
Doctor
Medicine
Help

Introduction

- Currently consulting doctors has become too expensive for most people that are below lower middle class.
- Medical interns also struggle with remembering all the medicine names for a particular disease or symptom
- In case of extreme emergency it helps to have an instant medicine recommender that could potentially save lives.
- In some cases patients might feel like getting a second opinion.

PROBLEM STATEMENT

Our goal is to build a medicine recommendation system that recommends medicines for the given descriptive symptoms.

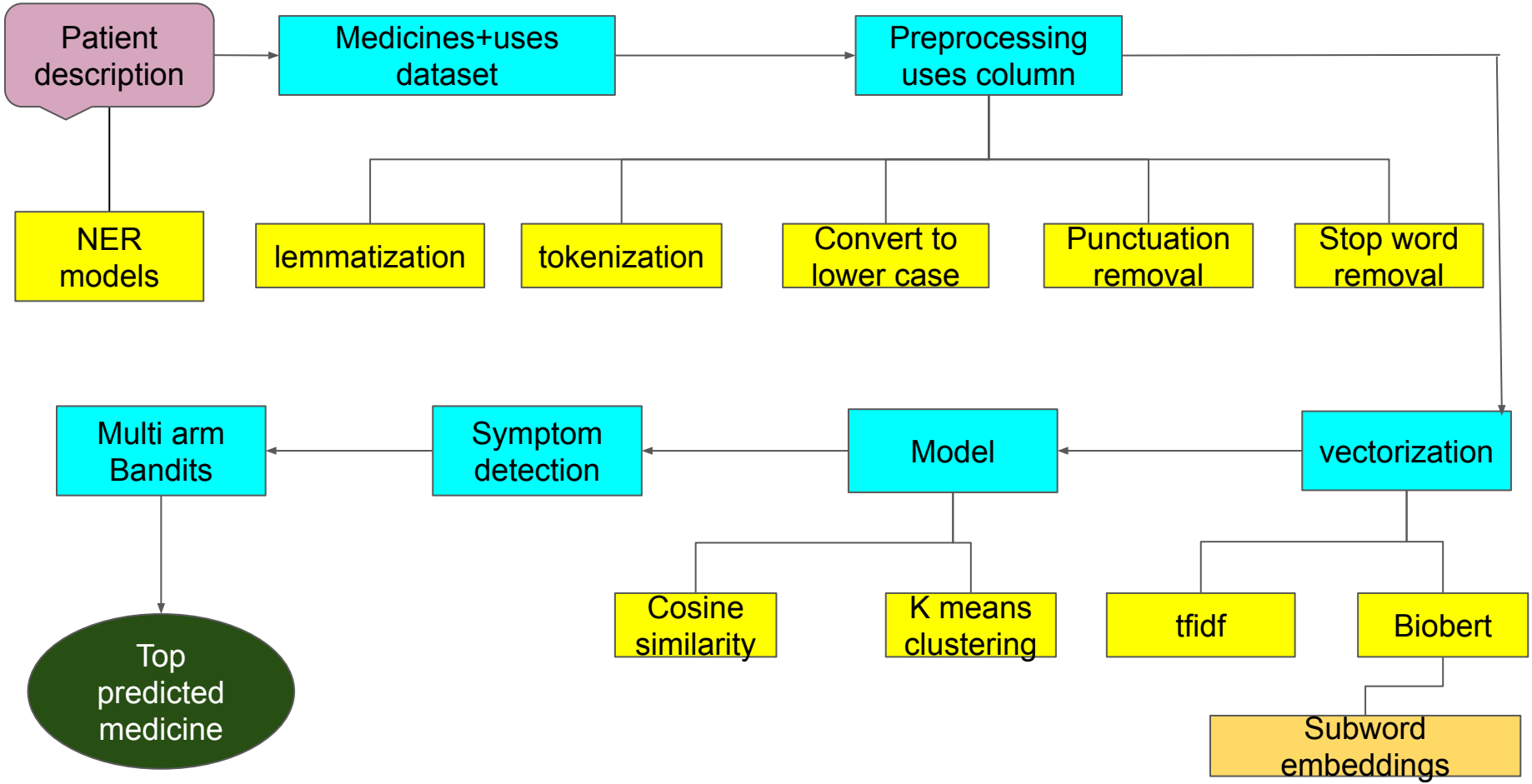
We wish to help people self identify diseases and recommend medicines to them

NOVELTY:

- Considering reviews in our MAB formula is one of our novelty’s

- Extracting symptoms from patient description is also one of the novelty.

