Drug Benefits Review

Team - Blockers

Using Text and Sentiment Analysis

Table of Content

Description of Drug Benefits Dataset and Problems

Pre-processing of Dataset

Explanation of Analysis

Outcomes from the Drug Dataset

Actionable Recommendations

Drug Dataset Description

Comprehensive Drug Comments Collection

- Over 4,000 comments from the FDA
- Insights into patient experiences and sentiments

Detailed Record Structure

- Includes drug name, rating, and treated condition
- Reviews on benefits and side effects

Analysis of Patient Perspectives

- Focus on benefitsReview column for trends
- Understanding consumer reactions to drugs

Key Columns in the Drug Dataset

ID	Drug Name	Rating	Condition	Benefits Review	Side Effects Review	Comments Review
Record ID	Name of the drug	Patient rating	Condition treated	Review of benefits	Description of side effects	Additional comments

FDA Comments and Columns

Column Name	Description
id	Record ID
drugName	Name of the drug
rating	Patient rating of the drug
condition	Condition treated by the drug
benefitsReview	Review of the drug's benefits by the patient
sideEffectsReview	Review of side effects experienced by the patient
commentsReview	Additional comments or reviews by the patient

Problems with Drug Dataset

Data Quality Concerns: Missing values, noise, inconsistencies, Capitalization, punctuation, and misspellings issues

Text Preprocessing Necessities: tokenization, stemming, lemmatization, and stop word

·**Subjectivity in Patient Reviews:** Reviews influenced by personal experiences and biases

Analysis Interpretation Difficulties

Topic Extraction Complexities

Time Consuming Data Analysis

Dataset Integrity Issues

Missing Crucial Data

Preprocessing Steps for Analysis

Steps for Pre-Processing 'benefitsReview' Text Data

Step	Action	Purpose
1	Convert to lowercase	Ensure consistency
2	Remove noise elements	Clean data
3	Tokenize text	Prepare for analysis
4	Lemmatize words	Normalize data
5	Remove stop words	Focus on valuable words
6	Handle missing data	Maintain data integrity

Consistency in Text Data
- **Convert text to lowercase to avoid case-sensitive duplication**

Cleaning Text Content
- **Remove URLs, special characters, numbers, and punctuation**

Tokenization
- **Split text into individual words for analysis**

Normalization
- **Lemmatize words to their base form**

Stop Words Removal
- **Eliminate common words that add little value**

Data Integrity
- **Handle missing data by dropping incomplete rows**

Analysis

- Text Analysis

- Sentiment Analysis

- N-gram Analysis

- Topic Analysis

- Word Frequency and Visualization

Text Mining Analysis

Top Words in Identified Topics	
Topic Number	Top Words
1	Word1, Word2, Word3, ..., Word10
2	Word1, Word2, Word3, ..., Word10
3	Word1, Word2, Word3, ..., Word10
4	Word1, Word2, Word3, ..., Word10
5	Word1, Word2, Word3, ..., Word10
6	Word1, Word2, Word3, ..., Word10

