

Influencer Analytics using Text, Social Media, and Web Analytics

A Comprehensive NLP and Sentiment Analysis Project

Overview

This project focuses on analyzing influencer performance using Text Analytics, Social Media Analytics, and Web Analytics techniques. It applies Natural Language Processing (NLP), sentiment analysis, content pattern extraction, and engagement modeling to identify authentic and high-performing influencers. The analysis is carried out in Python using Jupyter Notebook.

Objectives

- Analyze influencer-generated content through NLP
- Understand audience reactions using sentiment analysis
- Identify top keywords, topics, and hashtags
- Explore engagement behavior and audience demographics
- Estimate conversion potential by combining sentiment and engagement
- Build dashboards for comprehensive performance monitoring

Features and Use Cases

Use Case 1: Engagement Behavior Analysis

- Evaluates engagement rate patterns
- KDE density plots used to visualize distribution
- Synthetic enhancement ensures meaningful visuals even with limited data

Use Case 2: Audience Demographics

- Visualizes distributions for age groups, gender, and country
- Computes an activity score to assess country-level engagement
- Helps identify strong audience segments

Use Case 3: Content Pattern and NLP Analysis

- Full text cleaning pipeline including URL removal and tokenization
- Key word extraction using frequency analysis
- Hashtag identification to determine content themes and niches

Use Case 4: Conversion Score Estimation

- conversion_score calculated as engagement multiplied by sentiment
- Provides insight into potential campaign performance
- KDE chart used for smooth distribution visualization

Use Case 5: Performance Monitoring Dashboard

- Heatmap showing correlations among sentiment, engagement, clusters, and conversion
- Violin plots for sentiment grouped by content clusters
- Bubble chart comparing engagement and ROI metrics
- Enables a detailed influencer performance review

Technology Stack

- Python
- Jupyter Notebook
- Pandas, NumPy
- NLTK (VADER Sentiment Analyzer)
- Seaborn, Matplotlib
- Regex for text processing

Data Processing and Cleaning

The preprocessing steps include:

- Automatic detection of the most relevant text column
- Lowercasing and whitespace normalization
- Removal of URLs and special characters
- Tokenization and basic NLP preprocessing
- Fallback sentiment scoring when VADER is unavailable
- Synthetic generation of engagement and sentiment values when data is flat

Project Structure

```
project/
|
|   data/
|   |   influencer_data.csv
|
|   notebooks/
|   |   influencer_analytics_master.ipynb
|
|   README.md
|
|   requirements.txt
```

```
## How to Run the Project
```

1. Install required libraries:

```
pip install pandas numpy matplotlib seaborn nltk
```

2. Download the VADER lexicon if needed:

```
import nltk  
nltk.download('vader_lexicon')
```

3. Place your dataset (CSV) in the specified directory.

4. Open the Jupyter Notebook and execute cells in sequence.

```
## Key Outputs
```

- Engagement distribution plots
- Audience demographic breakdown
- Keyword and hashtag frequency charts
- Sentiment analysis results
- Conversion scoring dashboard
- Influencer performance comparison

```
## Future Improvements
```

- Integration of machine learning models for influencer recommendation
- Real-time API connections to platforms like Instagram, YouTube, Twitter
- Time-series analysis for influencer growth trends
- Deployment using Streamlit or Flask

```
## Author
```

Ragutu Deepika

MBA – Business Analytics

Project: Text, Social Media, and Web Analytics