



Social Media Sentiment Analysis

Austin Megalaitis, Owen Arnone, Mihir Khadka, Justin Nem

CEPS, University of New Hampshire, Durham, NH 03824

Fidelity Sponsors: Allen Detmer, Collette Powers, Navleen Singh



Introduction

- Social media can affect behavior around technical and financial fields, such as how last year social media posts had a large effect on Gamestop's stock price.
- Fidelity is aware of the growing influence of social media, and wishes to understand the overall sentiment about their mobile virtual assistant services and chat-bot services.
 - To achieve this goal, Fidelity wishes to perform Sentiment Analysis on publicly available social media posts.
- The goal of the project is to design and implement a sentiment model to identify trends and metrics to better understand Fidelity's customers, and their opinions on Fidelity products.
- This project should be able to give Fidelity a reasonable idea of how users are feeling about their product. It should identify any deficiencies or complaints so they can make adjustments to fit customer needs if necessary.
- When the user visits the dashboard, they should be able to understand the overall sentiment for their desired social media platform within 5 minutes.

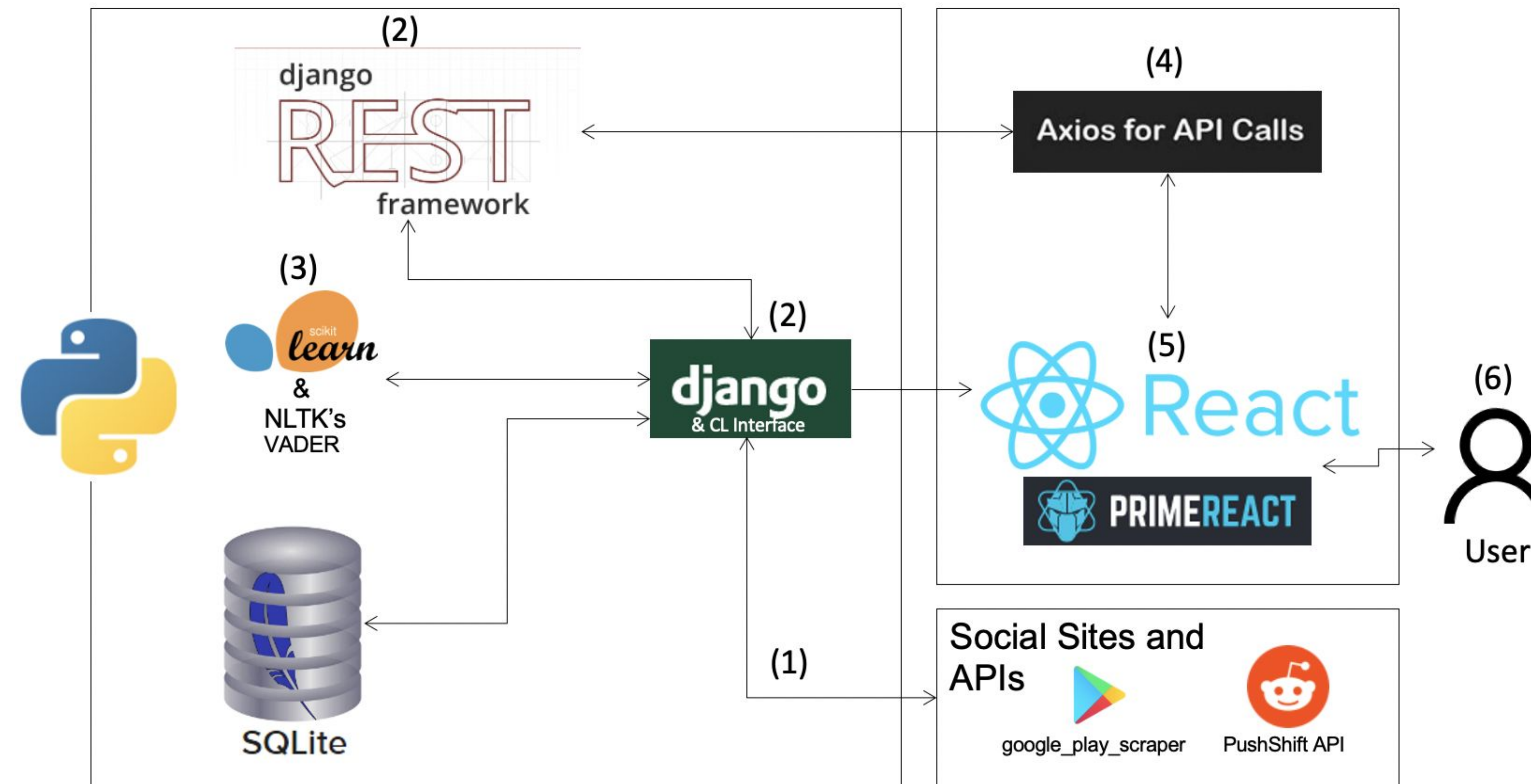
Requirements

The following features support the solution:

- The backend is able to scrape large amounts of data about Fidelity's virtual assistant and chat-bot services using public API's.
- After collecting 5,000 to 10,000 relevant entries, it is stored in a database.
- Create or use an existing machine learning model to measure accuracy or efficiency of the dataset.
- The product is able to take the entries stored in a database, and use a sentiment analysis tool to find the sentiment of the data.
- The product displays overall sentiment to a front-end dashboard.
- The system and the output for the web page must be organized to allow for ease of access for the customer, and the data in the web portal should be visualized in a meaningful way that the client can see and interpret the results easily.

