Application of text miner in web debate on president Trump

1 Introduction

Donald J. Trump won the president election. The supporters are happy to have a new president. But the antics are sad. Protest happened in several states. In this paper, text miner is applied to analyze the opinions of US citizen from website. The training data and score data are randomly collected from Donald Trump for President and Stop Trump respectively, by half-half, in Facebook website. It may not represent the population. It is just for academic study.

2 Variable Description

TargetField: Y represents support Donald J. Trump; N represents dislike him. C represents central stand.
TextField: The words posted by the Facebook users in Donald Trump for president and Stop Trump.

3 Method

Text mining is applied in this paper. Text mining, or text data mining, is text analytics which is a statistics method of deriving hi-quality information from text. Several statistical learning models can be applied in text mining. In this paper, classification tree is utilized. The goal is to turn text into data for analysis. In this paper, our interest is to distinguish text whether they support or antic the new president. The data are run under SAS enterprise miner. The software automatically cleans the text, removing meaningless words and punctuations, and retrieves informatic word via Natural Language Processing(NLP). 199 data are predefined by the author.

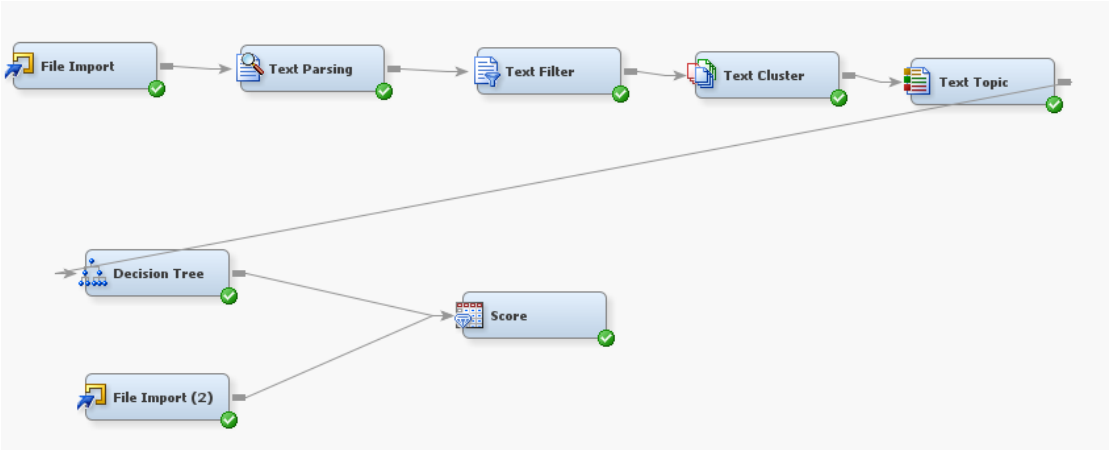

Figure 3.1 Diagram of nodes

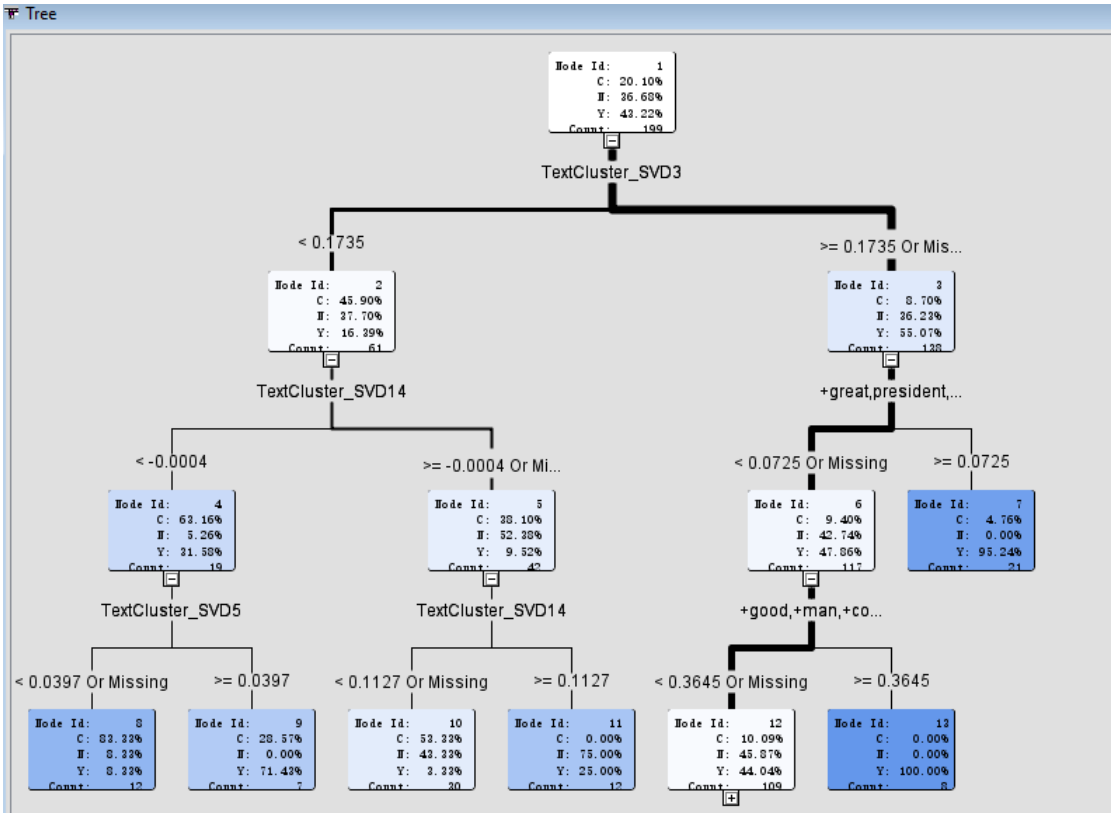

Figure 3.2 Classification Tree

4 Discussion

The result shows that 49 percent of sample users are supporters. Only 20 percent are antics. 31 percent are central stand. It seems that supporters are the major and that most of the antics start to question their protest activity. It is reasonable because the election result has been announced for three weeks. People are going to calm down and waiting for new day. Additionally, advanced classifications are made by the author, this may not be objective enough. The result may not

represent the reality. It is only for academic study in text mining.

Data Role=SCORE Output Type=CLASSIFICATION

Variable	Numeric Value	Formatted Value	Frequency Count	Percent
I_TargetField	.	C	31	31
I_TargetField	.	N	20	20
I_TargetField	.	Y	49	49

Table 4.1 output of score for new data.