PIPELINE OF OUR PROJECT



Datasets that we used

Medicine_Details.csv

- The dataset contains medicine names,their salt composition, manufacturer,uses and the reviews of medicines

	Medicine Name	Composition	Uses	Image URL	Excellent Review %	Average Review %	Poor Review %
0	Avastin 400mg Injection	Bevacizumab (400mg)	Cancer of colon and rectum Non-small cell lun...	https://onemg.gumlet.io/l_watermark_346,w_480,...	22	56	22
1	Augmentin 625 Duo Tablet	Amoxycillin (500mg) + Clavulanic Acid (125mg)	Treatment of Bacterial infections	https://onemg.gumlet.io/l_watermark_346,w_480,...	47	35	18
2	Azithral 500 Tablet	Azithromycin (500mg)	Treatment of Bacterial infections	https://onemg.gumlet.io/l_watermark_346,w_480,...	39	40	21
3	Ascoril LS Syrup	Ambroxol (30mg/5ml) + Levosalbutamol (1mg/5ml)...	Treatment of Cough with mucus	https://onemg.gumlet.io/l_watermark_346,w_480,...	24	41	35
4	Aciloc 150 Tablet	Ranitidine (150mg)	Treatment of Gastroesophageal reflux disease (...	https://onemg.gumlet.io/l_watermark_346,w_480,...	34	37	29

mtsamples.csv

	transcription	keywords
	SUBJECTIVE:, This 23-year-old white female presents with complaint of allergies. She used to have allergies when she lived in Seattle but she thinks they are worse here. In the past, she has tried Claritin, and Zyrtec. Both worked for short time b...	allergy / immunology, allergic rhinitis, allergies, asthma, nasal sprays, rhinitis, nasal, erythematous, allegra, sprays, allergic,
	PAST MEDICAL HISTORY:, He has difficulty climbing stairs, difficulty with airline seats, tying shoes, used to public seating, and lifting objects off the floor.	bariatrics, laparoscopic gastric bypass, weight loss programs, gastric bypass, atkin's diet, weight watcher's, body weight, laparoscopic gastric, weight

Symptoms _description.csv

- This dataset contains the symptoms and their description.

	symptom_Description
Symptoms	**Description**
Drug Reaction	An adverse drug reaction (ADR) is an injury caused by taking medication. ADRs may occur following a single dose or prolonged administration of a drug or result from the combination of two or
Malaria	An infectious disease caused by protozoan parasites from the Plasmodium family that can be transmitted by the bite of the Anopheles mosquito or by a contaminated needle or transfusion. Fal
Allergy	An allergy is an immune system response to a foreign substance that's not typically harmful to your body.They can include certain foods, pollen, or pet dander. Your immune system's job is to k
Hypothyroidism	Hypothyroidism, also called underactive thyroid or low thyroid, is a disorder of the endocrine system in which the thyroid gland does not produce enough thyroid hormone.
Psoriasis	Psoriasis is a common skin disorder that forms thick, red, bumpy patches covered with silvery scales. They can pop up anywhere, but most appear on the scalp, elbows, knees, and lower back
GERD	Gastroesophageal reflux disease, or GERD, is a digestive disorder that affects the lower esophageal sphincter (LES), the ring of muscle between the esophagus and stomach. Many people, incl

Extracting symptoms from description

- We used **'en_ner_bc5cdr_md'** model to find symptoms from description
- Trained on a combination of biomedical and chemical literature, it accurately identifies diseases, chemicals, and related entities.
- It specializes in recognizing biomedical and chemical entities in text

	the JNLPBA corpus.	
en_ner_bc5cdr_md	A spaCy NER model trained on the BC5CDR corpus.	Download
en_ner_bionlp13cg_md	A spaCy NER model trained on the BIONLP13CG corpus.	Download

Performance

Our models achieve performance within 3% of published state of the art dependency parsers and within 0.4% accuracy of state of the art biomedical POS taggers.

model	UAS	LAS	POS	Mentions (F1)	Web UAS
en_core_sci_sm	89.18	87.15	98.18	67.89	87.36
en_core_sci_md	90.08	88.16	98.46	68.86	88.04
en_core_sci_lg	89.97	88.18	98.51	68.98	87.89
en_core_sci_scibert	92.12	90.58	98.18	67.70	92.58

Model predictions

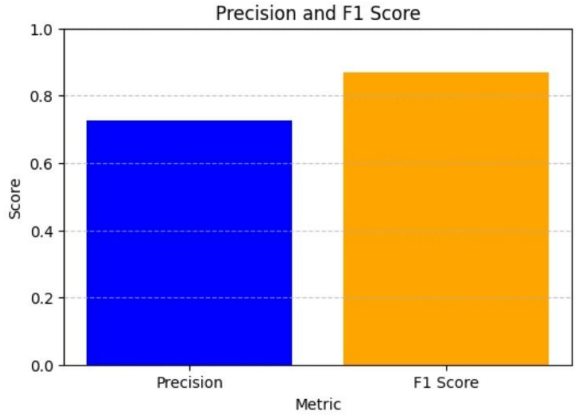
A 60-year-old man with a history of tubulovillous adenoma DISEASE underwent colonoscopy revealing a 2 cm sessile polyp DISEASE with high-grade dysplasia DISEASE , followed by negative sigmoidoscopies, but later presented with a 3 cm sessile polyp DISEASE , biopsied as villous adenoma DISEASE with low-grade dysplasia DISEASE , leading to MRI and EUS showing a T2 rectal carcinoma DISEASE involving mucosa and submucosa, prompting surgical resection with pathology revealing villous adenoma DISEASE without invasion and benign lymph nodes, followed by ileostomy reversal without recurrence.