Software Design Diagram

Django Back End - React Front End



1. Public API's are used for scraping text data from social media platforms.
2. Django models are used to store the data, and allow for ease of access. Django Rest endpoints used for access by webpage, Django Command Line Interface (CLI) used for executing backend applications.
3. The Python libraries NLTK and Scikit-Learn are used for machine learning sentiment analysis.
4. The web portal used axios to connect to the Django Rest endpoints.
5. React is used to build a single front end application to display data. PrimeReact is used in tandem to create a visually appealing website.
6. User interaction is through the frontend web portal.

Web Page

Figure 2: Sentiment Score Pie Chart

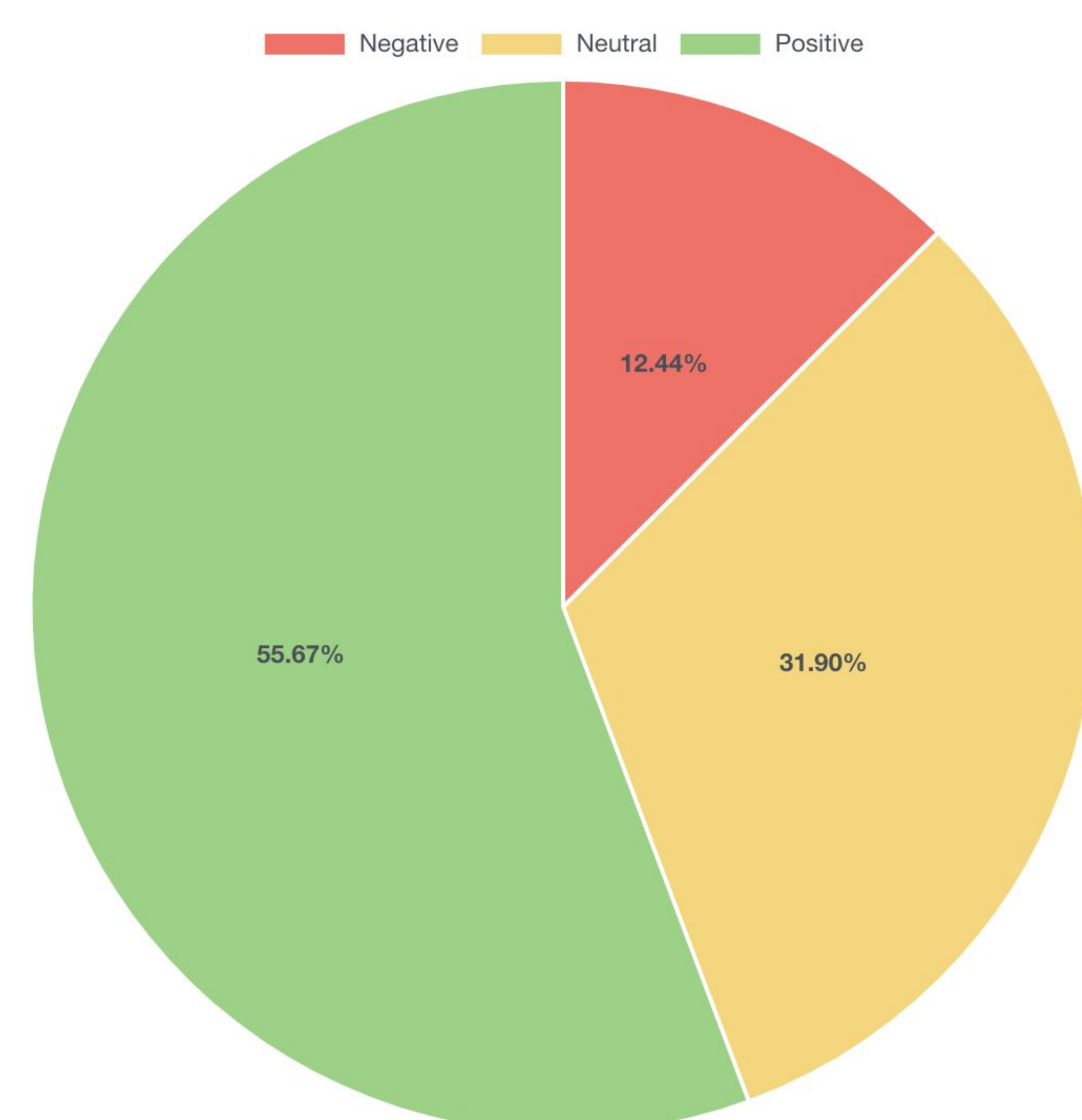
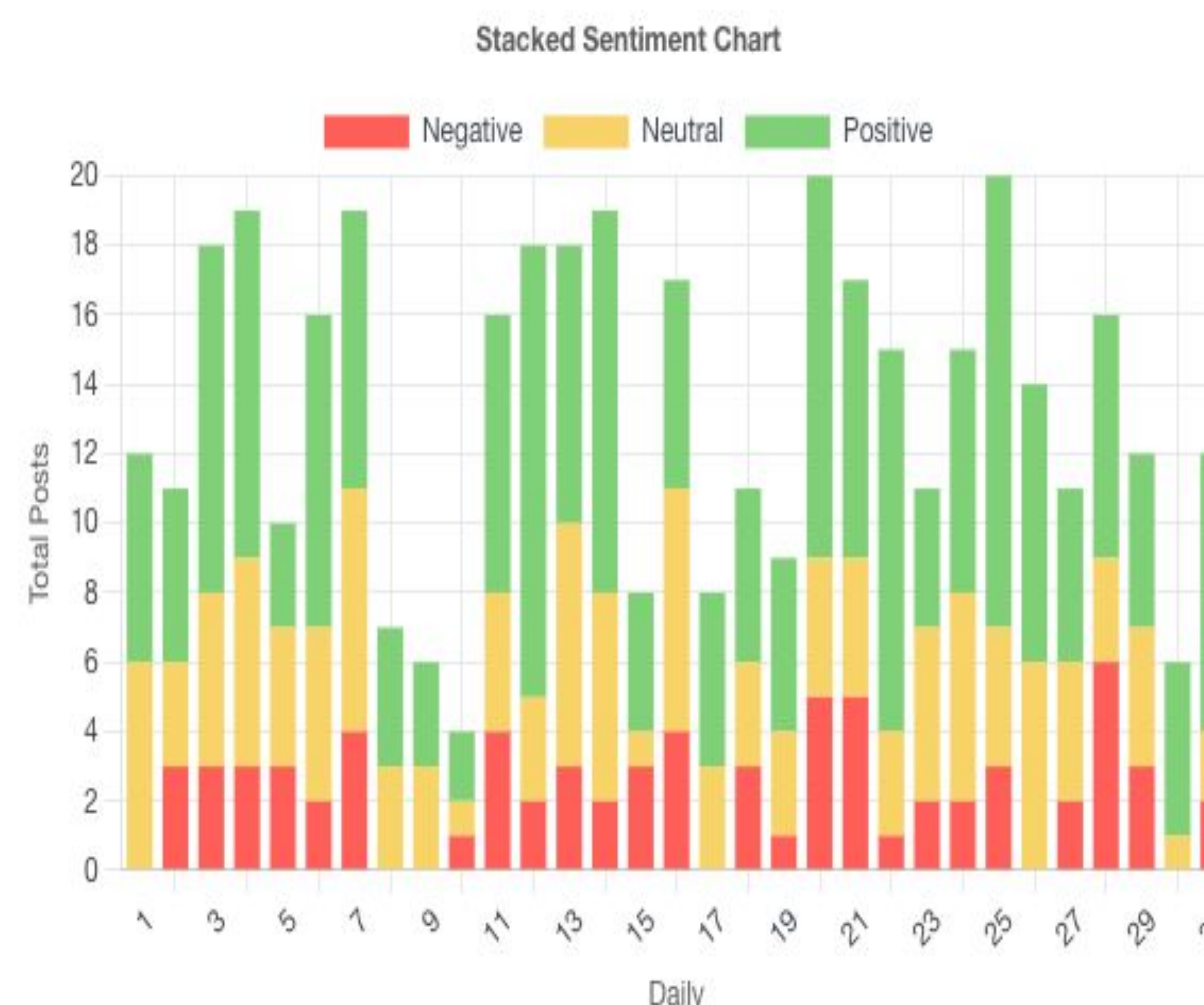


