**Sentiment Analysis and Drug Review Insights: A Data-Driven Approach**

**🏥 Project Overview**

This project focuses on understanding user sentiments and extracting insights from thousands of drug reviews. Using sentiment analysis, visualizations, and machine learning techniques, we classify user opinions as positive, negative, or neutral. The objective is to identify patterns in user feedback, assess drug effectiveness, and uncover common themes in user reviews.

**🎯 Objectives:**

1. Analyze user sentiments about drugs.
2. Visualize patterns in reviews, including ratings and review lengths.
3. Use machine learning models to classify review sentiments.
4. Identify common themes and topics using topic modeling (LDA).

**📊 Dataset Overview**

The dataset contains user reviews for a variety of drugs, including fields such as:

* **Drug Name**: The name of the drug being reviewed.
* **Condition**: The medical condition treated by the drug.
* **Review**: The user’s feedback.
* **Rating**: User rating of the drug, ranging from 1 to 10.
* **Useful Count**: Number of users who found the review helpful.
* **Review Length**: Number of words in each review.
* **Sentiment**: Classification of the review (Positive, Negative, Neutral).

**⚙️ Project Workflow**

**1. Data Exploration and Preprocessing**

* We begin by loading and inspecting the dataset for missing values, duplicate entries, and data consistency.
* **Text Cleaning**: The reviews are preprocessed by removing special characters, stopwords, and performing lemmatization.

**2. Exploratory Data Analysis (EDA)**

* **Sentiment Distribution**: A breakdown of positive, negative, and neutral reviews is provided, along with a pie chart and bar chart.
* **Top Reviewed Drugs**: Visualize the most frequently reviewed drugs.
* **Rating Distribution**: Analyze the distribution of ratings and observe which drugs receive the highest user ratings.
* **Review Length Distribution**: Plot the distribution of review lengths to identify trends in how users express their opinions.

**3. Sentiment Analysis**

* **Sentiment Scoring**: Using TextBlob, each review is assigned a sentiment score (ranging from -1 for negative to +1 for positive).
* **Sentiment Classification**: Reviews are classified into Positive, Negative, or Neutral based on the sentiment score.

**4. Machine Learning Models**

Several machine learning models are applied to predict the sentiment of reviews:

1. **Logistic Regression**: Provides a baseline model for sentiment classification.
2. **Decision Tree Classifier**: Models the decision-making process for sentiment classification.
3. **Random Forest Classifier**: Improves classification by averaging multiple decision trees.

**5. Model Evaluation**

* **Accuracy**: Evaluate the models based on accuracy, precision, recall, and F1 score.
* **K-Fold Cross-Validation**: Ensure model robustness through cross-validation.

**6. Topic Modeling**

* Using **Latent Dirichlet Allocation (LDA)**, key themes in the reviews are identified. The top topics include:
  + Medication side effects
  + Anxiety treatment
  + Birth control and its effects
  + Pain management
  + Skin conditions

**🔍 Key Insights and Observations**

* **Sentiment Distribution**: Approximately 63.7% of the reviews are positive, 34.9% are negative, and 1.4% are neutral.
* **Rating Trends**: Users tend to rate drugs highly, with ratings of 9 and 10 being the most frequent. Lower ratings (1-8) are less common.
* **Top Drugs**: Drugs such as **Escitalopram** and **Mirena** received the highest number of reviews, indicating their popularity and frequent use.
* **Common Themes**: Topic modeling revealed recurring themes around medication side effects, anxiety management, and birth control, which provide valuable insights for pharmaceutical companies and healthcare professionals.

**🚀 Conclusion**

Through sentiment analysis and machine learning, this project provides valuable insights into user experiences with various drugs. The findings can help healthcare professionals, patients, and pharmaceutical companies better understand drug effectiveness, side effects, and overall user satisfaction.