Analysis of Social Media Data

Q1: AI

- Can use a data set from Kaggle or build your own
- Eg: Sentiment140 dataset with 1.6 million tweets (kaggle.com)Links to an external site.
- Perform exploratory data analysis to understand the data and the distribution of the sentiments.
- Use ChatGPT Plus to generate tweets related to a topic or keyword, and then analyze the sentiment of those tweets.
- Compare the realism of the generated and real tweets using a metric called perplexity

Solution 1: The details to this Question are implemented and answered here: https://github.com/aditipatil0711/SJSU_Masters_Assignments/tree/d27353fc829db5588a099148563bd49a834f6e94/CMPE272_Enterprise_Software_Platforms/Assignment_6

Q2BI: What systems would you implement/setup to answer following questions? Explain the process in detail:

- What are the main topics or themes that people are talking about on social media related to your business or industry?
- How do people feel about your products, services, brand, competitors, or industry trends? What are the sentiment and emotions expressed in their posts?

Solution:

To answer these questions effectively, especially in the context of the FinTech domain, the domain we are working one, we would need to set up a system that combines social media monitoring, natural language processing (NLP), and sentiment analysis. Here's a detailed explanation on the systems and processes we consider implementing:

1. Social Media Monitoring and Data Collection
Tools and Systems:

- **Social Media APIs**: Use APIs from platforms like Twitter, Facebook, LinkedIn, etc., to collect posts, comments, and discussions about your business or industry.
- **Web Scraping Tools**: If API access is limited, consider web scraping tools to gather data from social media and forums.
- **Data Storage**: Set up a database to store collected data. Options include SQL databases like PostgreSQL or NoSQL databases like MongoDB.

Process:

- Determine Keywords: Identify keywords, hashtags, and phrases relevant to your business and industry (e.g., specific product names, company name, industry buzzwords).
- Data Collection: Regularly query social media APIs or scrape web pages for these keywords.

- Data Preprocessing: Clean the collected data for analysis. This includes removing irrelevant posts, correcting typos, and standardizing text format.

2. Natural Language Processing (NLP) for Topic Analysis
Tools and Systems:

- **NLP Libraries**: Use libraries like NLTK or SpaCy in Python for text processing and topic modeling.
- **Topic Modeling Algorithms**: Implement algorithms like Latent Dirichlet Allocation (LDA) or Non-negative Matrix Factorization (NMF) for extracting themes.

Process:

- Text Tokenization: Break down the content into individual words or phrases.
- Stopword Removal: Remove common words that add little semantic value.
- Topic Modeling: Apply LDA or NMF to identify clusters of words that represent the main topics being discussed.

3. Sentiment Analysis
Tools and Systems:

- **Sentiment Analysis Models**: Leverage pre-built models like VADER or build custom models using machine learning libraries like scikit-learn or TensorFlow.
- **Visualization Tools**: Use tools like Tableau or Matplotlib in Python for visualizing the results.

Process:

- Sentiment Classification: Use the model to classify the sentiment of each post as positive, negative, or neutral.
- Emotion Detection: For more nuanced insights, implement or use existing models to detect emotions like happiness, frustration, or disappointment.
- Aggregation and Visualization: Aggregate sentiment data to get an overall sense of public sentiment. Visualize trends over time or differences between products, services, or competitors.

4. Continuous Monitoring and Reporting
Tools and Systems:

- **Automated Scripts**: Set up automated scripts to regularly collect, process, and analyze new data.
- **Dashboard Tools**: Implement dashboard tools like Power BI or Google Data Studio for real-time monitoring.

Process:

- Automate Data Pipeline: Automate the data collection, processing, and analysis pipeline to regularly update insights.
- Real-time Dashboard: Develop a dashboard that showcases key metrics like dominant sentiment, trending topics, and comparative analysis between your brand and competitors.
- Regular Reporting: Set up a schedule for periodic in-depth analysis and reporting to stakeholders.

Conclusion for a FinTech Professional:
As someone in the FinTech domain, this approach allows you to tap into the wealth of unstructured data on social media to gain insights into customer perceptions, emerging trends, and competitive landscape. The combination of social media monitoring, NLP, and sentiment analysis offers a comprehensive view of public opinion and market dynamics, crucial for strategic decision-making in the fast-evolving FinTech industry. Implementing these systems will require a blend of technical setup (like coding and database management) and analytical skills (like data analysis and interpretation).