

In order to gain insight into the details of what people are talking on social media related to our business, we would set up a comprehensive BI system that integrates data collection, NLP, sentiment analysis and advanced data visualization techniques.

Detailed steps

- 1) **Data Collection** : The first step in the process would be collect data from the various social media platforms like Twitter, Facebook and Instagram. The data would be from the posts, comments, the shared post, likes and dislikes for posts. If there are APIs provided by the apps then we will use the APIs to get the relevant data. WebScraping techniques can be used on the sites where APIs do not provide sufficient access. In the data collection process we need to ensure that we comply with the privacy laws and the also Terms of Service for the platforms from which we are collecting data. We need to comply with the data protection regulations like GDPR, CCPA. We need to make sure that whatever data we have collected does not have any personal identifiable information without consent. We must maintain anonymity and privacy of the user.
- 2) **Data Storage** : Once we have collected the data we need to store it in a robust database like MySQL/PostgreSQL or a cloud based solution like Amazon RDS.
- 3) **Data Cleaning** : Once we have the data the next step would be perform Data Cleaning. This would involve removal of irrelevant content, duplicate entries, filtering out non-textual elements, handling missing values.
- 4) **Data Transformation** : This would be done to normalize and transform the data to a consistent format. This would include tokenizing, case normalization, removing special characters and stop words.
- 5) **Feature Engineering** : This would involve extraction of relevant features like text, hashtag, mentions or specific keyword that are related to our business.
- 6) **NLP for Topic Detection** : We would apply NLP techniques to analyze the textual data. We would implement topic modeling algorithms to identify the prevalent topics in social media. We would perform keyword analysis to extract and analyze frequently used terms using TF-IDF statistics
- 7) **Sentiment Analysis**: We would use a sentiment analysis model to categorize the post into sentiments like positive, negative or neutral. Algorithms like logistic regression, SVM can be used for building custom models.
- 8) **Data Visualization** : Data visualization tools like Tableau can be used to create insightful visualizations. This would include sentiment distribution over time, topic frequency maps, word clouds. We can also develop interactive dashboards which would allow data to be explored in real time. Tools like Apache Kafka can be used for real time data streaming
- 9) **Monitoring** : We would need to continuously monitor the performance of our NLP and sentiment analysis model.