A 57-year-old woman with a history of stage IIA breast cancer DISEASE treated with lumpectomy, axillary node dissection, chemotherapy, radiation and anti-estrogen CHEMICAL therapy presented with acute shortness of breath DISEASE after 17 years. She had tachycardia DISEASE , tachypnea DISEASE , hypertension DISEASE , jugular venous distention DISEASE , pulmonary rales DISEASE and peripheral edema DISEASE . Investigations revealed severe left ventricular dysfunction DISEASE with an LVEF of 0.20 but no evidence of blocked arteries or heart attack. Her heart function improved significantly with medical therapy and she remained asymptomatic on a low dose of metoprolol CHEMICAL after a year.

A 37-year-old man with Friedreich's ataxia DISEASE , a nerve and muscle disorder DISEASE that can weaken the heart, arrived at the ER with severe abdominal pain DISEASE . Despite having a previously diagnosed weak and enlarged heart (cardiomyopathy DISEASE), further tests revealed a surprising culprit for his pain DISEASE : a blockage in the artery supplying blood to his right kidney (renal infarction DISEASE). Intriguingly, imaging showed normal-looking vessels. Doctors meticulously ruled out other causes of his pain DISEASE , including metabolic disorders DISEASE , inflammatory conditions, and blood clotting DISEASE problems. He improved with medications that addressed both the blockage (blood thinners) and the resulting inflammation DISEASE (antibiotics), while his existing heart and blood pressure medications were adjusted. This case highlights a rare but serious complication in Friedreich's ataxia - renal artery embolism DISEASE . Early diagnosis and treatment, as seen here, are essential to prevent further kidney damage DISEASE and ensure a better outcome.

Testing model

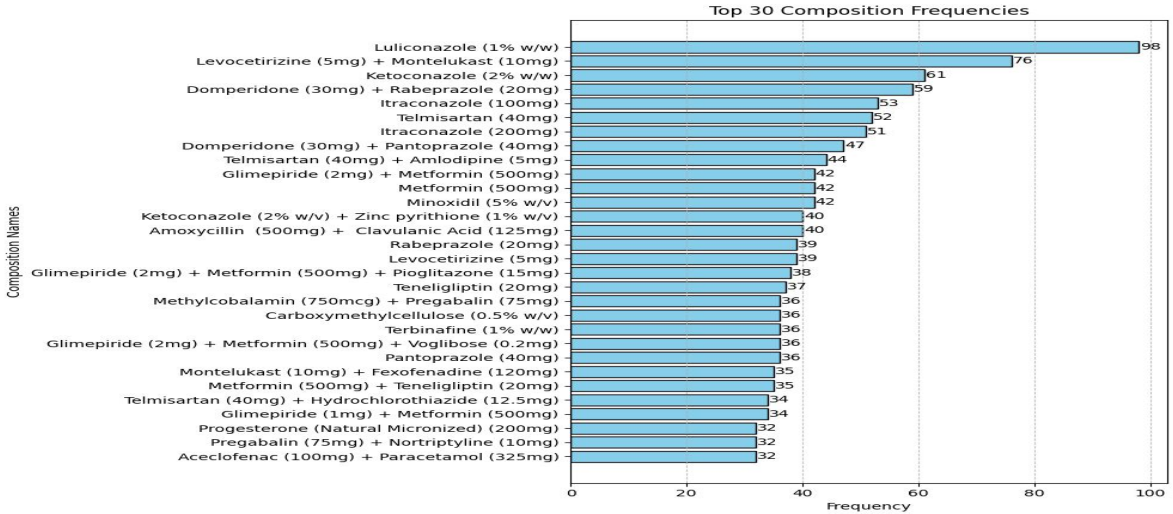
- Using **mtsamples.csv,** which has both description and keywords in it, we tested our model.
- This dataset has keywords which are medical terms
- We used our model on all transcriptions in that dataset
- We checked whether every keyword in the dataset is in our prediction.
- We tried to get the percentage of keywords which our model identified and get f1 scores.



