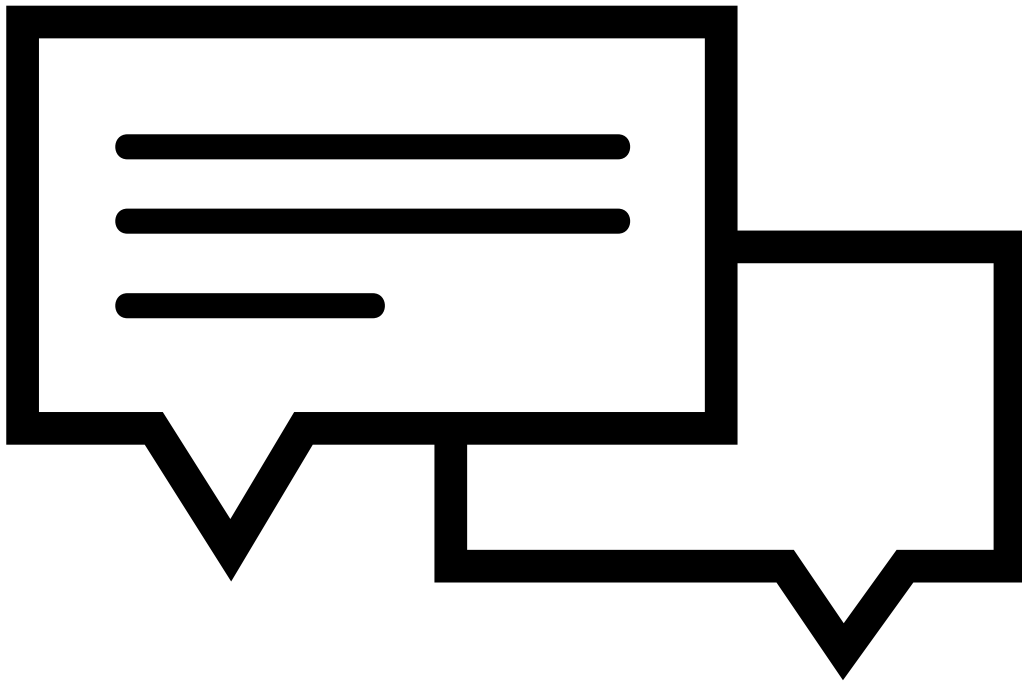


MBD OCT 2018 - TERM III,  
GROUP C, SECTION 0-1

PROFESSOR IDAFEN SANTANA  
PEREZ



# **NATURAL LANGUAGE PROCESSING & TEXT MINING**

**GROUP FINAL ASSIGNMENT**

JUNE 28TH 2019

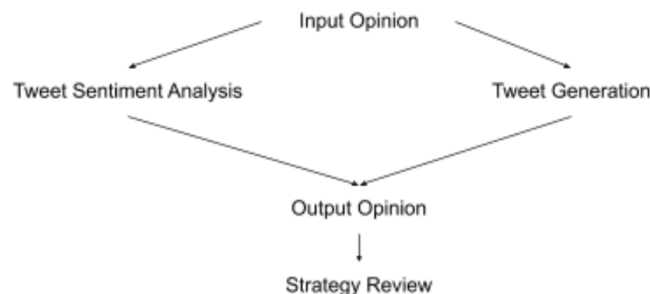
CAMILLE BLAIN-COALLIER  
IRUNE MAURY ARRUE  
JUAN PABLO GARCIA GOMEZ  
DEMETRIS PERDIKOS  
NAYLA FAKHOURY  
RAFAEL HERNANDEZ  
ALFONSO BUGAG

## **Business Proposal - Application**

During government elections, speeches and debates are always at the heart of news articles. More than that, speeches are a way to convey a candidate's message to their audience, giving them reason to elect their candidate. Campaign organizers need to strategically think about their audience to get the most value out of their candidate's speeches. For that, there is a formula for the right speech. "A persuasive speech takes time. Candidates can use friendly audiences as a proving ground for introductions, key points, conclusions, and more subtle elements of a stump speech." (Throughlinegroup) In addition, the speech needs to be given at the right time and with the right speaking skills.

