Project Proposal

Sentiment Analysis on Patient Drug Reviews

Yuval Elisha - 209168988
Nikol Jabotinski - 325133619

Problem Statement

Motivation: Understanding patient experiences with medications through unstructured text analysis.

Input: Unstructured patient reviews of drugs and conditions.

Output: Categorization of reviews into 3 sentiment classes (positive, negative, neutral).

Relevant NLP Task: Sentiment Analysis using LLMs.

Challenges:

- Analyzing informal and subjective patient language.
- Handling sentiment polarity without explicit labels.
- Addressing data sparsity for rare conditions.

Training and Test Data

Data Type: Unstructured, unlabeled 273,915 patient reviews on drugs and medical conditions.

Source: Patient Insights: 2.8Lakh Drug & Condition Reviews from Kaggle.

Preprocessing Steps:

- Cleaning and tokenization.
- Removing irrelevant data (e.g., stopwords, non-text characters).
- Using embeddings for sentiment classification.

Input Example

*"On Abilify for Bipolar and it is the first time in my entire life that **I have felt stable**.*
***I have no side effects** from this medication.*
*I take it at night and it makes me a bit sleepy **which is good** for my sleep anyways.*
*I have tried a few other mood stabilizers without success and finally tried Abilify and **it works fantastically for me.**"*

*"I had so many **horrible experiences** with meds for my bipolar.*
*It scared me to try one more but I'm now on 2mg of abilify once daily and **found my old self again** after a very long time **living an unbearable life.***
*The only side effect is slight insomnia and I'm losing the ton of weight that I gained while on seroquel. Abilify **saved my life.**"*

Evaluating Model Performance

Evaluation Metrics:

- Sentiment Agreement Score (evaluating model consistency).
- Precision, Recall, F1-score for sentiment classification.

Comparison:

- Baseline: Lexicon-based sentiment analysis (word-matching approach).
- Proposed: LLM-based sentiment classification using embeddings & context understanding.

Train/Test Strategy:

- Unsupervised approach (evaluating model-generated sentiment vs. human-labeled subset).
- Using patient-reported adverse reactions as a control metric.

Thank you!

Any questions?