Precision: 0.7256013745704467
F1 Score: 0.8691088701378884

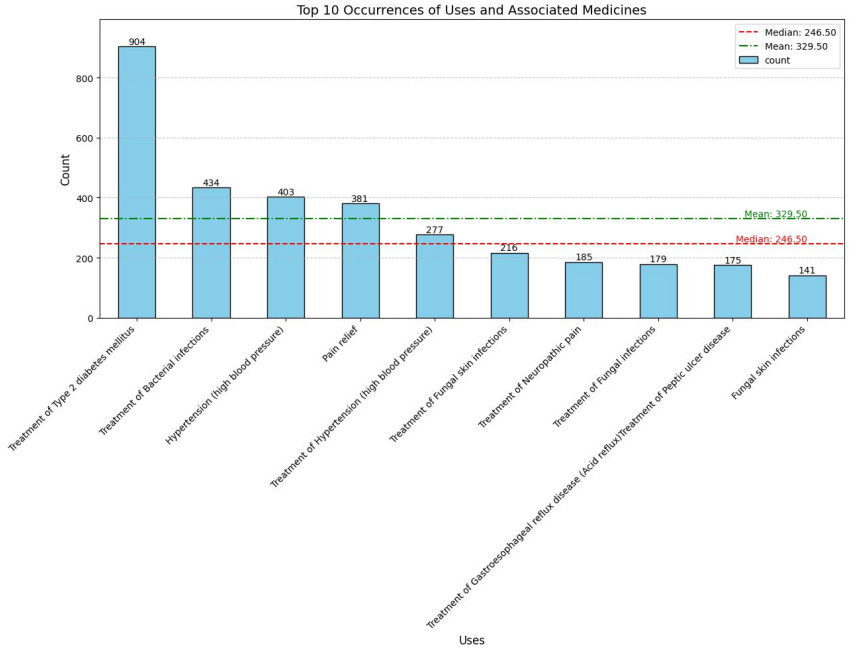
Exploratory Data Analysis

- We found “84” duplicate rows in the data set, Since 84 is very small comparative to total rows(11800) ,we dropped the rows.
- We printed the top 30 salt compositions.



- We are first finding out the unique uses and then printing the top 10 occurring uses and the medicine associated with it.

```
Use: Treatment of Type 2 diabetes mellitus
Number of Occurrences: 904
Medicines: ['Amaryl 1mg Tablet', 'Ajaduo 25mg/5mg Tablet', 'Amaryl 2mg Tablet', 'Azulix 1 MF Tablet PR', 'Azulix 2 MF Tablet PR', 'Amaryl MV 2mg
----------
Use: Treatment of Bacterial infections
Number of Occurrences: 434
Medicines: ['Augmentin 625 Duo Tablet', 'Azithral 500 Tablet', 'Amoxyclav 625 Tablet', 'Azee 500 Tablet', 'Althrocin 500 Tablet', 'Azithral 200 L
----------
Use: Hypertension (high blood pressure)
Number of Occurrences: 403
Medicines: ['Amlosafe 3D Tablet', 'Arbitel-Trio 50 Tablet ER', 'Arbitel-Trio 25 Tablet ER', 'Amlokind-L Tablet', 'Amlokind-H Tablet', 'Amlong-H T
----------
Use: Pain relief
Number of Occurrences: 381
Medicines: ['Altraday Capsule SR', 'Aldigesic-SP Tablet', 'Aceclo Plus Tablet', 'Acemiz Plus Tablet', 'Aldigesic P 100mg/325mg Tablet', 'Acenac-P
----------
Use: Treatment of Hypertension (high blood pressure)
Number of Occurrences: 277
Medicines: ['Arkamin Tablet', 'Amlopres-AT Tablet', 'Amlovas-AT Tablet', 'Amlopres TL Tablet', 'Amlip AT Tablet', 'Amlong-A Tablet', 'Angicam-Bet
----------
Use: Treatment of Fungal skin infections
Number of Occurrences: 216
Medicines: ['Abzorb Antifungal Soap', 'Abzorb Syndet Bar', 'Afitra Cream', 'Afitra-L Cream', 'Canazole Cream', 'Dermicee Cream', 'Ebernet Cream',
----------
Use: Treatment of Neuropathic pain
Number of Occurrences: 185
Medicines: ['Axogurd-SR Tablet', 'Axogurd NT Tablet SR', 'Axinerve-P Tablet SR', 'Axogurd  Capsule', 'Arigaba NT 100 Tablet', 'Arigaba NT Tablet'
----------
Use: Treatment of Fungal infections
Number of Occurrences: 179
Medicines: ['AF 400 Tablet', 'AF 150 Tablet DT', 'AF 200 Tablet', 'Alcros 100 Capsule', 'Alcros 200 Capsule', 'AF 300 Tablet', 'AF 50 Tablet DT',
----------
Use: Treatment of Gastroesophageal reflux disease (Acid reflux)Treatment of Peptic ulcer disease
Number of Occurrences: 175
Medicines: ['Aciloc 150 Tablet', 'Aciloc 300 Tablet', 'Aciloc Injection', 'Acera-D Capsule SR', 'Anleo-DSR Capsule', 'Aciloc Only Oral Liquid', '
```