- Text Mining Techniques
  - Utilization of CountVectorizer for token matrix
  - Employment of LDA for theme identification
- Insights from 'benefitsReview'
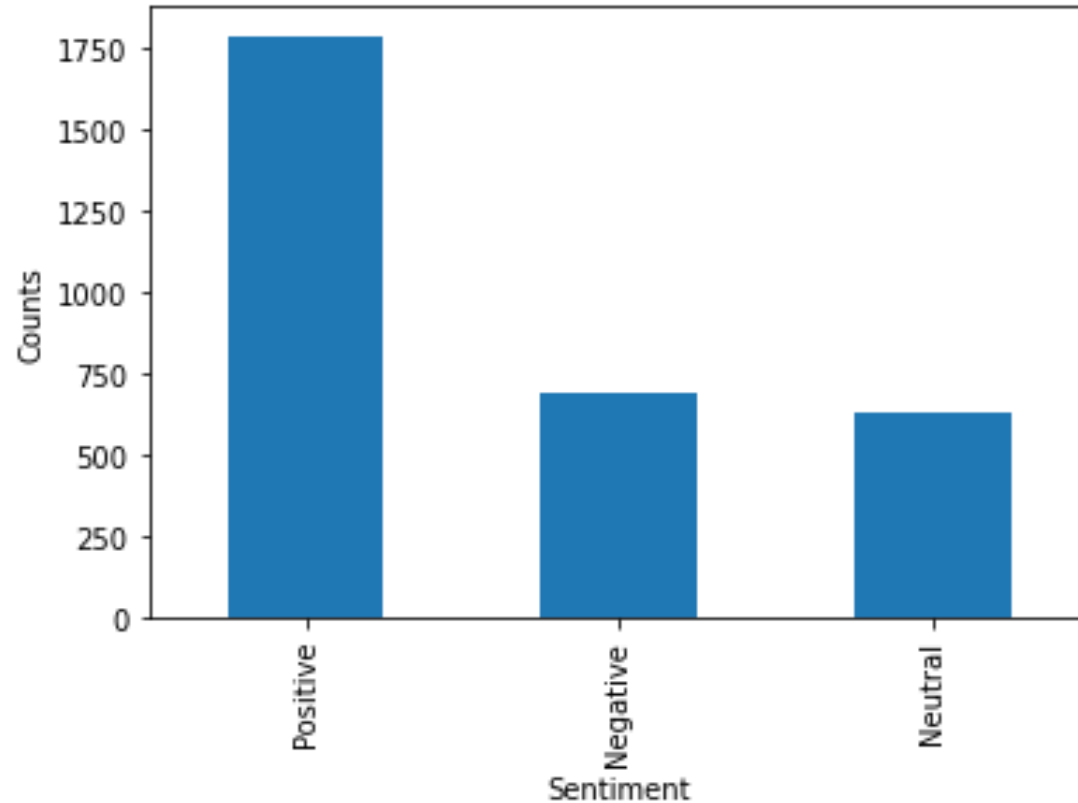 - Top 10 words for 6 topics reveal dominant themes
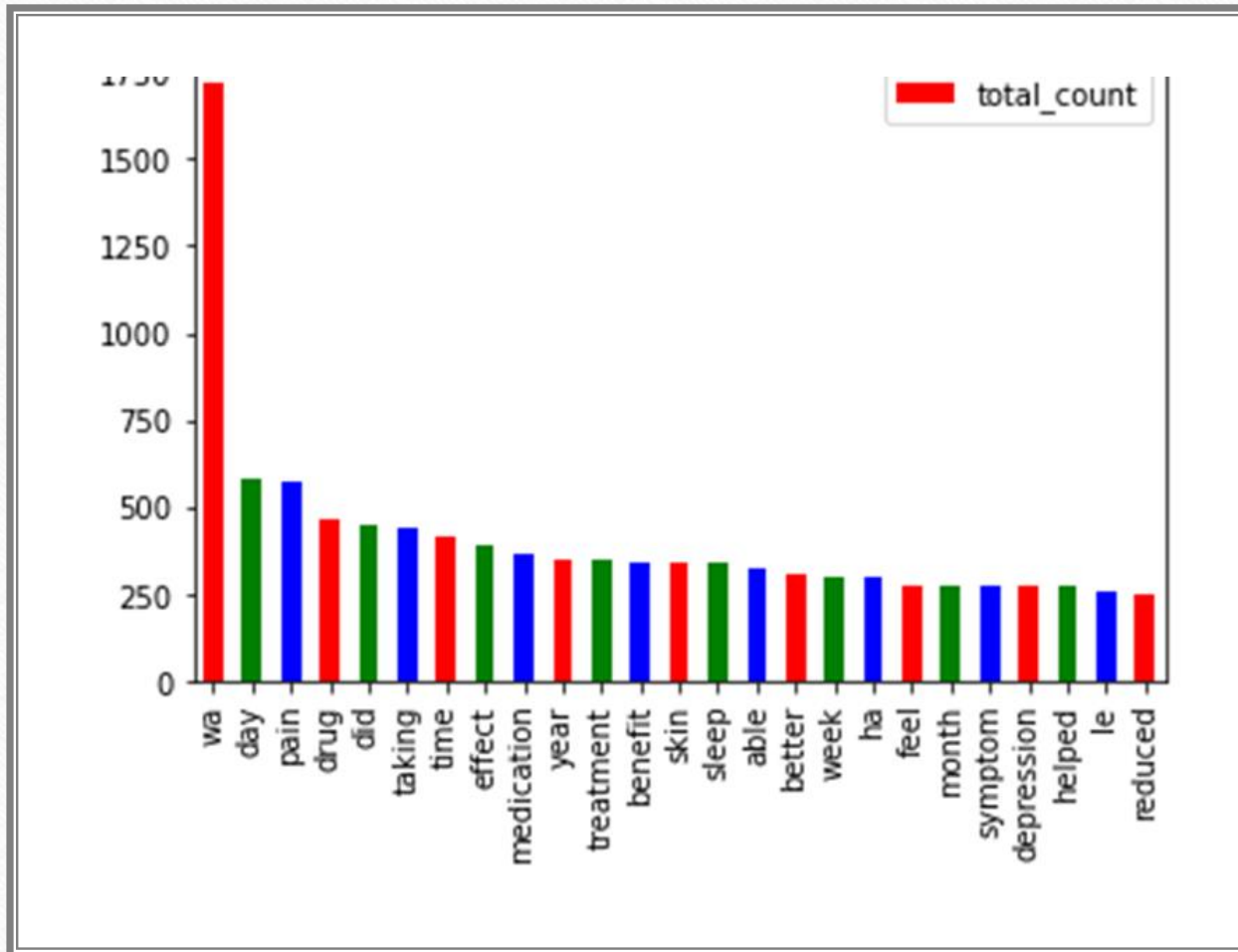
- Textual analysis

