AI BASED MEDICAL DIAGNOSIS WITH MEDICINE RECOMMENDATIONS

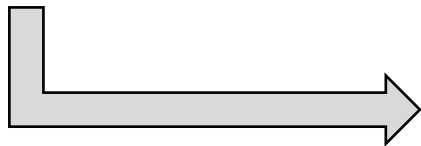By: Janvi Patel

INTRODUCTION

- Accurate diagnosis is a fundamental aspect of global healthcare systems. In the US, approximately 5% of outpatients receive an incorrect diagnosis, with errors being particularly common for serious medical conditions, and carrying the risk of serious patient harm.

- Diagnosing disease can be a tricky task as there are many common symptoms related to different diseases and accurately predicting the disease is very important here.

- The conventional method used for diagnosing disease is somewhat error prone. Using the AI technique to diagnosis the disease reduces the chances of errors and increases the accuracy of identifying the disease based on the symptoms and using the knowledge

DATASET FOR PREDICTION

- Disease Symptom Dataset from Kaggle : 4920, 18

- After preprocessing : 4920, 8

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5	Symptom_6	Symptom_7
count	4920	4920	4920	4920	4572	3714	2934	2268
unique	41	34	48	54	50	38	32	26
top	Fungal infection	vomiting	vomiting	fatigue	high_fever	headache	nausea	abdominal_pain
freq	120	822	870	726	378	348	390	264

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5	Symptom_6	Symptom_7
0	Fungal infection	itching	skin_rash	nodal_skin_eruptions	dischromic _patches	0	0	0
1	Fungal infection	skin_rash	nodal_skin_eruptions	dischromic _patches	0	0	0	0
2	Fungal infection	itching	nodal_skin_eruptions	dischromic _patches	0	0	0	0
3	Fungal infection	itching	skin_rash	dischromic _patches	0	0	0	0
4	Fungal infection	itching	skin_rash	nodal_skin_eruptions	0	0	0	0

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5	Symptom_6	Symptom_7
0	Fungal infection	1	3	4	0	0	0	0
1	Fungal infection	3	4	0	0	0	0	0
2	Fungal infection	1	4	0	0	0	0	0
3	Fungal infection	1	3	0	0	0	0	0
4	Fungal infection	1	3	4	0	0	0	0

DATASET FOR RECOMMENDATION

- Unique Diseases: 708

- Unique Drugs: 2675

drugName	Disease	review	rating	usefulCount
Voltaren	zen Shoulde	"The medication has been somewhat helpful with...	5	22
Diclofenac	zen Shoulde	"The medication has been somewhat helpful with...	5	22
Aleve	zen Shoulde	"One every 8 hours dulls the pain but sufficie...	5	31
Aleve	zen Shoulde	"Very little relief. I finished PT and after ...	2	6
Relafen	zen Shoulde	"Very helpful for my frozen shoulder pain with...	9	38

Drug	Disease	Review	Rating	UsefulCount	Symptoms
Sulfamethoxazole / trimethoprim	urinary tract infection	"I have bad side effects from all antibiotics,...	1	17	[burning_micturition, bladder_discomfort, f...
Levofloxacin	urinary tract infection	"This last Monday I was detected by my doctor ...	3	24	[burning_micturition, bladder_discomfort, f...
Sulfamethoxazole / trimethoprim	urinary tract infection	"I am currently suffering from recurring cysti...	1	7	[burning_micturition, bladder_discomfort, f...
Nitrofurantoin	urinary tract infection	"I have been taking 4 a day for last 6 days, 2...	5	10	[burning_micturition, bladder_discomfort, f...
Macrobid	urinary tract infection	"I wish I would have read these reviews before...	1	28	[burning_micturition, bladder_discomfort, f...

FINAL DATASET

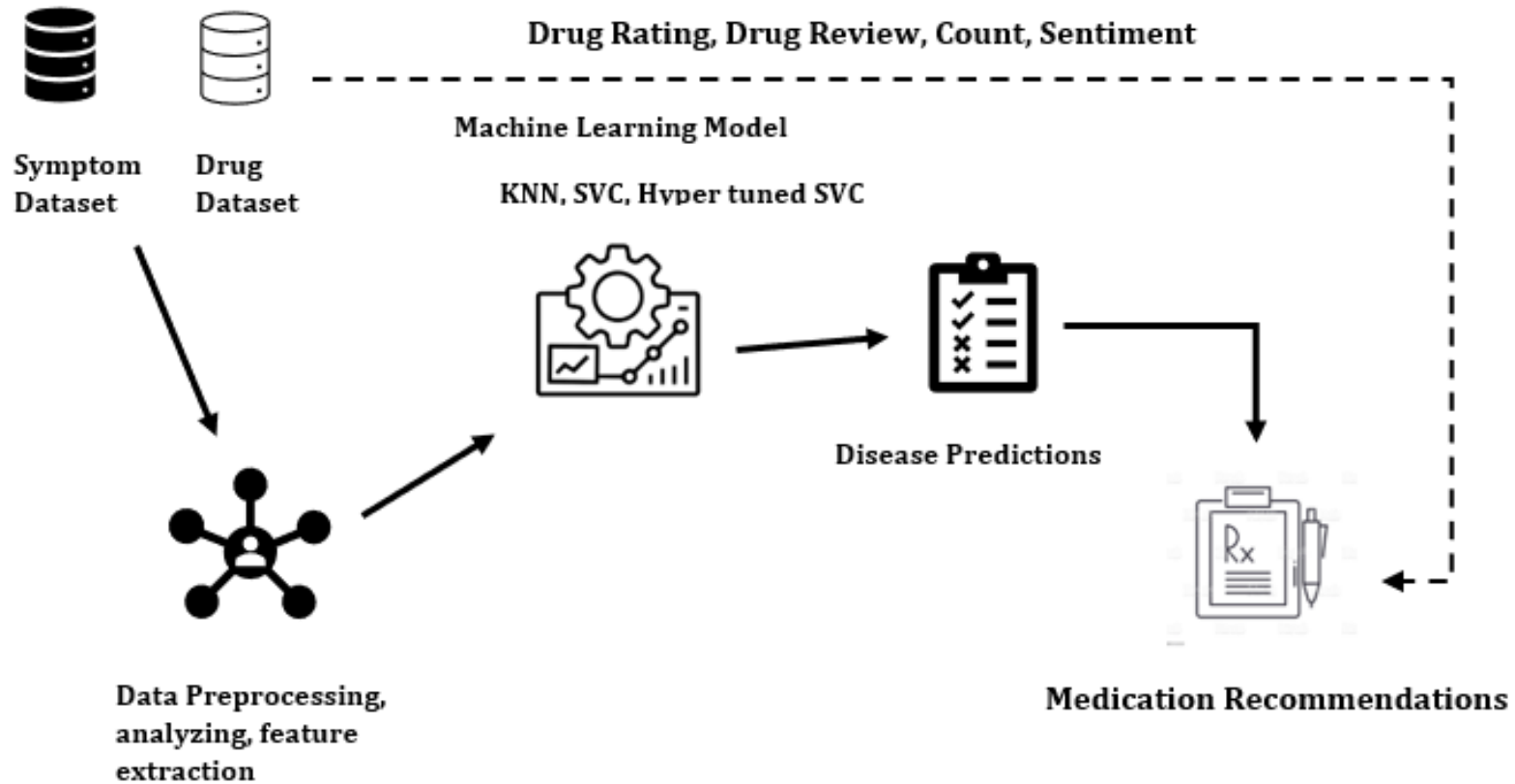| Drug	Rating_Weighted_Avg	Disease	Symptoms	Review	Sentiment	Rating	UsefulCount
Cefixime	10.0	urinary tract infection	[' burning_micturition', ' bladder_discomfort'...	"My daughter was born with urinary reflux, the...	Negative	10	7
Doribax	10.0	urinary tract infection	[' burning_micturition', ' bladder_discomfort'...	"It is proved itself it is an excellent drug o...	Positive	10	1
Doripenem	10.0	urinary tract infection	[' burning_micturition', ' bladder_discomfort'...	"It is proved itself it is an excellent drug o...	Positive	10	1
Lactobacillus acidophilus	10.0	urinary tract infection	[' burning_micturition', ' bladder_discomfort'...	"This has been a miracle for me, and it was by...	Negative	10	189
Macrodantin	10.0	urinary tract infection	[' burning_micturition', ' bladder_discomfort'...	"Started experiecing pain peeing and went for ...	Positive	10	28

```
No. of unique diseases:  12

array(['acne', 'migraine', 'gerd', 'psoriasis', 'heart attack',
       'urinary tract infection', 'pneumonia', 'malaria', 'hepatitis c',
       'hyperthyroidism', 'gastroenteritis', 'hepatitis b'], dtype=object)
```

```
No. of unique drugs:  286

Output exceeds the size limit. Open the full output data in a text editor
array(['Absorica', 'Acanya', 'Accutane',
        'Acetaminophen / aspirin / caffeine',
        'Acetaminophen / aspirin / caffeine / salicylamide', 'Aciphex',
        'Acitretin', 'Acnex', 'Aczone', 'Adalimumab', 'Adapalene',
        'Adapalene / benzoyl peroxide', 'Adoxa', 'Aldactone',
        'Almotriptan', 'Altabax', 'Altace',
        'Aluminum hydroxide / magnesium hydroxide / simethicone', 'Amerge',
        'Ammonium lactate / halobetasol', 'Amnesteem', 'Amoxicillin',
        'Amoxicillin / clavulanate', 'Amoxil', 'Amrix', 'Aspirin',
        'Aspirin / butalbital / caffeine', 'Atovaquone',
        'Atovaquone / proguanil', 'Atralin', 'Augmentin', 'Avelox',
```

SYSTEM MODEL

Symptom Dataset

Drug Dataset

Drug Rating, Drug Review, Count, Sentiment

Machine Learning Model

KNN, SVC, Hyper tuned SVC

Disease Predictions

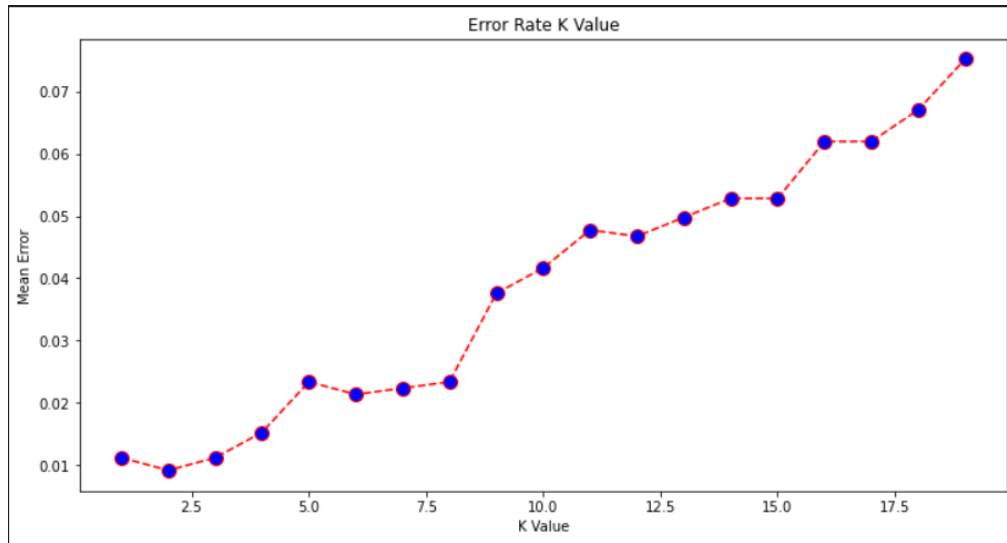Medication Recommendations

Data Preprocessing, analyzing, feature extraction

KNN MODEL

- Group all the similar symptoms so that it will effectively predict the associate disease.

- Since k nearest neighbor works on the principle of proximity i.e how similar the data points are this model will helps us to effectively predict the disease as the symptoms are closely related to each other of a disease.

- For this model distance measure between symptoms is calculated and only those symptoms which have smaller distances will be grouped for predicting the associate disease.

KNN MODEL RESULTS

	precision	recall	f1-score	support
(vertigo) Paroymsal Positional Vertigo	1.00	1.00	1.00	18
AIDS	1.00	1.00	1.00	30
Acne	1.00	1.00	1.00	24
Alcoholic hepatitis	1.00	1.00	1.00	25
Allergy	1.00	0.96	0.98	24
Arthritis	1.00	1.00	1.00	23
Bronchial Asthma	1.00	1.00	1.00	33
Cervical spondylosis	0.96	1.00	0.98	23
Chicken pox	1.00	1.00	1.00	21
Chronic cholestasis	1.00	1.00	1.00	15
Common Cold	1.00	1.00	1.00	23
Dengue	1.00	1.00	1.00	26
Diabetes	1.00	1.00	1.00	21
Dimorphic hemmorhoids(piles)	1.00	1.00	1.00	29
Drug Reaction	1.00	1.00	1.00	24
Fungal infection	1.00	1.00	1.00	19
GERD	1.00	1.00	1.00	28
Gastroenteritis	0.93	1.00	0.96	25
Heart attack	1.00	1.00	1.00	23
Hepatitis B	1.00	1.00	1.00	27
Hepatitis C	1.00	1.00	1.00	26
Hepatitis D	0.96	0.96	0.96	23
Hepatitis E	0.93	0.90	0.91	29
...				
accuracy			0.99	984
macro avg	0.99	0.99	0.99	984
weighted avg	0.99	0.99	0.99	984

KNN Model

F1-score: 98.89

Accuracy: 98.88

SVC MODEL

- SVM tries to classify cases by finding a separating boundary called hyperplane.

- This technique has recently been used to improve methods for detecting diseases in clinical settings.

- SVM is a powerful supervised algorithm that works best on smaller datasets but on complex ones.

SVC MODEL RESULTS

	precision	recall	f1-score	support
(vertigo) Paroymsal Positional Vertigo	0.94	0.94	0.94	32
AIDS	1.00	1.00	1.00	39
Acne	1.00	1.00	1.00	41
Alcoholic hepatitis	1.00	1.00	1.00	36
Allergy	0.88	1.00	0.93	35
Arthritis	1.00	1.00	1.00	36
Bronchial Asthma	0.96	0.98	0.97	44
Cervical spondylosis	1.00	0.91	0.95	32
Chicken pox	1.00	1.00	1.00	35
Chronic cholestasis	1.00	0.93	0.97	30
Common Cold	1.00	1.00	1.00	31
Dengue	1.00	1.00	1.00	40
Diabetes	0.89	1.00	0.94	33
Dimorphic hemmorhoids(piles)	0.98	1.00	0.99	45
Drug Reaction	1.00	1.00	1.00	35
Fungal infection	1.00	1.00	1.00	28
GERD	1.00	1.00	1.00	41
Gastroenteritis	0.81	0.95	0.87	40
Heart attack	1.00	0.69	0.82	36
Hepatitis B	1.00	1.00	1.00	46
Hepatitis C	1.00	1.00	1.00	40
Hepatitis D	0.94	0.86	0.90	35
Hepatitis E	0.95	0.79	0.86	47
...				
accuracy			0.97	1476
macro avg	0.97	0.97	0.97	1476
weighted avg	0.97	0.97	0.97	1476

SVC Model
F1-score: 96.88 %
Accuracy: 96.88 %

GRIDSEARCHCV MODEL

```python
param_grid = {'C': [0.1, 1, 10, 100],
              'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
              'gamma':['scale', 'auto'],
              'kernel': ['linear']}
```

```
Fitting 5 folds for each of 8 candidates, totalling 40 fits
Best Parameters: {'C': 10, 'gamma': 'scale', 'kernel': 'linear'}
```

```
GridSearchCV Model

F1-score:  97.0

Accuracy:  97.02 %
```

	precision	recall	f1-score	support
(vertigo) Paroymsal Positional Vertigo	0.88	0.96	0.92	24
AIDS	1.00	1.00	1.00	24
Acne	1.00	1.00	1.00	24
Alcoholic hepatitis	1.00	1.00	1.00	24
Allergy	0.83	1.00	0.91	24
Arthritis	1.00	1.00	1.00	24
Bronchial Asthma	0.96	1.00	0.98	24
Cervical spondylosis	1.00	1.00	1.00	24
Chicken pox	1.00	1.00	1.00	24
Chronic cholestasis	1.00	0.88	0.93	24
Common Cold	1.00	1.00	1.00	24
Dengue	1.00	1.00	1.00	24
Diabetes	0.92	1.00	0.96	24
Dimorphic hemmorhoids(piles)	1.00	1.00	1.00	24
Drug Reaction	1.00	1.00	1.00	24
Fungal infection	1.00	1.00	1.00	24
GERD	1.00	1.00	1.00	24
Gastroenteritis	0.88	0.96	0.92	24
Heart attack	1.00	0.83	0.91	24
Hepatitis B	1.00	1.00	1.00	24
Hepatitis C	1.00	1.00	1.00	24
...				
accuracy			0.97	984
macro avg	0.97	0.97	0.97	984
weighted avg	0.97	0.97	0.97	984

RESULTS

```
--------------------------------------------------------------------
Symptoms:  ['continuous_sneezing', 'scurring', 'skin_peeling', 'silver_like_dusting', 0, 0, 0]
--------------------------------------------------------------------
The predicted disease is: Acne
--------------------------------------------------------------------
Recommended drugs for Acne are:  ['Aldactone' 'BenzEFoam' 'Benzoyl peroxide / sulfur']
--------------------------------------------------------------------
```

```
----------------------------------------------------------------------------------------
Symptoms:  ['muscle_wasting', 'yellowish_skin', 'dark_urine', 'throat_irritation', 'drying_and_tingling_lips', 'muscle_pain', 'altered_sensorium']
----------------------------------------------------------------------------------------
The predicted disease is: Migraine
----------------------------------------------------------------------------------------
Recommended drugs for Migraine are:  ['Methergine' 'Imitrex Nasal']
----------------------------------------------------------------------------------------
```

CONCLUSIONS

The prediction of the diseases largely affect the dataset available. Due to which the model was only able to predict only 12 diseases out of 41 diseases.

For prediction of diseases the KNN model gave the highest accuracy of 98% as it efficiently correlated the symptoms for the diseases.

For future work a better dataset can be created with the help of a domain expert to check whether the model still gives best accuracy.