- We removed the stop words and we lemmatized .
- The column we got after removing the stop words from the uses column was added to the dataframe as Processed_uses.

```python
# Downloading NLTK resources
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')

# Initializing the WordNet Lemmatizer
lemmatizer = WordNetLemmatizer()

# Define additional stop words to remove
additional_stopwords = ['treatment', 'prevention','pain','infection']

# Function to preprocess text
def preprocess_text(text):
    # Convert text to lowercase
    text = text.lower()
    # Tokenize text
    words = word_tokenize(text)
    # Remove stopwords and lemmatize the words
    filtered_words = [lemmatizer.lemmatize(word) for word in words if word not in stopwords.words('english') and word not in additional_stopwords
    # Concatenate words without spaces
    return " ".join(filtered_words)

# Assuming you have a DataFrame 'clean_df' with a column 'Uses'
# Applying the preprocess function to each element in the column 'Uses'
clean_df['Processed_Uses'] = clean_df['Uses'].apply(preprocess_text)

# Show the processed column
print(clean_df[['Uses', 'Processed_Uses']])
```

```python
# Define a function to remove the word "infection" from a string if present
def clean_word(word):
    # Define words to remove
    words_to_remove = ["infection", "pain", "treatment", "prevention", "disorder",
                       "infectiontreatment", "infectionstreatment", "disease",
                       'wellbeing','need','due','pm','benign','certain','due',
                       'functional','severe','affecting']

    # Remove specified words from the word
    for word_to_remove in words_to_remove:
        word = word.replace(word_to_remove, "")

    # Check if "treatment" or "infection" appears at the end of the word
    if word.endswith("treatment"):
        # Remove "treatment" from the end of the word
        word = word[:-len("treatment")]
    elif word.endswith("infection"):
        # Remove "infection" from the end of the word
        word = word[:-len("infection")]

    return word

# # Test the function
# word = "infectiontreatment"
# cleaned_word = clean_word(word)
# print(cleaned_word)  # Output: ""


# Apply the function to the column 'Processed_Uses'
clean_df['Processed_Uses'] = clean_df['Processed_Uses'].apply(clean_word)
```

Vectorization

Tfidf:

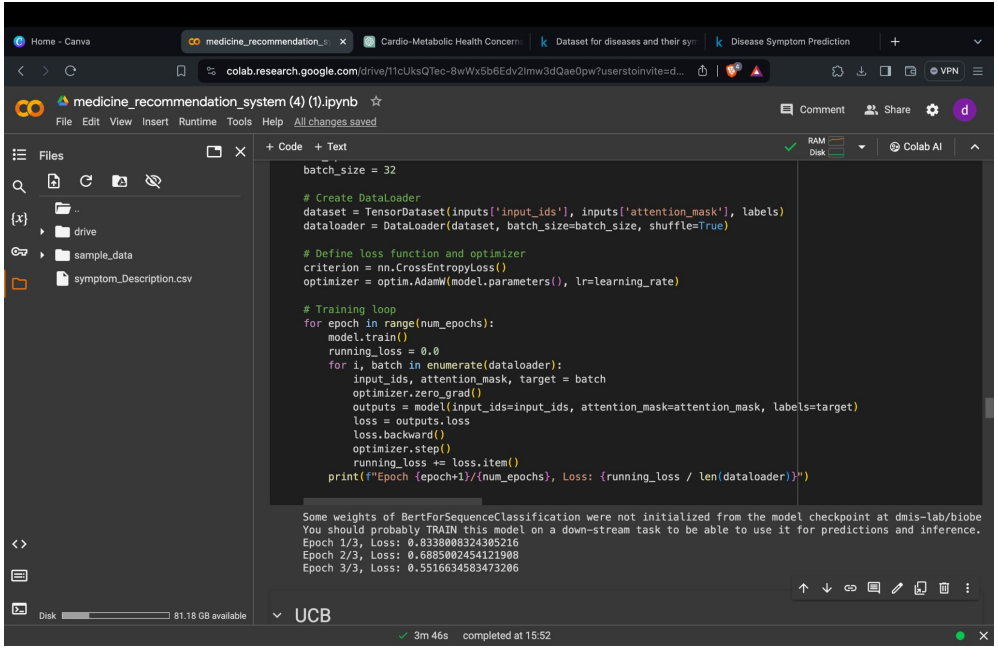
- **Tf-idf measures the term importance in a document by considering its frequency and rarity in a corpus**
- **We applied Tf-idf to the processed uses column of the data frame.**
- **It falls short when given out of vocabulary words, so bio bert was chosen instead.**

Bio-bert

- We used bio bert model to calculate the vectors for the uses column
- It is trained on extensive biomedical literature and datasets.
- It utilizes a transformer-based model with self-attention mechanisms.
- It can be fine-tuned for tasks like named entity recognition and relation extraction in the biomedical domain.

- **We used Symptoms _description.csv to fine tune the bio bert model to get better vector representation .**

- **And we used sub word embeddings for the words that are not in the vocabulary of our pretrained model.**



Cosine Similarity:

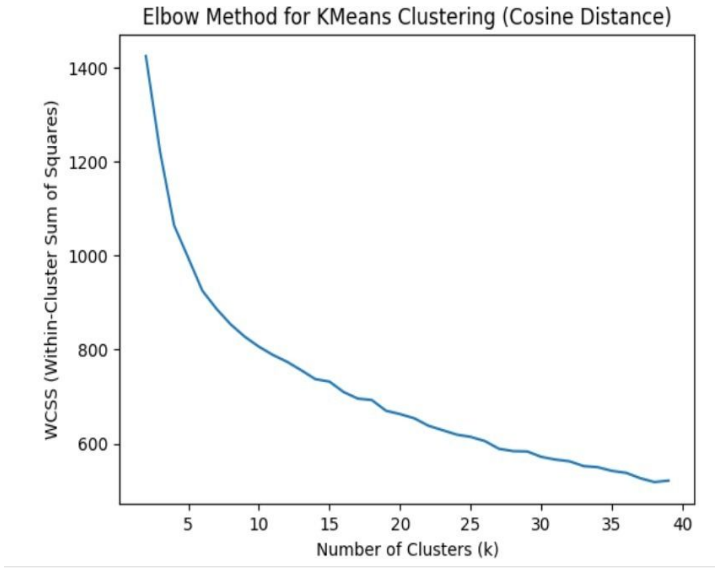
- We get a vector for the new word using the vectorization method discussed above.

- We then compare the vector with already produced vectors for the processed uses column, now we perform cosine similarity and the medicine corresponding to the most similar use is given as prediction.

```
from scipy.spatial.distance import cosine

def cosine_similarity_with_word(new_word, existing_texts):
    # Calculate embedding for the new word
    new_word_embedding = get_text_embedding(new_word)

    # Calculate similarity with each existing text
    similarities = []
    for existing_text_embedding in clean_df['Uses_embedding']:
        # print(len(existing_text_embedding))
        similarity = 1 - cosine(new_word_embedding, existing_text_embedding)
        similarities.append(similarity)

    return similarities
```

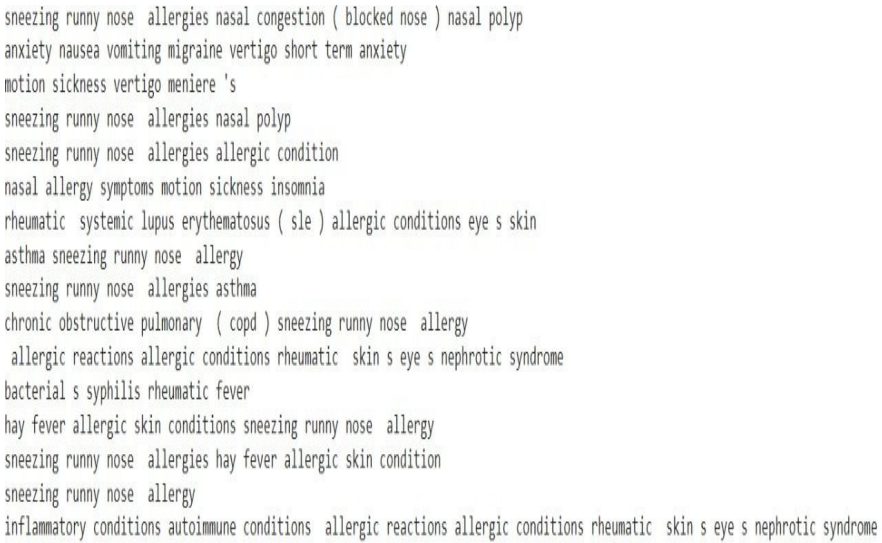
K Means

- We applied K Means on processed uses to cluster them.
- We kept distance as cosine similarity between words as distance when we applied it.
- We used an elbow plot to obtain the number of clusters and then we also observed clusters manually.
- We finalized the number of clusters to be 27 after manually observing them for different number of clusters



Observation of clusters

- From these, we infer that similar diseases fell into same cluster and our vectorization approach is also good



Multi Arm Bandits(UCB)

- In our dataset, we have multiple medicines for a single symptom.
- We used this to identify better medicine among them.
- For reward purpose, we are asking user's feedback for reward and are storing it in a list for each medicine
- We are doing personalized medication and are more focussed on giving better medicines for frequent diseases

$$UCB(i,t) = \hat{\mu}_E(i) + \sqrt{\frac{\alpha \ln T.}{N_i(t).}}$$

- We modified the formula to give weight to reviews since we wish to avoid medicines with bad reviews.

$$UCB(i,t) = \hat{\mu}_E(i) + \sqrt{\frac{\alpha \ln T.}{N_i(t)}}$$


```
# Calculate the upper confidence bound (UCB) for the medicine
ucb = 0.85 * avg_reward + (alpha * math.sqrt(math.log(n)) / n) + 0.15 * (review / 100)
```

	Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses	words_list	Rewards
9	Hicet-AX Tablet SR	Cetirizine (10mg) + Ambroxol (75mg)	Treatment of Cough	https://onemg.gumlet.io/l_watermark_346,w_480,...	Micro Labs Ltd	67	33	0	cough	[cough]	[1]

Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses	words_list	Rewards	Uses_emb
Hicet-AX Tablet SR	Cetirizine (10mg) + Ambroxol (75mg)	Treatment of Cough	https://onemg.gumlet.io/l_watermark_346,w_480,...	Micro Labs Ltd	67	33	0	cough	[cough]	[1, 0]	[0.0012 -0.0070 0.0014

```
ean_df[clean_df['Medicine Name'] == 'Lincotus DX  Dry Cough Formula']
```

