behaviour on Twitter and how he used Twitter to combat "fake news". A visual analysis on his behaviour between DATE and DATE will be conducted to reveal the patterns in his behaviour in this timeframe. This leads to the reserach question:

What was trump's behaviour on social media, prior to his announcement of his presidency. And how did it change, during and after his political campaign.

Furthermore, Trump's political campaign was heavily focused on combating illegal immigration. One of the primary targets of Trump's statement was Mexico. This leads to the research question:

How did Trump's political statements on Twitter affect the American and Meixcan currency?

4 Theoretical Framework

For this research paper, an usupervised data mining technique has been chosen. With this method there are no pretetermined attributes or a specific target. This research aims to use induction and deduction to find correlations among data. Thus revealing a structure, which was otherwise hidden.

Clustering maybe? When distributing fake news and stuff? idk

(Reveal groups and categories, previously unaware of) -¿ Further datamining discover new correlations

Association: Co-occurences among events. -¿ Example, Trump becomes president and the currency falls, why?

Asociation rules "50 percent of the activity is negative?"

Feature extraction: If violense, action, heroism is an attribute of a film then "action" may be trhe feature. You can say that news can have t he feature fake or real?

5 Methodology

To answer the research questions two different types of analysis is presented

Furthermore, the CRISP framework has been used in a modified fashion, in order to better illustrate the context of this research paper.

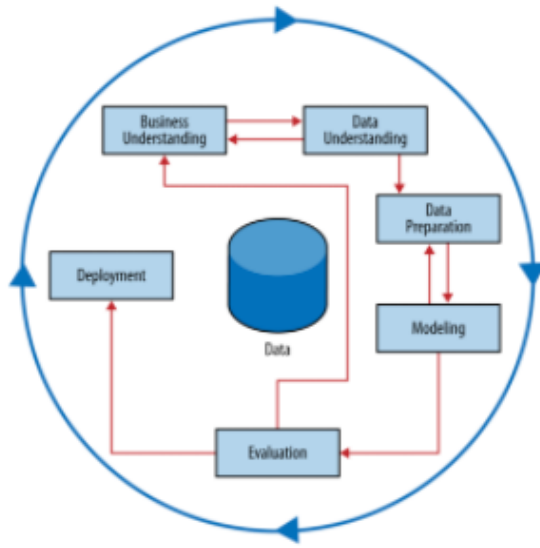


Figure 3: real, local caption for refrence

The model serves to illustrate how the data was processed throughout the proeject in order to yield a final re-
sult/answer to the resarch questions. The model is iterative and serves to explore the Twitter data, so a better
understanding of the data can be reached. This process has been conciously iterated throughout the project.
"Business understanding" is in this context the overall problem formulation and research questions that we wish
answered.
For every iteration a deeper understanding of the data is achieved, which affects the understanding of the overall
research framework.
"Data preperation" is where we merge, cleanse and integrate multiple data sources. In this research paper, Alteryx
has been used as shown in Table 1 later in the report. "Modelling has used Tableau, which is a visual analytics
tool. This makes it easier to explore and understand the patterns presented in the data.
"Evaluation" Here we evaluate, whether the data model that has been built answers the overall research questions.
After the model has been evaluated and validated, the model will go through "deployment", which in this sense
will be the final conclusion to the research questions.

To summarize, Table 1 shows the purpose and use of each datamining tool used in this research paper:

Table 1: My caption

Tool:	Purpose:	Use:
Alteryx	Data Analytics tool that provide data mining solutions by helping with data	Lorem ipsumLorem ipsumLorem ipsumLorem ipsum Lorem ipsum Lorem ipsumLorem ipsumLorem ipsum
Tableau	This tool helps with visual data representation.	Lorem ipsumLorem ipsumLore m iLorem ipsumpsumdasdLo remLorem ipsum ipsum

5.1 Data Acquisition and Dataset Description

The datasets in this research paper has been collected from Trump's Twitter page, ranging from 2009 to 2017.

The first datasets consists of Twitter data on Trump's primary twitter channel. This dataset will mainly be used to analyze Trump and his Campaign team's behaviour.

The second dataset consists of Historical data on the American and Mexican currency between Trump's announcement to 14/07/2017

The third dataset is collected through Sentione and will be used to analyze the sentiment towards Trump's campaign. Ranging from before his inauguration, until 02/11/2017.