SYMPTOMS MATCHER
AND
DOCTOR RECOMMENDATION
MEDICAL ASSISTANT

PROJECT REPORT 1

By: SHREYAS OM

Date: 21st February 2023

Abstract:

The Covid-19 pandemic has brought about a lasting impact on various sectors of society, ranging from personal routines to major industries. One of the most significant changes has been observed in the field of medicine. With the growing demand for reliable medical assistance, many individuals have become hesitant to frequently visit hospitals due to various inconveniences, such as unnecessary tests and the potential risk of exposure. To address this concern, there is a need for a solution that simplifies and streamlines medical assistance, allowing individuals to receive care from the comfort of their own homes, utilizing the power of AI and machine learning.

Problem Statement:

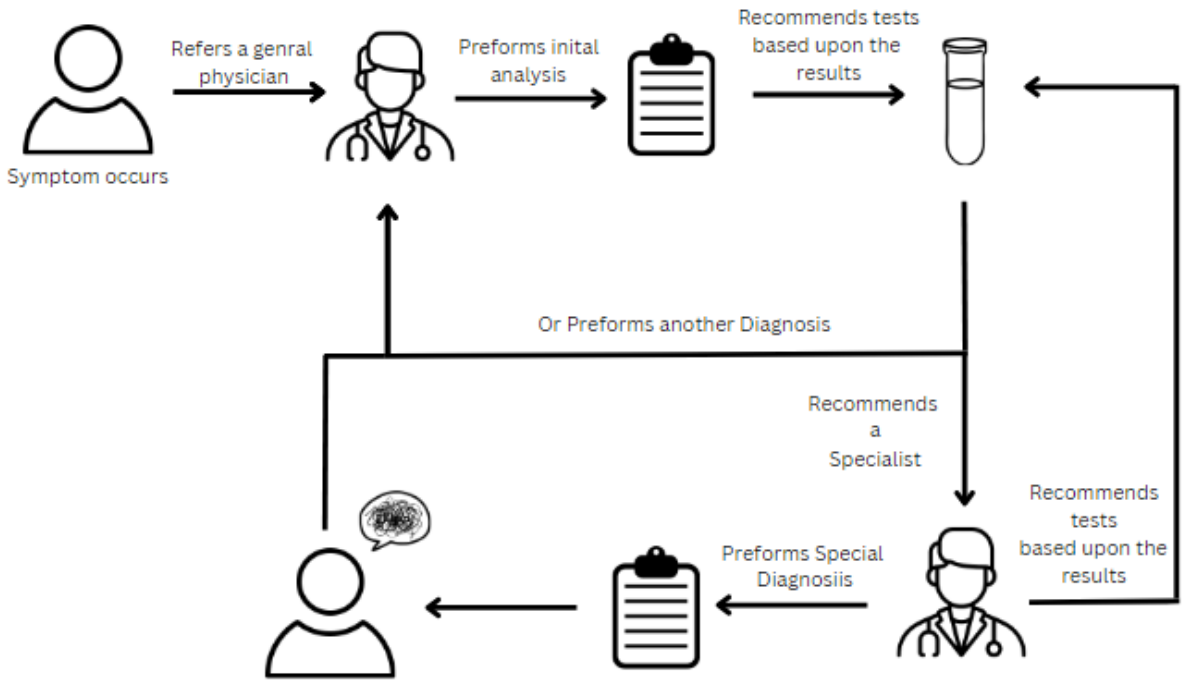
The current medical system especially in India as it stands is a very lengthy process involving various trips to doctors and many tests and visits to medical centers.

Whenever someone suffers from any kind of symptoms, first they must refer to a general physician. The physician performs an initial analysis of the patient's symptoms. Then some tests are performed if the problem persists, followed by another visit to the general physician, who either performs another analysis or depending on that recommends them to a specialized doctor for the said disease.

This redundancy off puts people from receiving proper care and many required checkups and tests are put on hold. A part of this is also because many peoples avoid visiting hospitals and medical centers often due to the risk of exposure to pathogens and diseases.

Added on top of this comes the extravagant fees of big hospitals and doctors associated with them. People either don't know or cannot trust standalone doctors and clinics and this creates a divide between doctors required to affiliate themselves with hospitals and patients forced to pay high fees for them.

This divide in the system can be solved by automating a major part of the system and applying AI to it to analyze patient data to recommend appropriate doctors to them based on their specific diseases and symptoms. This would cut short the process of repetitive visits to hospitals and give people quick and concrete results.