- Top 25 words
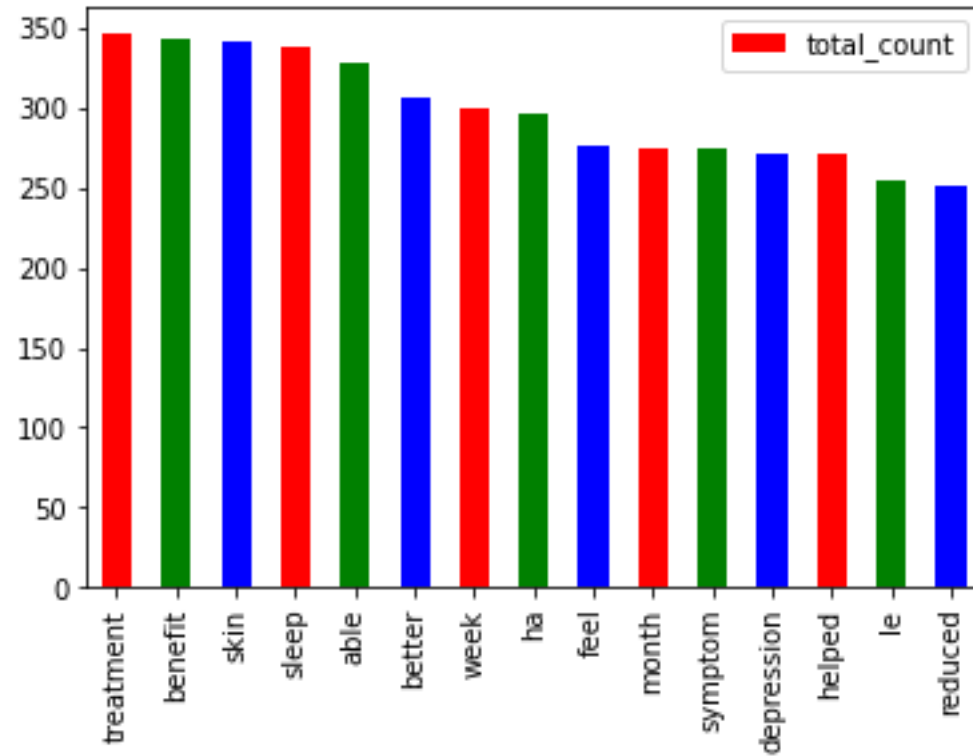
- Textual analysis

- After excluding top 10 to get your **Middle Group**

Sentiment Classification

Review Sentiment Classification
- Utilized TextBlob library to calculate polarity scores
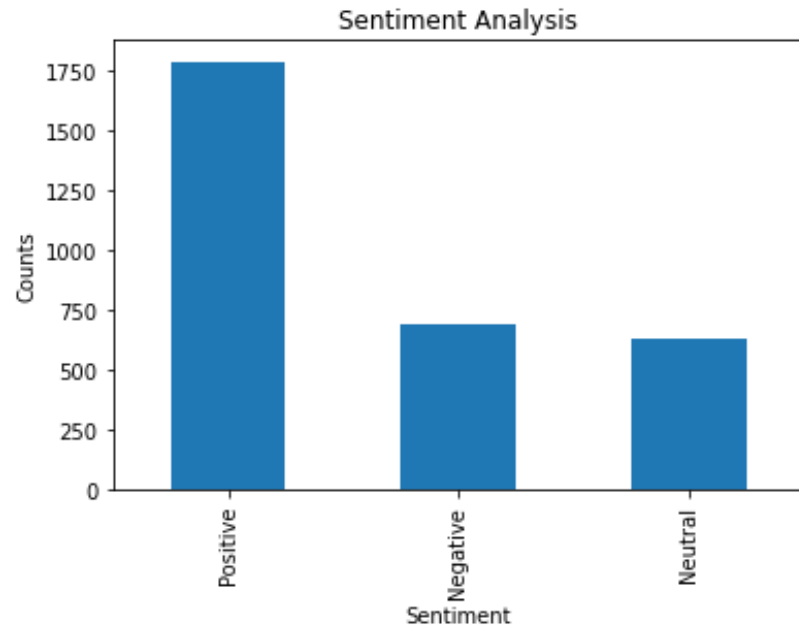- Categorized reviews based on sentiment polarity

Sentiment Distribution Visualization
- Displayed using a bar chart
- Provided an overview of drug sentiment

Data Export for In-Depth Analysis
- Positive and negative reviews saved in separate CSV files

Sentiment analysis

- Positive Sentiment Reinforcement
  - Utilize drugs with the most positive reviews in marketing
  - Incorporate patient testimonials showcasing benefits
  - Employ promotional materials that highlight effectiveness

N-gram and Topic Analysis

N-gram Analysis:

- The code performs n-gram analysis using both sci-kit-learn's Count Vectorizer and NLTK (Natural Language Toolkit) libraries.

- It extracts the most frequent 2-gram and 3-gram phrases, providing insights into common word combinations discussed in the reviews.

Topic Analysis:

- Conducted topic modeling to identify common themes in the benefits reviews.

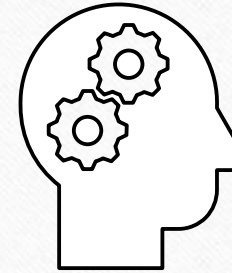- This helped us understand the main topics discussed by consumers.

