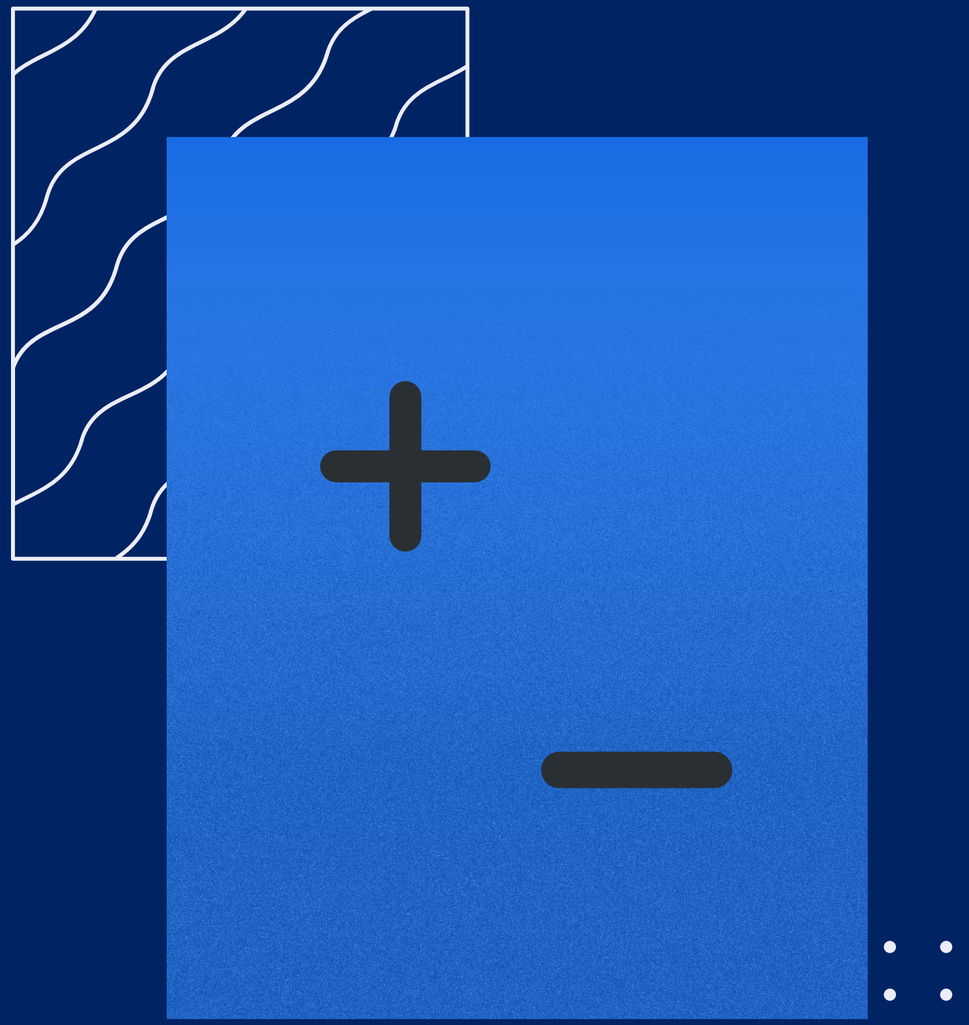
Analyzing Sentiment through Natural Language Processing

USING DATA FROM TWITTER



Agenda

- Business Objective
- Data Background
- Methods
- Results
- Recommendations

Objective

Builld a model that can accurately classify the sentiment (positive, negative, neutral) of a tweet based on its contents.

The resulting model will be used to classify tweets from future tech conferences and analyze how sentiment changes from event to event.

DATA BACKGROUND

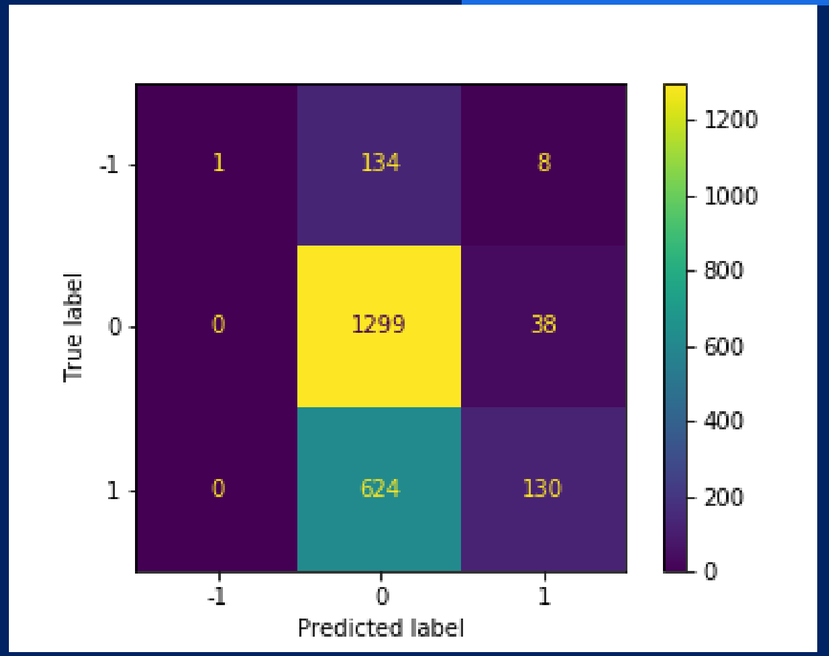
- Sourced from Twitter from SXSW 2013
- Contains tweets about the SXSW tech events and product announcements
- Most of the tweets are about Apple and Google, and many mention specific products
- Manually labeled for classification