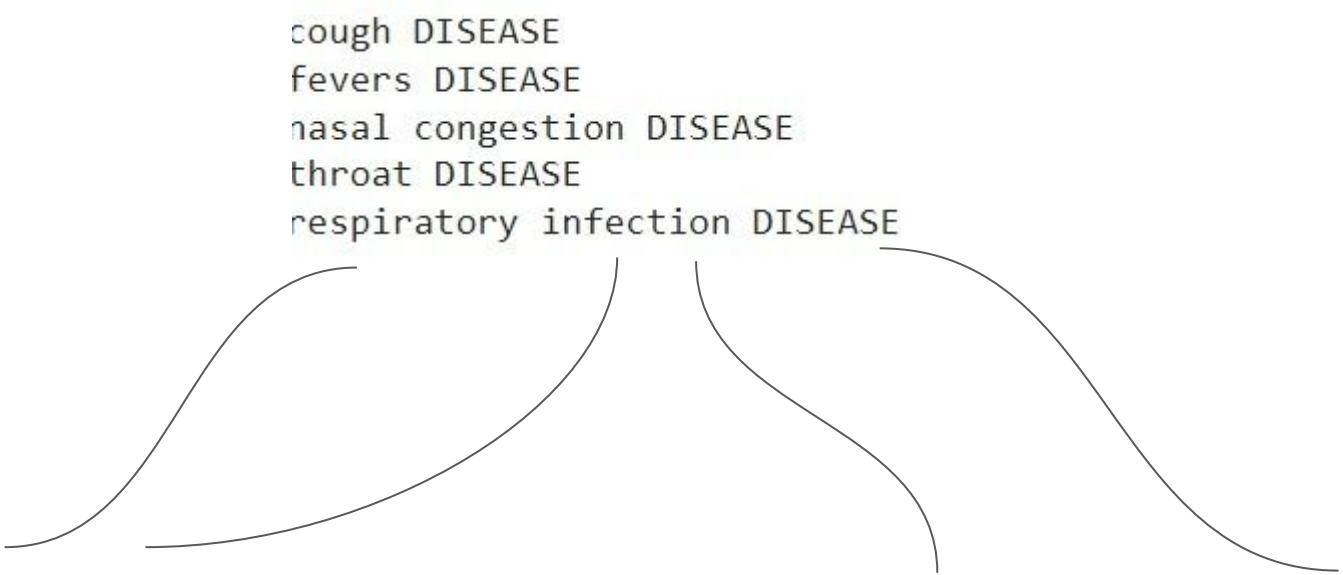
	Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses	words_list	Rewards
07	Lincotus DX Dry Cough Formula	Ambroxol (15mg/5ml) + Dextromethorphan Hydrobr...	Treatment of Cough	https://onemg.gumlet.io/l_watermark_346,w_480,...	Macleods Pharmaceuticals Pvt Ltd	58	17	25	cough	[cough]	[1]

Reward list getting updated and medicine recommendation changing when the user gives feedback as 0.

Predictions

The patient suffers from a persistent cough DISEASE that refuses to subside, along with a stuffy nose filled with congestion. To make matters worse, they experience recurring fevers DISEASE , with their temperature spiking especially at night. As if that weren't enough, they also describe a bothersome nasal congestion DISEASE , a sensation of mucus constantly draining down the back of their throat DISEASE . These symptoms paint a picture of a potential respiratory infection DISEASE that warrants further investigation.

Extracted symptoms



Pain relief Fever

Nasal congestion (blocked nose)

