Automated Essay Grading

A Kaggle Challenge

kaggle

The Learning Agency Lab

E. WASSERMAN

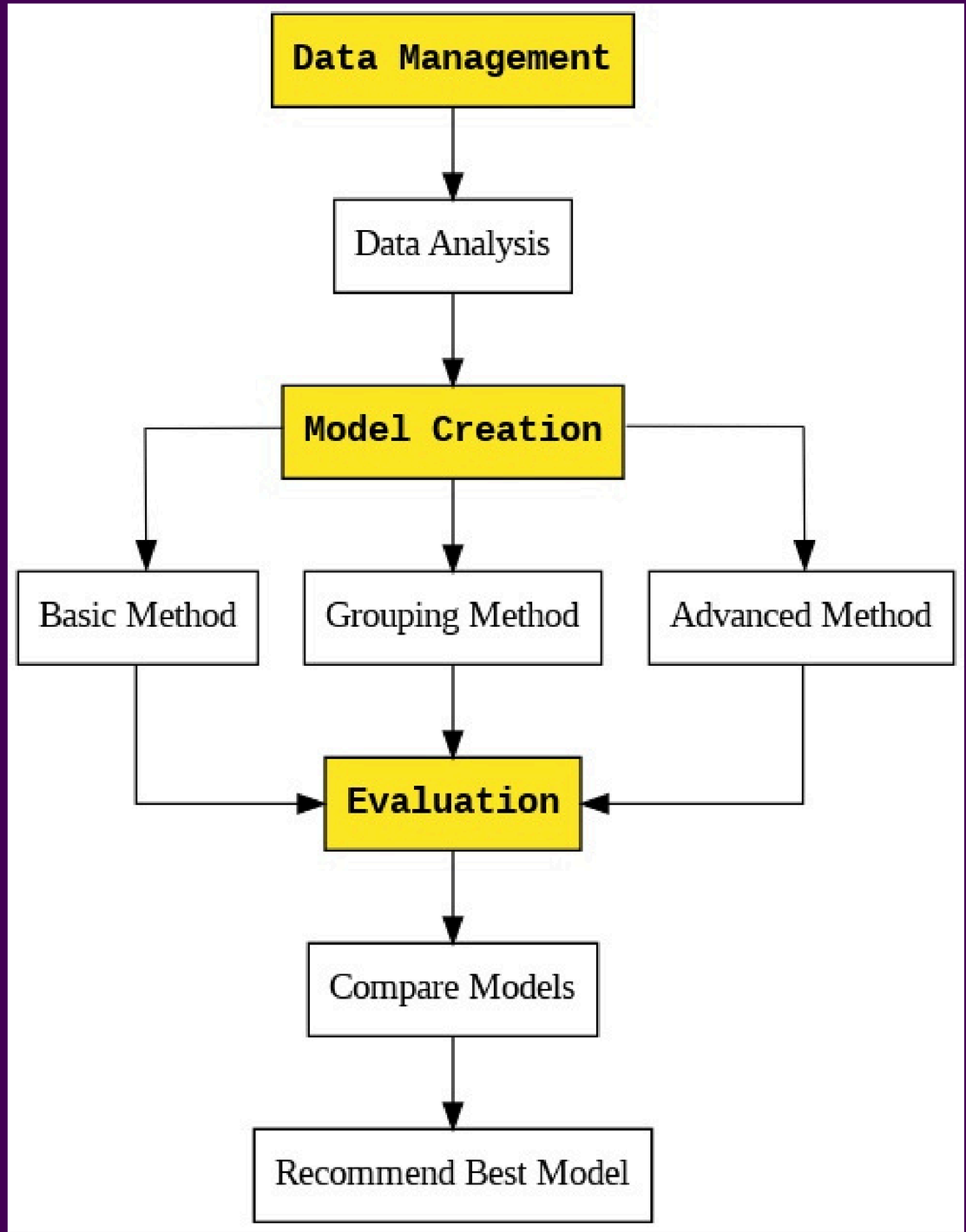
BUSINESS PROBLEM

1. What are the optimal test settings for the models?
2. How does the model build impact grading accuracy?

DATA UNDERSTANDING

- Source: Kaggle (open commons)
- 6,000 argumentative essays
- Essay scores range: 1 to 6

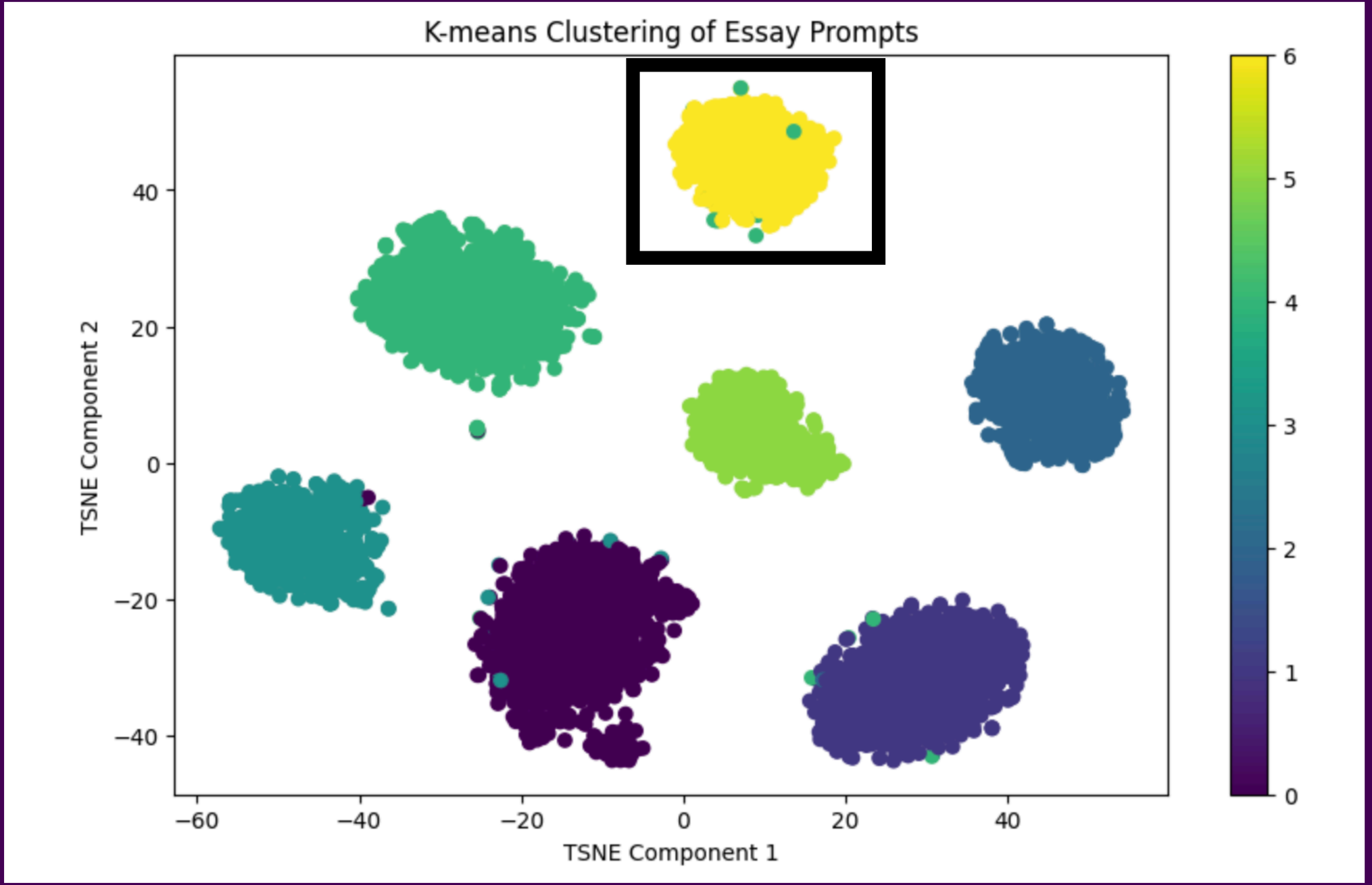
METHODOLOGY



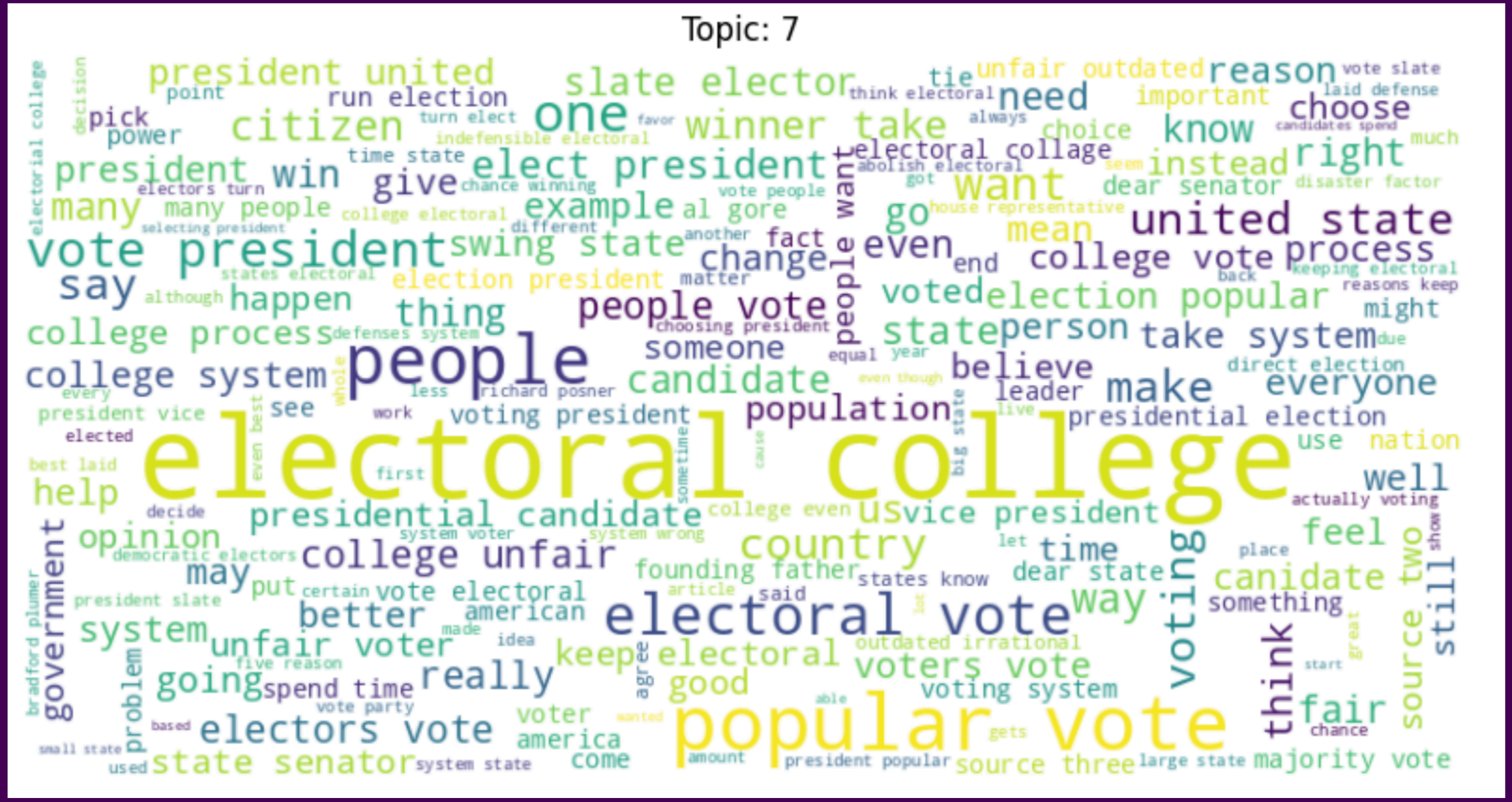
RESULTS

Vector	Model	Sample Size	Train Accuracy	Test Accuracy
TF-IDF	Neural Network	6000	88.5%	41.4%
TF-IDF	K-Means	6000	100.0%	41.2%
BERT	BERT	2000	84.6%	52.8%