METHODOLOGIES

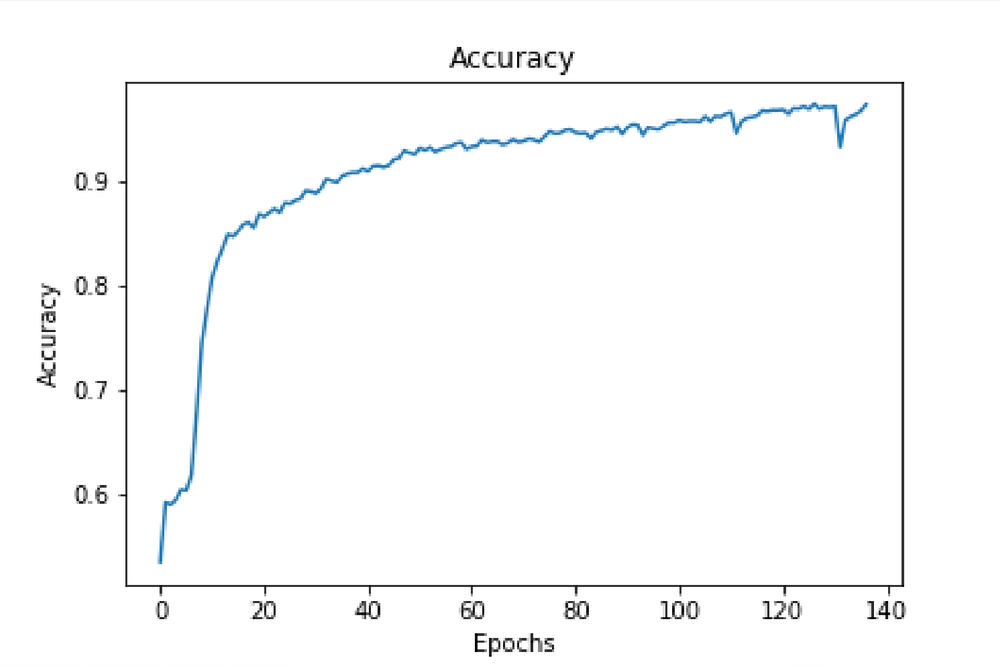
In order to process text data, it was converted into a format that can be read by ML models.

Several different methodologies were evaluated on their accuracy. Most had trouble predicting negative or positive sentiment.

A more complex model architecture, a neural net, was built out using a model that was pre-trained on an immense amount of text to understand word representations.

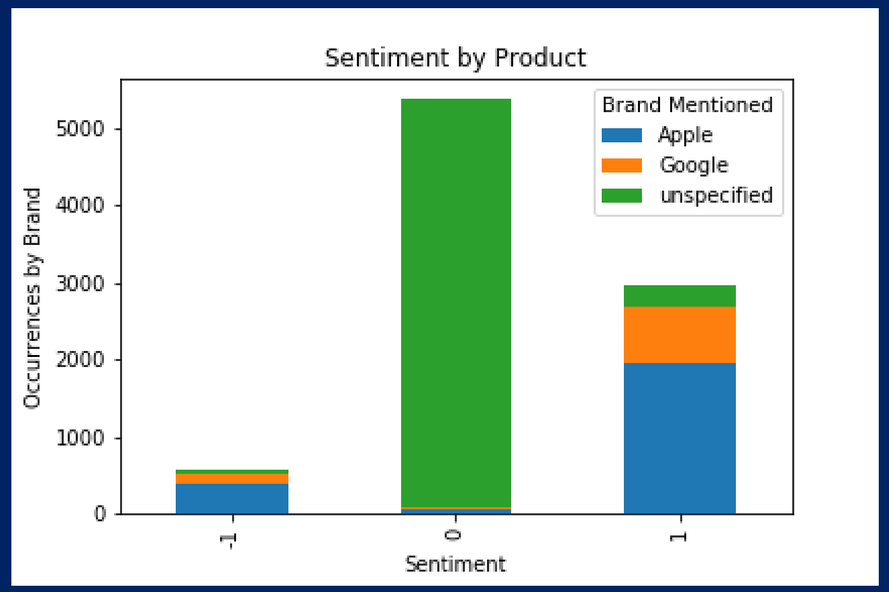


RESULTS



- The neural nets using pre-trained word embeddings achieved high accuracy at 98% on the data it was trained on
- However, it struggled when tested with a 64% accuracy score.
- All model types struggled with test data but this was the best performer overall.

RECOMMENDATIONS/FUTURE WORK



Move forward with neural net model to evaluate sentiment around future events and conferences related to tech as it has learned off of a specified data set.

In order to improve this model and make in more generalizable, it is recommended to gather more data that has a higher percentage of labels that indicate positive or negative sentiment.

Adding more data could also allow the use of a simpler model, which could help to reduce runtimes.

THANK YOU!