Figure 3: Stacked Bar Chart of Positive, Neutral and Negative Sentiment Scores



Results

Scikit-Learn & NLTK were implemented using the data collected from the scrapers APIs.

- Most common accuracy of Scikit-Learn machine learning model is around 88%, with a high of 90% and low of 84% based upon the random sampling.
- In a selected sampling, the recall was 94%, and precision was 90%.
- Conducted sentiment analysis using NLTK. End results are displayed in bar graph form to display how many reviews are positive, negative, and neutral.

The design and organization of the frontend is still minimal, with the majority of the focus going on the data analysis visualization.

Figure 4: Word Cloud



Next Steps

The team has made great strides with this project, but it could be expanded by doing the following:

- Add more layers to the Scikit-Learn model to refine, and improve accuracy. It is possible that different machine learning methods would have improved accuracy.
- Use alternate Machine Learning methods to classify and train the dataset, such as PyTorch and Tensorflow to observe differences in model accuracy.
- Explore more social media platforms such as Twitter to get more reviews about Fidelity products. Sentiment analysis could vary from platform to platform.
- Create more visuals to give the customer more ways to view and interpret the sentiment analysis data. These visuals will include a line graph, pie chart, and word cloud. Possibly another type of visual may be created for data visualization.
- Focus on the user friendliness of the website.