

Commonwealth_Sales(Forage)

Anonymization

- Remove the **customer_id** and **current_location** column. Removal of data (redaction) that does not have much informational value is a valid data privacy technique.
- Mask the **username** column to hide the real username.
- Replace the original name column with a fake name. Replacing real values for fake values is a valid data privacy technique.
- Mask the **email** column to hide the real email address.
- Add noise to the **date_registered** and **birthdate** columns to hide the real value. Adding noise protects the real value by adding random noise to the actual value.
- Categorise the **salary** and **age** columns into bins. This categorisation hides the original values and preserves the distribution.
- The **credit_card_provider** and **credit_card_expire** have been tokenised. This step converts the categorical value of the columns into a different random value while preserving the distribution.
- The **credit_card_number** and **credit_card_security_code** have been masked.
- The **employer** and **job** columns have also been tokenised to preserve the original distribution.
- The **residence** and **address** have been replaced with fake values.

How to do Twitter Sentiment Analysis?

Common steps to perform Twitter sentiment analysis include:

1. Sort Twitter data
2. Clean Twitter data
3. Develop Twitter sentiment analysis model

4. Analyze Twitter data for positive/negative sentiments
5. Visualize the insights

Sort Twitter Data

The first step in sentiment analysis is the collection and sorting of data. There is a sea of data on Twitter, it is important to pick the data that is most relevant to the problem you are looking to solve or the thing you wish to find out. Only relevant data can be used to train the sentiment analysis model and test whether the model performs satisfactorily on Twitter data. Another important aspect to cover is what type of tweets you are looking to analyze – historical or current. To sort this data, you first need to extract it from Twitter. For this, you can use some of the following platforms:

- Zapier to create an automated workflow between Twitter and Google Sheets, for example.
- IFTTT to collect Twitter data without any code.
- Export Tweet to track hashtags, keywords, etc. in real-time or look for historical tweets and mentions.
- Tweet Download to collect tweets from one's own account including the mentions and replies.
- Twitter API to access and analyze public tweets about keywords, brand mentions, hashtags, or tweets from particular people.
- Tweepy, the python library, for reaching Twitter API and gathering data from there.

Twitter Data Cleansing

Once the data is gathered and sorted, it then needs to be cleaned before it can be used to train the Twitter sentiment analysis model. Twitter data is mostly unstructured, so the cleaning process involves removing emojis, special characters, and unnecessary blank spaces. The process also includes getting rid of duplicate tweets, making format adjustments, and also removing very short tweets – those less than three characters. Clean data can give more accurate results.

Develop Twitter Sentiment Analysis Model

There are different Machine Learning platforms that can help one build and implement a Twitter sentiment analysis model. These platforms can provide access to pre-trained or ready-to-train models. You can use your Twitter data to train these models. For developing a model, you need to go through the following steps:

- Choose the type of model you want to build, for example, a classifier model that helps classify text into pre-defined tags.
- Decide the type of classification. In this case, it would be Sentiment Analysis.
- Import the relevant Twitter data to train the model.
- Tag the data as positive, negative, or neutral, for example, to train your model.
- Test your model.

Analyze Twitter Data for Sentiments

Once your model is trained and gives satisfactory test results, it is ready for deployment. Now, you simply need to connect your Twitter data with your sentiment analysis model. There are several ways to accomplish it. One way to do it is to analyze a particular file of new or unseen tweets and classify them. Another way is to integrate Twitter data with Zapier and Google sheets and analyze this data using your model.

Insight Visualization

There are tools that help visualize your data results and make them easy to interpret and digest. These attractive visualization tools such as Google Data Studio, Looker, Tableau, etc. create visual reports including charts, graphs, and data tables that are easily understood by a larger number of people.

ChatGPT Enhancing Twitter Sentiment Analysis

Social media platforms like Twitter serve as a goldmine of public opinion. Analyzing the sentiment behind tweets not only provides insights into individual reactions but also paints a broader picture of public perception. With the advent of advanced AI models like **ChatGPT**, Twitter sentiment analysis has reached new heights. Here's how ChatGPT is making a significant impact:

- **Contextual Understanding:** Twitter is a hub of concise, context-heavy conversations. ChatGPT excels in understanding the intricacies of language, including slang, idioms, and abbreviations commonly used on Twitter. Its ability to grasp contextual nuances ensures accurate sentiment interpretation, even in the midst of brief and cryptic tweets.
- **Emotion Decoding: Emojis,** GIFs, and emoticons are integral to Twitter communication. ChatGPT can decipher the emotional tone conveyed through these symbols, adding an extra layer of depth to sentiment analysis. Understanding not just the words but also the emotions expressed enhances the overall accuracy of the analysis.
- **Real-Time Engagement:** Twitter operates in real-time, with trends and topics changing swiftly. ChatGPT's capacity for instant responses enables businesses and researchers to engage with Twitter users dynamically. By interacting with users in real-time, ChatGPT gathers live feedback, enriching the sentiment analysis process with fresh, up-to-the-minute data.
- **Language Adaptability:** Twitter is a global platform where tweets are posted in numerous languages. ChatGPT's multilingual capabilities allow it to handle tweets in various languages, ensuring a comprehensive analysis regardless of the linguistic diversity of the tweets being examined.
- **Continuous Learning:** ChatGPT is continually learning from the vast array of conversations it encounters. This ongoing learning process enables it to adapt to evolving language trends, slang, and emerging terminologies on Twitter. Consequently, its analysis remains relevant and reflective of the current linguistic landscape.

By incorporating ChatGPT into Twitter sentiment analysis, businesses, researchers, and analysts gain access to a tool that not only understands the language intricacies of tweets but also decodes the underlying emotions. This deeper understanding not only improves the accuracy of sentiment analysis but also provides actionable insights for informed decision-making.