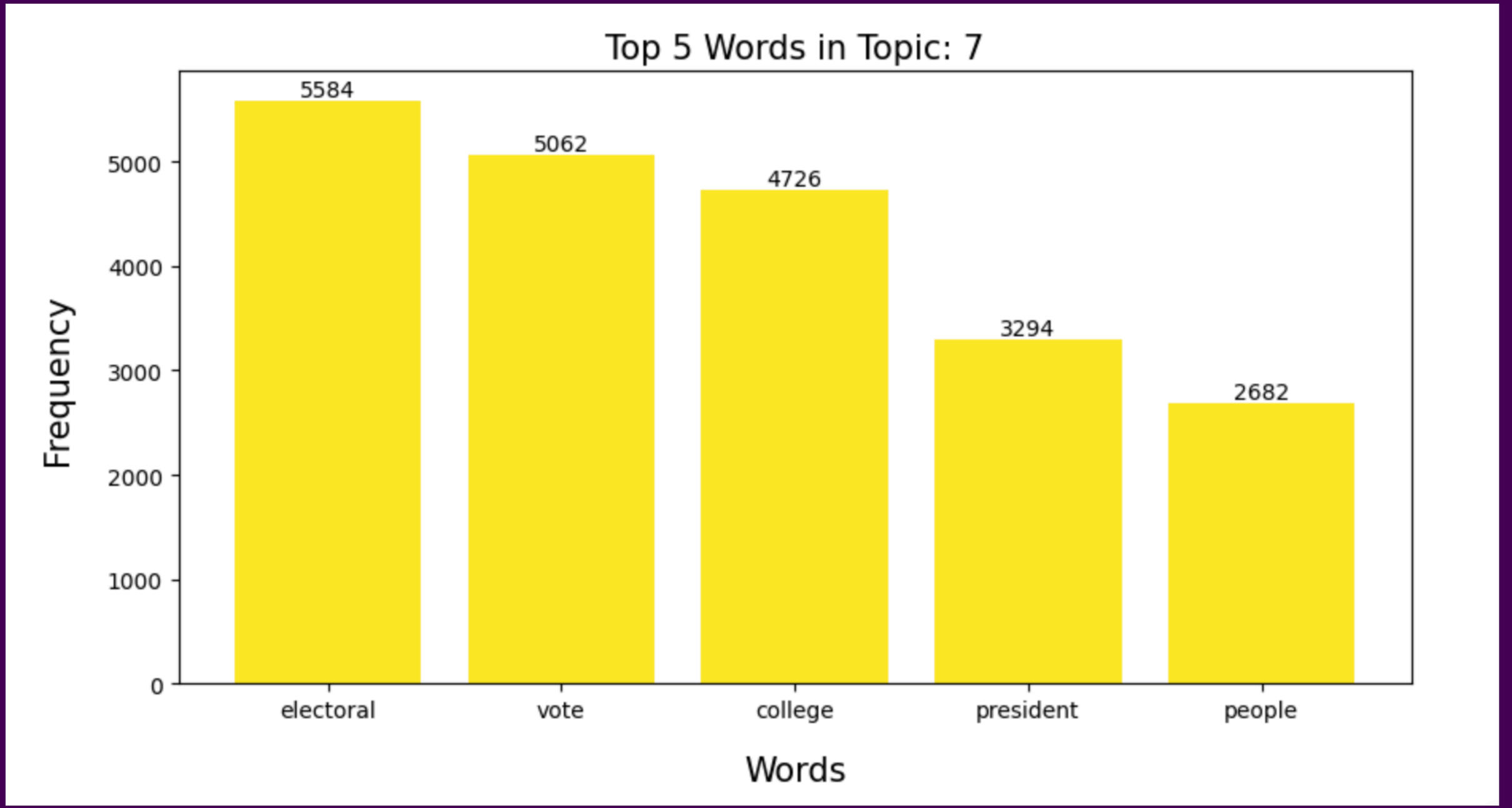
GROUPING TOPICS



ESSAY 7 WORD CLOUD



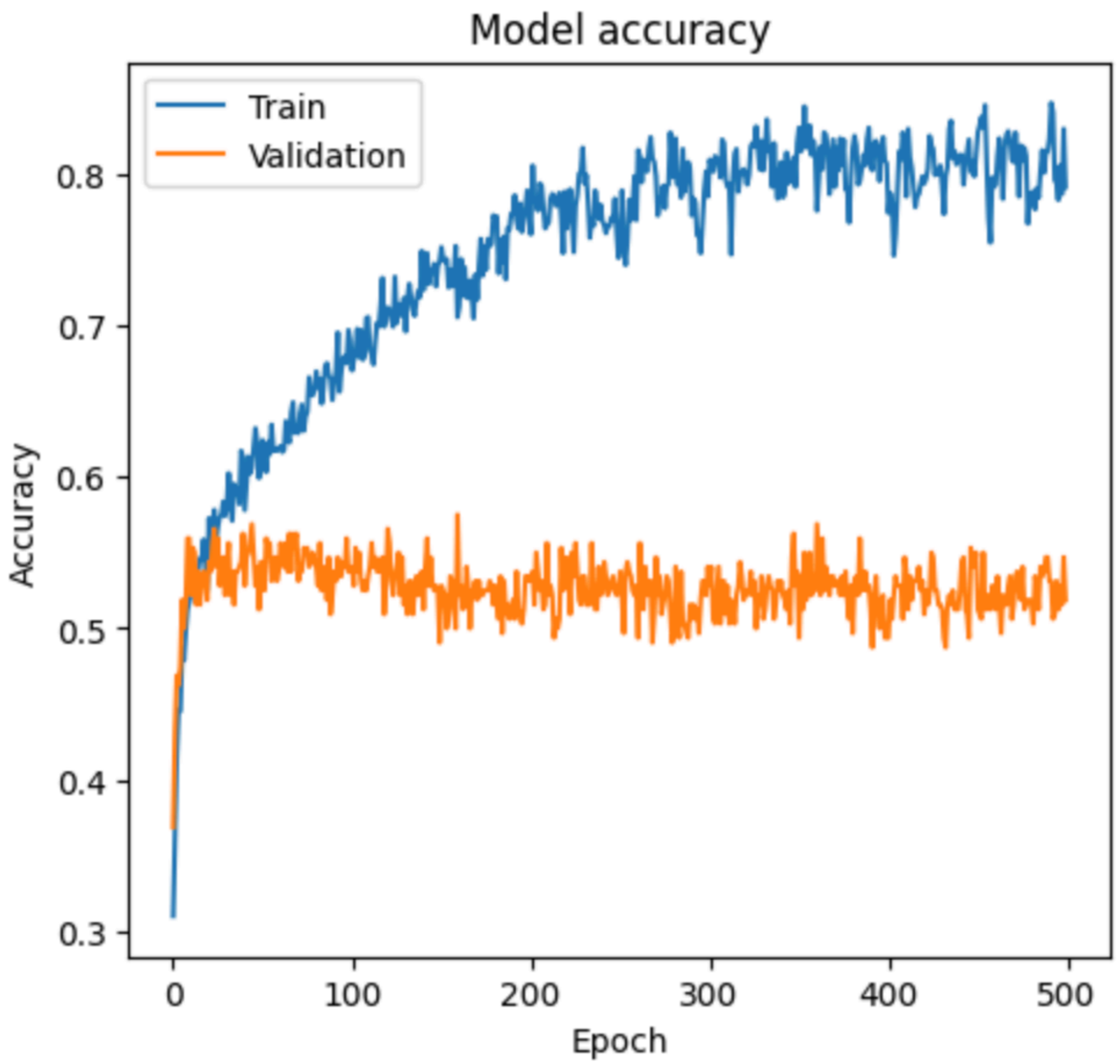
ESSAY 7 TOP WORDS

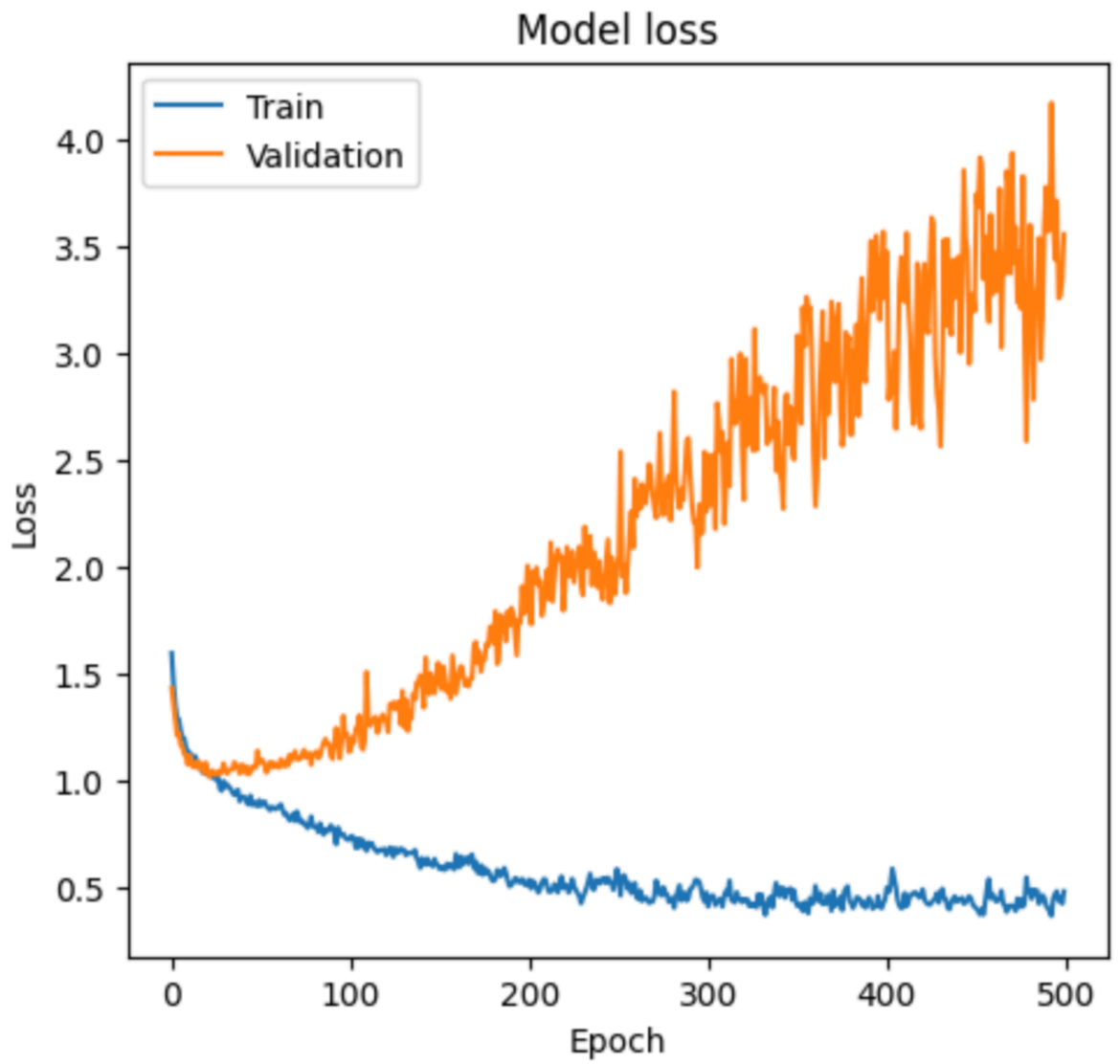


RESULTS

Vector	Model	Sample Size	Train Accuracy	Test Accuracy
TF-IDF	Neural Network	6000	88.5%	41.4%
TF-IDF	K-Means	6000	100.0%	41.2%
BERT	BERT	2000	84.6%	52.8%

BEST MODEL


Model accuracy
Train
Validation
Accuracy
0.8
0.7
0.6
0.5
0.4
0.3
0
100
200
300
400
500
Epoch


Model loss
Train
Validation
Loss
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
100
200
300
400
500
Epoch

RECOMMENDATIONS

- Adopt the BERT-Based Mode
- Address Overfitting in Neural Network Models
- Increase Training Data Size

LIMITATIONS

- Inconsistent Grading
- Limited Dataset Variety
- Bias in the Model

NEXT STEPS

- Make the Model Bigger and Smarter
- Train the Model Longer and Better
- Enhance the Model’s Understanding

THANK YOU

QUESTION AND ANSWER