Patient Experience Clustering: Insights, Findings, and Recommendations

Abstract

This research explores patient experiences with various drugs by analyzing their reviews and ratings. Using clustering techniques, we identify patterns in drug effectiveness and provide actionable recommendations for improving patient outcomes. The study utilizes a dataset containing patient reviews, drug names, conditions, ratings, and other relevant information. By segmenting drugs based on patient feedback, this research aims to enhance treatment decisions for healthcare providers and patients alike.

1. Introduction The effectiveness of pharmaceutical drugs varies among patients, influencing satisfaction levels and treatment adherence. Understanding these variations through data-driven clustering techniques allows for better recommendations and improved patient experiences. This research applies clustering algorithms to group drugs based on patient reviews and ratings, offering insights into drug efficacy and patient satisfaction trends.

2. Research Objectives The primary objectives of this study are:

1. To clean and preprocess patient drug review data for meaningful analysis.
2. To explore the distribution of patient conditions, drugs, and ratings.

3. To apply clustering techniques to identify patterns in drug effectiveness.
4. To provide drug recommendations based on patient ratings and clustering results.
5. To visualize and interpret findings for practical healthcare applications.

3. Methodology

3.1 Data Collection and Preprocessing The dataset comprises 161,297 entries with seven key attributes: Unnamed: 0, drugName, condition, review, rating, date, and usefulCount. Missing values were detected in the condition column (899 missing entries), which were removed to ensure data integrity. The date column was converted to a datetime format for chronological analysis. Additionally, an effectiveness metric was created to classify drugs into three categories:

- Effective: Rating ≥ 8
- Moderately Effective: Rating between 5 and 8
- Ineffective: Rating < 5

3.2 Exploratory Data Analysis (EDA) After preprocessing, the dataset contained 160,398 rows. The average drug rating was 6.99, with a standard deviation of 3.27, indicating significant variability in patient experiences. The most frequently treated conditions were:

1. Birth Control
2. Depression
3. Anxiety
4. Pain

5. Acne

The most frequently reviewed drugs were:

1. Levothyroxine Sodium
2. Sertraline
3. Bupropion
4. Metformin
5. Gabapentin

3.3 Clustering Approach K-Means Clustering was employed to group drugs based on patient ratings, forming five clusters. Key findings include:

- **Cluster 0**: Drugs for contraception, hyperhidrosis, and smoking cessation with high patient satisfaction (e.g., Drysol for Hyperhidrosis had a rating of 10.0).
- **Cluster 4**: Drugs for asthma and diabetes with lower patient satisfaction (e.g., Formoterol for Asthma had a rating of 4.0).

3.4 Recommendation System A function was developed to recommend the top three drugs for a given condition within a cluster. For example, in Cluster 0, the top-rated drugs for Birth Control were:

1. Drysol (Rating: 10.0)
2. Depo-Provera (Rating: 9.2)
3. Chantix (Rating: 9.0)

These recommendations provide valuable insights for patients and healthcare professionals in selecting effective treatments.

4. Key Insights and Discussion

- **High-Rated Drugs**: Drugs like Drysol and Depo-Provera exhibited high patient satisfaction, making them ideal choices.
- **Low-Rated Drugs**: Chronic condition treatments such as those for Asthma and Diabetes received lower ratings, suggesting potential concerns regarding efficacy or side effects.
- **Cluster Differences**: Cluster 0 contained drugs for preventive and lifestyle conditions, while Cluster 4 focused on chronic illnesses, reflecting diverse patient experiences.

5. Challenges and Solutions

- **Data Quality Issues**: Missing values in the condition column were handled by dropping incomplete rows.
- **Variability in Ratings**: Clustering helped categorize drugs based on similar patient experiences to minimize discrepancies.
- **Lack of Additional Features**: Future studies could incorporate side effects, patient demographics, and prescription adherence data for a more comprehensive analysis.

6. Conclusion and Future Work This research successfully applied clustering techniques to analyze patient experiences with various drugs, providing valuable insights into drug effectiveness and patient satisfaction. The findings emphasize the importance of data-driven healthcare decisions. Future work could explore additional features, such as side effects and longterm adherence patterns, to enhance clustering accuracy and recommendation effectiveness.

7. Recommendations

- **For Healthcare Providers**: Leverage clustering insights to guide treatment decisions and improve patient management strategies.
- **For Patients**: Use the recommended drugs for specific conditions to make informed choices.
- **For Researchers**: Investigate low-rated drugs further to understand patient dissatisfaction and potential formulation improvements.

This study highlights the role of patient feedback in optimizing drug selection and underscores the potential of machine learning in personalized healthcare.