Treatment of Cough

Treatment of Respiratory tract infection

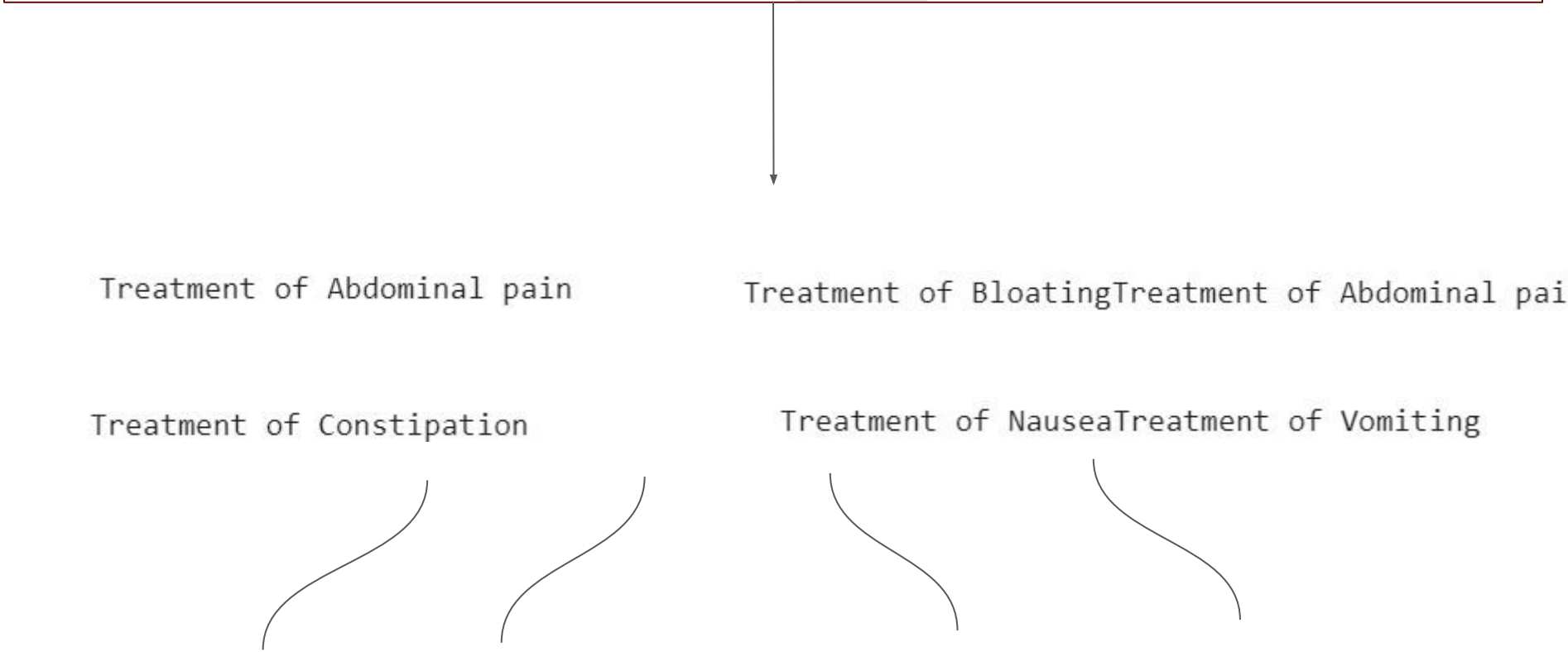
Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses
Hicet-AX Tablet SR	Cetirizine (10mg) + Ambroxol	Treatment of Cough	https://onemg.gumlet.io/l_watermark_346,w_480,...	Micro Labs Ltd	67	33	0	cough

Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses
Xylone Adult Nasal Spray	Xylometazoline (0.1% w/v)	Nasal congestion (blocked nose)	https://onemg.gumlet.io/l_watermark_346,w_480,...	West-Coast Pharmaceutical Works Ltd	50	17	33	nasal congestion (blocked nose)

Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses
Hhcepo-CV Tablet	Cefpodoxime Proxetil (200mg) + Clavulanic Acid...	Treatment of Respiratory tract infection	https://onemg.gumlet.io/l_watermark_346,w_480,...	Hegde and Hegde Pharmaceutical LLP	100	0	0	respiratory tract

Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses
PA 12 Tablet SR	Paracetamol (300mg) + Paracetamol (700mg)	Pain relief Fever	https://onemg.gumlet.io/l_watermark_346,w_480,...	Lincoln Pharmaceuticals Ltd	67	0	33	relief fever

The patient complains of persistent abdominal pain DISEASE , particularly in the upper abdomen. This discomfort often worsens on an empty stomach or at night and may be temporarily relieved by antacids CHEMICAL . Along with the pain DISEASE , the patient experiences frequent bloating DISEASE . Additionally, there is occasional constipation DISEASE and nausea DISEASE . These symptoms have been ongoing for several months and tend to worsen after consuming spicy or fatty DISEASE foods.



Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses
Emesafe Oral Spray	Ondansetron (2mg)	Treatment of NauseaTreatment of Vomiting	https://onemg.gumlet.io/l_watermark_346,w_480,...	Wallace Pharmaceuticals Pvt Ltd	100	0	0	nausea vomiting

Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses
All Klear Granules	Lactitol (10gm) + Ispaghula (3.5gm)	Treatment of Constipation	https://onemg.gumlet.io/l_watermark_346,w_480,...	Tablets India Limited	100	0	0	constipation

Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses
Dimol 40 Tablet	Dimethicone (40mg)	Treatment of BloatingTreatment of Abdominal pain	https://onemg.gumlet.io/l_watermark_346,w_480,...	Wallace Pharmaceuticals Pvt Ltd	47	29	24	bloating abdominal

Medicine Name	Composition	Uses	Image URL	Manufacturer	Excellent Review %	Average Review %	Poor Review %	Processed_Uses
Fenadol-DV Tablet	Drotaverine (80mg) + Aceclofenac (100mg)	Treatment of Abdominal pain	https://onemg.gumlet.io/l_watermark_346,w_480,...	Dolvis Bio Pharma Pvt Ltd	67	33	0	abdominal

Thank you

IMT2021035, Kushal dasari
IMT2021042, Raghunadh
IMT2021034, Prachoday.