Intuitive Visualizations
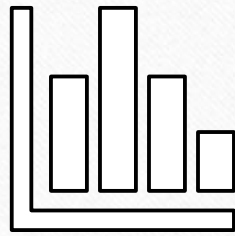
Visual Data Analysis in Drug Reviews

Word clouds showcase frequently mentioned terms

Frequency plots highlight common phrases

Consumer Feedback Interpretation

Visualizations offer intuitive topic comprehension

Themes in reviews are easily identified

Word Frequency and Visualizations
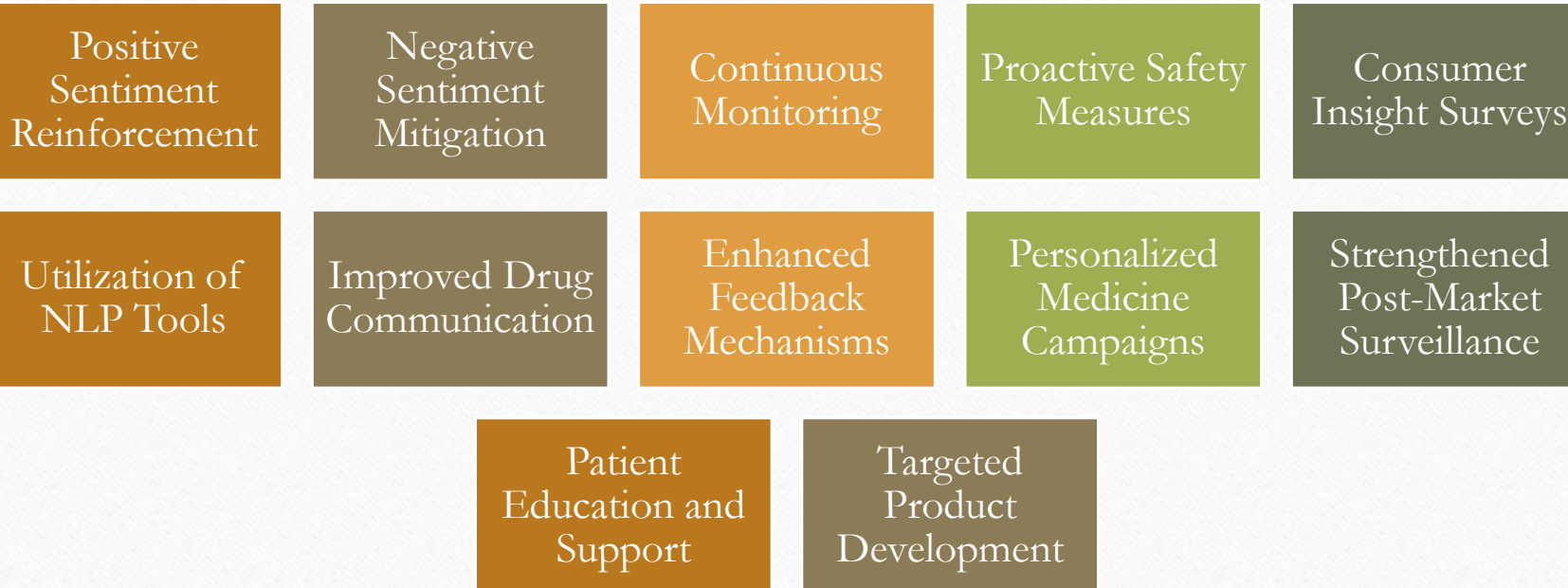

Visual Analysis of Drug Reviews

- Word clouds and frequency plots used for visualization
- Highlighting most-used words and phrases

Intuitive Understanding of Consumer Discussions

- Visualizations reveal prominent topics and themes

Justifiable and Actionable Recommendations

Positive Sentiment Reinforcement	Negative Sentiment Mitigation	Continuous Monitoring	Proactive Safety Measures	Consumer Insight Surveys
Utilization of NLP Tools	Improved Drug Communication	Enhanced Feedback Mechanisms	Personalized Medicine Campaigns	Strengthened Post-Market Surveillance
	Patient Education and Support	Targeted Product Development		

Presented by: