

Team-07

# NLP Project Evaluation-1

## Pharmaceutical Drug Matching Assistance

"SE22UARI201" "SE22UARI103" "SE22UARI147" "SE22UARI186"

# 1. Motivation

- 1> Because of the large number of medications accessible, the pharmaceutical business confronts difficulties matching drugs with similar ingredients, names, or effects.
- 2> Incorrect drug matching can result in errors while prescribing or administering medications, which can have major ramifications for patient safety.
- 3> An NLP-based drug matching system would assist healthcare providers in rapidly and precisely matching pharmaceuticals, decreasing human error and increasing efficiency.



## 2. Problem Statement

- 1> The problem is to accurately match pharmaceutical medications with similar names, active components, or functions.
- 2> Drug names might differ depending on brand or generic, and minor spelling differences can lead to confusion.
- 3> There is a need for an automated system that can efficiently match medications, lowering the risk of human mistake in drug identification and protecting patients.



# 3. Proposed Pipeline

01.

Data Collection: Gather datasets of drug names, ingredients, and related information from pharmaceutical databases.

02.

Text preprocessing involves cleaning up the text data via tokenization, stop word removal, and stemming/lemmatization.

03.

Entity Recognition: Use named entity recognition (NER) to detect important medicine names, components, and dosages.

# 3. Proposed Pipeline

04.

Similarity Matching: Using cosine similarity, Levenshtein distance, or other metrics, compare medications based on text descriptions.

05.

Classification: Create a classification model to predict if two medications are compatible based on their characteristics.

06.

Evaluation: To test the system's accuracy, use precision, recall, and F1 scores.

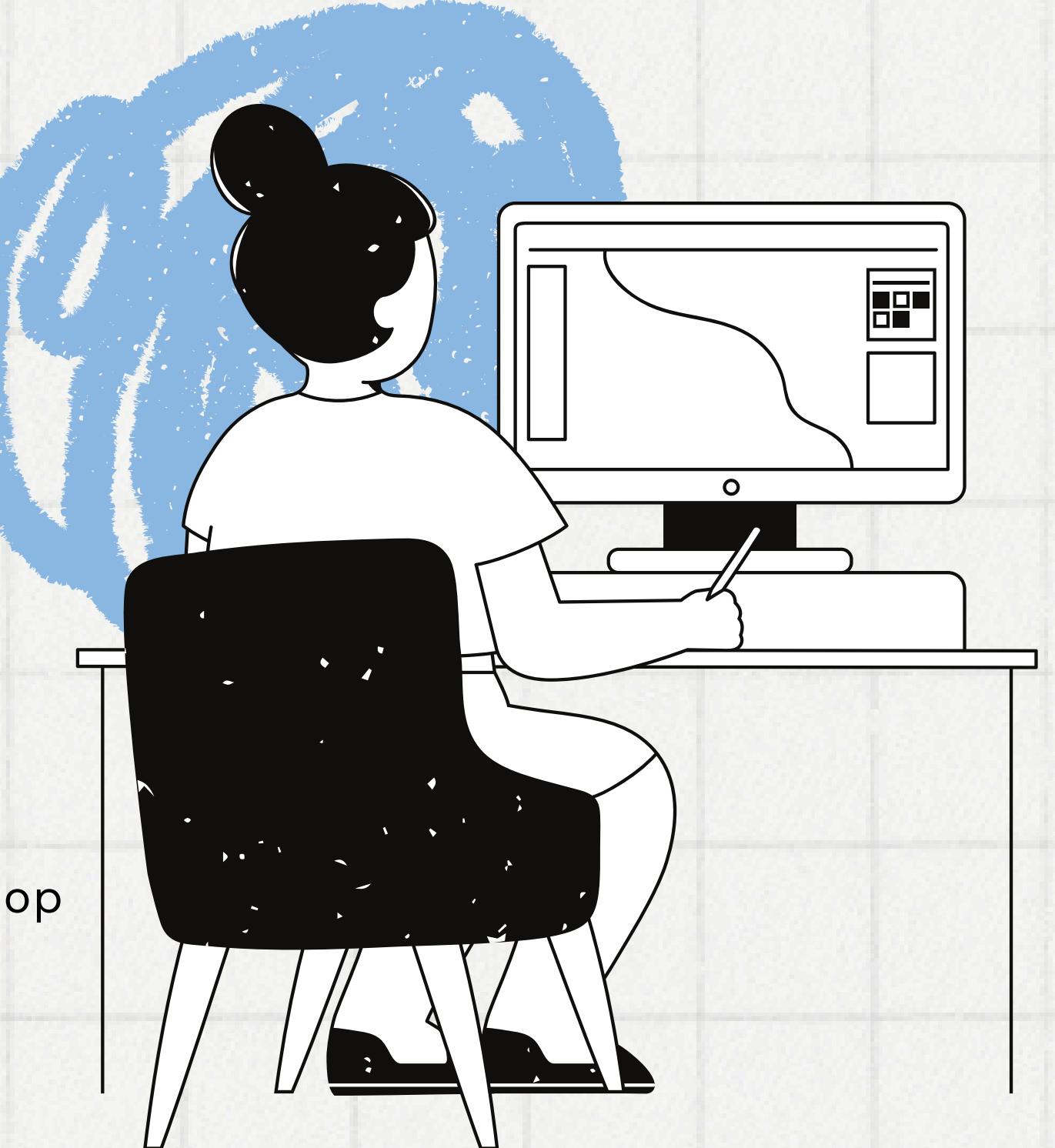
# 4.Timeline (5Weeks)

Week 1:

- Problem Understanding & Data Collection: Gather drug-related datasets (names, ingredients, descriptions) and familiarize yourself with the problem domain.

Week 2:

- Text Preprocessing & Exploration: Perform tokenization, stop word removal, stemming/lemmatization, and exploratory data analysis (EDA).



# 4.Timeline (5Weeks)

Week 3:

- Similarity Matching: Implement similarity measures like cosine similarity or Levenshtein distance for comparing drug names and descriptions.

Week 4:

- Model Building & Classification: Develop a classification model to predict matches between drugs based on their features.

Week 5:

- Evaluation & Finalization: Evaluate the model's performance, fine-tune as needed, and prepare the final report and presentation.





## 5. Expected Outcome/Application

- A functional NLP system that can accurately match pharmaceutical drugs based on textual data.
- The system will reduce errors in drug identification and prescribing, improving patient outcomes.
- It can be integrated into hospital management systems, pharmaceutical databases, or mobile health apps for real-time assistance in drug matching.

# Dataset

<https://www.kaggle.com/datasets/jithinanievarghese/drugs-side-effects-and-medical-condition>



```
import pandas as pd
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
data = pd.read_csv('/content/drugs_side_effects_drugs_com.csv')
nltk.download('punkt')
nltk.download('stopwords')
def preprocess_text(text):
    """Preprocess text by tokenizing, lowercasing, and removing stopwords."""
    if isinstance(text, str):
        tokens = word_tokenize(text.lower())
        tokens = [word for word in tokens if word.isalnum() and word not in stopwords.words('english')]
        return ' '.join(tokens)
    else:
        return ''
data['processed_side_effects'] = data['side_effects'].apply(preprocess_text)
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
```

**Thank you  
very much!**

**Team-07**