**Sentiment Analysis on Instagram Comments**

**Project Overview**

This project aims to perform sentiment analysis on Instagram comments to gain insights into user sentiments, detect spam and fake comments, evaluate emoji usage, and identify majority opinions about a post. The project uses a dataset containing fields such as comment\_id, created\_at, text, user\_id, and username. By leveraging Natural Language Processing (NLP) techniques, the project provides a comprehensive sentiment breakdown of user comments.

**Dataset Description**

The dataset includes the following fields:

* comment\_id: Unique identifier for each comment.
* created\_at: Timestamp indicating when the comment was made.
* text: Content of the comment.
* user\_id: Unique identifier for the user who made the comment.
* username: Username of the commenter.

**Technologies and Libraries Used**

* **Python**: Core programming language.
* **Jupyter Notebook**: Development environment.
* **Pandas**: Data manipulation and analysis.
* **NLTK (Natural Language Toolkit)**: For text preprocessing, stopword removal, and lemmatization.
* **VADER (Valence Aware Dictionary and sEntiment Reasoner)**: For efficient sentiment analysis of social media text.

**Workflow**

**1. Data Loading and Preprocessing**

* Imported libraries: pandas, nltk, vaderSentiment.
* Downloaded NLTK resources like stopwords and wordnet.
* Removed stopwords and lemmatized text to standardize comments.

**2. Text Processing**

* Tokenized text to extract meaningful words.
* Applied lemmatization to reduce words to their root forms (e.g., "running" to "run").
* Evaluated emoji usage to understand emotional cues.

**3. Sentiment Analysis**

* Utilized the VADER library for sentiment classification into positive, negative, and neutral sentiments.

**4. Spam and Fake Comment Detection**

* Identified patterns in spammy and fake comments using textual analysis and user activity metrics.

**5. Visualization and Insights**

* Created visualizations for sentiment distribution.
* Highlighted majority opinions about posts based on aggregated sentiment scores.

**How to Run the Project**

**Prerequisites**

1. Install Python 3.9 or higher.
2. Install required libraries:
3. pip install pandas nltk scikit-learn vaderSentiment
4. Download the dataset and place it in the project directory.

**Steps to Run**

1. Open the comment\_sentiment\_nlp\_project.ipynb file in Jupyter Notebook.
2. Run all cells sequentially.
3. View outputs for sentiment analysis and spam detection.

**Future Scope and Enhancements**

* **Deep Learning Models**: Implement advanced models like BERT for improved sentiment classification.
* **Real-time Analysis**: Integrate APIs for live comment scraping and analysis.
* **Emoji Sentiment Mapping**: Create a dedicated model for interpreting emojis more effectively.
* **VADER Optimization**: Fine-tune VADER thresholds to improve social media sentiment detection.

**Feel free to contact me by DM for any suggestions, discussions or anything**

-Vicky Patil

[Vickyvpatil25@gamil.com](mailto:Vickyvpatil25@gamil.com)