AI for Social Media Trend Analysis

Aniakwa Nathan

August 16, 2025

Overview

This document presents a Python-based analysis of social media trends using sample posts related to AI in social media. The script loads sample data, preprocesses text by removing stopwords and special characters, counts word and hashtag frequencies, and generates visualizations. The full code, results, and visualization descriptions are provided below.

Full Code

The following is the complete code from the Jupyter notebook:

```
import pandas as pd
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from collections import Counter
import matplotlib.pyplot as plt
from wordcloud import WordCloud
import re

# Download required NLTK data
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('punkt_tab')

def load_sample_data():
    """Load sample social media data (simulating posts)"""
    # Sample data - in a real project, this would come from an
       API or database
    posts = [
        "Loving the new AI features on this app! #AI #Tech",
        "AI is changing how we create content! #
           ArtificialIntelligence",
        "Just saw an amazing AI-generated video! #SocialMedia #AI
           ",
        "The future is AI-driven social platforms #TechTrends",
        "Anyone tried the new AI filters? They're awesome! #AI #
           Filters",
```

```
        "AI is making social media so much fun! #Tech #SocialMedia",
        "New AI tools for content creators are  #AI #ContentCreation",
    ]
    return pd.DataFrame(posts, columns=['post'])

def preprocess_text(text):
    """Clean and preprocess text data"""
    # Convert to lowercase
    text = text.lower()
    # Remove URLs, mentions, and special characters
    text = re.sub(r'http\S+|@\w+|#[^\s]+|[^\w\s]', '', text)
    # Tokenize
    tokens = word_tokenize(text)
    # Remove stopwords
    stop_words = set(stopwords.words('english'))
    tokens = [word for word in tokens if word not in stop_words]
    return tokens

def analyze_trends(df):
    """Analyze social media posts for trends"""
    # Apply preprocessing to all posts
    df['tokens'] = df['post'].apply(preprocess_text)

    # Count word frequencies
    all_words = []
    for tokens in df['tokens']:
        all_words.extend(tokens)

    word_freq = Counter(all_words)

    # Get hashtags
    hashtags = []
    for post in df['post']:
        hashtags.extend(re.findall(r'#(\w+)', post.lower()))

    hashtag_freq = Counter(hashtags)

    return word_freq, hashtag_freq

def visualize_trends(word_freq, hashtag_freq):
    """Create visualizations for trends"""
    # Plot top 5 words
    plt.figure(figsize=(10, 5))
    top_words = dict(sorted(word_freq.items(), key=lambda x: x[1], reverse=True)[:5])
    plt.bar(top_words.keys(), top_words.values())
    plt.title('Top 5 Words in Social Media Posts')
    plt.xlabel('Words')
    plt.ylabel('Frequency')
```

```
    plt.savefig('word_freq.png')
    plt.close()

def main():
    # Load data
    df = load_sample_data()

    # Analyze trends
    word_freq, hashtag_freq = analyze_trends(df)

    # Visualize results
    visualize_trends(word_freq, hashtag_freq)

    # Print top trends
    print("Top 5 Words:")
    for word, count in sorted(word_freq.items(), key=lambda x: x[1], reverse=True)[:5]:
        print(f"{word}: {count}")

    print("\nTop 5 Hashtags:")
    for hashtag, count in sorted(hashtag_freq.items(), key=lambda x: x[1], reverse=True)[:5]:
        print(f"#{hashtag}: {count}")

if __name__ == "__main__":
    main()

# Additional visualization code for hashtags
df = load_sample_data()
word_freq, hashtag_freq = analyze_trends(df)

plt.figure(figsize=(10, 5))
top_hashtags = dict(sorted(hashtag_freq.items(), key=lambda x: x[1], reverse=True)[:5])
plt.bar(top_hashtags.keys(), top_hashtags.values())
plt.title('Top 5 Hashtags in Social Media Posts')
plt.xlabel('Hashtags')
plt.ylabel('Frequency')
plt.savefig('hashtag_freq.png')
plt.close()

# Word cloud generation
wordcloud = WordCloud(width=800, height=400, background_color='white').generate_from_frequencies(word_freq)
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud of Social Media Posts')
plt.savefig('wordcloud.png')
plt.close()
plt.show()
```

Analysis Results

The analysis processes 7 sample posts about AI applications in social media. After preprocessing (lowercasing, removing URLs, mentions, hashtags, special characters, and stopwords), the key trends are identified through frequency counts.

Top 5 Words

Word	Frequency
ai	5
new	3
content	2
social	2
loving	1

Table 1: Top 5 Words in Social Media Posts

Top 5 Hashtags

Hashtag	Frequency
ai	4
tech	2
socialmedia	2
artificialintelligence	1
techtrends	1

Table 2: Top 5 Hashtags in Social Media Posts

Visualizations

The script generates three visualizations (saved as PNG files in the original notebook):

- **Top 5 Words Bar Chart**: A bar chart displaying the words ‘ai’, ‘new’, ‘content’, ‘social’, and ‘loving’ on the x-axis, with bar heights corresponding to frequencies 5, 3, 2, 2, and 1 on the y-axis.
- **Top 5 Hashtags Bar Chart**: A bar chart displaying the hashtags ‘ai’, ‘tech’, ‘socialmedia’, ‘artificialintelligence’, and ‘techtrends’ on the x-axis, with bar heights corresponding to frequencies 4, 2, 2, 1, and 1 on the y-axis.
- **Word Cloud**: A visual representation where words are sized by frequency. ‘ai’ appears largest, followed by ‘new’, and smaller words like ‘content’, ‘social’, ‘loving’, ‘features’, ‘changing’, ‘create’, ‘saw’, ‘amazing’, ‘generated’, ‘video’, ‘future’, ‘driven’, ‘platforms’, ‘tried’, ‘filters’, ‘awesome’, ‘making’, ‘media’, ‘fun’, ‘tools’, ‘creators’.