Yusiqi Wang

Professor Diego Klabjan

Data Science and Analytics

March 18, 2020

# Executive Summary

**Process Taken:**

This study aims to develop a Question and Answer Systems which quickly gain information from daily business news for 2013 and 2014. The QA system was built on Elasticsearch in order to generate answers to the following questions:

• Which companies went bankrupt in month X of year Y?

• What affects GDP? What percentage of drop or increase is associated with this property?

• Who is the CEO of company X?

The process should be classifying question, querying the Elasticsearch, finding answer (keyword) related to the question, and combining the keywords in answer and question to appear as the final result.

**Report for Text Analysis**

# Data Preprocessing

# Ergodic the TXT file. Set stop words. Establish the Elasticsearch and insert the article to Elasticsearch sentence by sentence.

# Raise Entities

# Extract entity to find the Keywords in the question. Use the nltk.word\_tokenize ,pos\_tag and ne\_chunk to raise entities

# 手机屏幕截图 描述已自动生成

# Use compile transfer string to Unicode, then transform the questions.

手机屏幕截图

描述已自动生成

# Classify Questions

# Use cosine distance to measure similarity among questions. Classify questions of different types to determine the corresponding model.

# Bankrupt questions corresponds to question type 0. GDP questions corresponds to question type 1. CEO questions corresponds to question type 3.

手机屏幕截图

描述已自动生成

# Define Functions to answer questions

# Build three separate functions to answer each type of the questions. In each function, we define rules to determine which article and sentence have the highest score to our questions.

手机屏幕截图

描述已自动生成

社交网站的手机截图

描述已自动生成

# Business insight

# For the Q&A System part, if we can increase the size of our corpus and add function in our model, we can quickly get our answer for our problem, which can increase our business efficiency. And it is easy for us to know economic Condition for whole society or financial situation for some Companies.