Executive Report: Interactive Analytics Web Application for Sentiment Analysis

Our team has developed an interactive analytics web application using Streamlit and Google Cloud to analyze sentiment in Reddit posts. Our application is designed to handle real-time data, making use of Google Cloud resources for data collection and machine learning workflows. The application has been successfully deployed on Streamlit Cloud.

Model and Dataset:

The model was trained on a Kaggle dataset containing over 35,000 Reddit posts. The dataset underwent preprocessing, including text cleaning, tokenization, and lemmatization. The SentimentIntensityAnalyzer from the NLTK library was utilized to calculate sentiment scores, resulting in a training accuracy of .98, and a testing accuracy of .91.

We tested the model on a dataset of almost one thousand Reddit posts from r/NBA, achieving an accuracy of approximately 83%, and an F1 score of about .9. This demonstrates the model's effectiveness in analyzing sentiment on the real-world data that we are planning on collecting using Google Cloud.

Google Cloud Integration:

We utilized Google Cloud resources to automate data collection and machine learning workflows. The Streamlit app reads data directly from the Firestore database.

Streamlit and Streamilt Cloud:

Our team created a streamlit app that showed our model’s training and testing performance on the intial training dataset we used. There is also a slider that allows the user to change the test/train split and see the impact on the model’s performance.

Our app also collects data from the Google Firestore and shows our model’s current accuracy and F1 scores on the up-to-date reddit posts. Our app also contains two wordclouds, of words that are most common in negative posts and words most common in postitive posts. Finally, we have a feature where the user can input a team or player name and will be shown the current number of positive and negative posts for that player or team, from the database of posts from our Google Cloud storage.

Conclusion:

Our team successfully developed an interactive analytics web application that effectively analyzes sentiment in Reddit posts using Streamlit and Google Cloud. This application demonstrates our team's ability to create valuable solutions for real-world analytics problems, and we believe it has great potential for various applications in business and beyond.