However, gaining further insights on your audience is the key tool to create better speeches. Today, most people advocate their opinions and feelings on social media, especially through Twitter's platform. Our purpose is to understand the overall and aggregated view of these people during the 2016 Grand Old Party (GOP) Debate. We propose to gain insights in two steps; a quantitative and a qualitative approach. First, we will understand/predict the overall sentiment of our audience through tweeted data, scoring the actual sentiment model. Second, we will predict the future tweets regarding some topics. With both tools combined, we can really understand the level of sentiment (output opinion) for a certain person or a topic. This will allow campaign organizers to understand the effects of certain speeches and topics and also, allowing them to proactively review their strategy in the short-term/long-term, if need be. Modifying the direction or content of a speech (for a candidate) is an example of a campaign reviewing and taking action to realign their strategy. This will in turn, increase the candidate's chances at winning the elections.

Steps	Output	Reason	Python Notebook
Tweet Sentiment Analysis	Quantitative	How do they feel?	1_Sentiment_Analysis_Final
Tweet Text Generation	Qualitative	What do they say?	2_Tweet_Generation_RNN
Speech Text Generation	Qualitative	What should I say now?	3_Text_Generation



In addition, we've completed this analysis while comparing two methods. We have used the traditional NLP approach (taught in class) as well as the Deep Learning approaches (Neural Network and Recurrent Neural Network) to understand what models and techniques generate the best results for all instances; sentiment analysis and text generation.

The reason why we decided to approach the same problem with different techniques (standard versus deep learning) is because “for a long time, the majority of the methods used to study NLP problems employed shallow machine learning models and time-consuming, hand-crafted features. This lead to problems such as the curse of dimensionality since linguistic information was represented with sparse representations (high-dimensional features). However, with the recent popularity and success of word embeddings (low dimensional, distributed representations), neural-based models have achieved superior results on various language-related tasks as compared to traditional machine learning models like SVM or logistic regression.”(Medium)

With the latest advancements in deep learning, we can foresee that many of the traditional tasks of natural language processing will become easily implemented through these new tools. “Text Generation is one such task which can be architected using deep learning models, particularly Recurrent Neural Networks.” (Medium)

## **Results**

The results obtained in our analysis all seem to support one argument: the standard NLP approach for analyzing sentiment analysis and text generation, for this dataset, was more successful than the neural network approaches. With higher scores in accuracy for the sentiment analysis, the standard NLP approach also seems to be generating Tweets and speeches that are more coherent to the average user.

However, this approach may be oversimplified when conducting more complicated topics and datasets. We are aware that deep learning methods, like RNN, are powerful tools at extending the capacity of normal neural-based NLP models. Overall, combining all these powerful techniques will provide better tools to fully understanding the complexity of languages.

## **Conclusion**

In conclusion, we have used various Natural Language Processing techniques to solve a specific question: during presidential elections, could candidates be better equipped with tools that would allow them to make better marketing campaigns? What if candidates had better insights on the impact of their speeches and could modify their texts, if need be, with those insights?

Now, even though this analysis was completed for the 2016 GOP Debate, the same approach can be taken for other business problems, which include, the launch of a new product. By performing a Tweet sentiment analysis, the company can understand the consumers’ opinion. Then, when generating Tweets, the companies will be able to predict some keywords assigned to certain topics. This again, will allow people to adjust their strategy (for example: speeches) to better reflect the needs of their consumers, hence, increase their customer lifetime value in the long run.

## Sourced Information

1. The Right Speech Formula For Winning Political Campaigns.  
<https://www.throughlinegroup.com/2012/07/22/the-right-speech-formula-for-winning-political-campaigns/>
2. Deep Learning For NLP: An Overview Of Recent Trends. <https://medium.com/dair-ai/deep-learning-for-nlp-an-overview-of-recent-trends-d0d8f40a776d>
3. Language Modelling and Text Generation Using LSTMs – Deep Learning.  
<https://medium.com/@shivambansal36/language-modelling-text-generation-using-lstms-deep-learning-for-nlp-ed36b224b275>