Customer Need:

Currently, whenever a patient comes up with some symptoms, they first visit a general physician in their vicinity, the general physician does some initial analysis of the symptoms and recommends some tests, and based on the test results they are given some medicine if the symptoms persist the patient again visits the general physician who either recommends another set of tests or finally recommends them to a specialist, where again the cycle continues. This is a very redundant but true face of the current medical system which makes it difficult for patients especially the ones with diseases that are not as apparent.

The patient needs a system in place that is open to use and free of cost, and that can analyze their symptoms appropriately and based on that recommend them an appropriate specialized doctor. This would cut short the constant cycle of visits to the hospital and general physician and hasten up the treatment process making the results more efficient.

Market Need:

The lack of any assistance through AI and ML in the medical industry results in the overburdening of doctors as they have to work from scratch from jotting down the symptoms to analyzing them. Using this system doctors won't have to work with scratch data and would have an initial analysis of the patient.

Also, many local doctors are not affiliated with a big hospital and run individual clinics, such doctors sometimes go unnoticed. This way they also can gain valued patients and an opportunity to come into the limelight.

They require a system that can let patients know as well as trust their clinics and their doctors and allow them to gain access to such patients. This would help them establish their medical centers and at the same time make healthcare more accessible to all.

Target Specification:

The target of this service would be every single patient suffering from any kind of disease. But initially, this service would appeal the most to individuals who are skeptical about visiting medical centers with the risk of exposure to various diseases. Also, people who want to hasten up the procedure without sacrificing any due diligence or causing any kind of mistreatment.

In terms of doctors rather than aiming at a big organization, it will recommend local doctors and doctors with specialization in various niches of medical sectors. This will also take off load from the general physicians and reduce the redundancy in the medical sector.

External Research

- "Artificial Intelligence in Healthcare: Past, Present and Future" by Fei Jiang et al. (2019): The paper concludes that AI has the potential to revolutionize healthcare by improving diagnosis accuracy, predicting disease progression, and developing more effective treatments. However, challenges such as data privacy and ethical considerations must be addressed to ensure that AI is used responsibly and transparently.
- "Machine Learning for Medical Diagnosis: History, State of the Art and Perspective" by Andrew L. Beam et al. (2020): The paper concludes that machine learning has the potential to significantly improve medical diagnosis accuracy. However, challenges such as ensuring data quality and addressing ethical considerations must be addressed to ensure that machine learning is used responsibly and effectively.
- "Deep Learning for Healthcare: Review, Opportunities and Challenges" by Yuji Fan et al. (2020): The paper concludes that deep learning has

significant potential to improve healthcare outcomes, particularly in areas such as medical image analysis and disease diagnosis. However, challenges such as data privacy and regulatory considerations must be addressed to ensure that deep learning is used responsibly and effectively.

- "Artificial Intelligence in Healthcare: A Comprehensive Review" by John McCarthy et al. (2020): The paper concludes that AI has the potential to significantly improve healthcare outcomes in areas such as personalized medicine and clinical decision-making. However, challenges such as data privacy and ethical considerations must be addressed to ensure that AI is used responsibly and effectively.
- "Machine Learning in Medicine: A Practical Introduction" by Alvin Rajkumar et al. (2018): The paper concludes that machine learning has significant potential to improve healthcare outcomes, particularly in areas such as predicting hospital readmissions and identifying patients at risk of sepsis. However, challenges such as ensuring data quality and addressing ethical considerations must be addressed to ensure that machine learning is used responsibly and effectively.

Bench Marking

Following are a few of the existing products in the healthcare sector that apply machine learning and Ai to assist individuals

- **Babylon Health**(https://www.babylonhealth.com/): Babylon Health uses AI to provide a symptom checker and virtual consultations with healthcare professionals. Users can input their symptoms into the app, and the AI algorithm provides a list of possible conditions that may be causing the symptoms. The app also allows users to book virtual consultations with healthcare professionals if needed. Drawback: The product doesn't show nearby professional doctors and neither makes face to face appointments makes this product sub-optimal due to lack of due diligence. Online consultation while good for first opinions do not

- **Ada Health**(https://ada.com/): Ada Health uses AI to provide personalized health assessments based on a user's symptoms. Users input their symptoms into the app, and the AI algorithm provides a list of

possible conditions and personalized health recommendations based on the user's age, gender, medical history, and other factors. Drawback: Ada Health is a good product for information and personalization of patient data, but in the health, sector recommending anything related to health without consulting a trained professional can be dangerous and in some cases lethal to the user.

- **Infermedica**(https://infermedica.com/): Infermedica provides an AI-powered platform for healthcare providers to triage and diagnose patients based on their symptoms. The platform uses AI to analyze patient data, identify potential conditions, and provide diagnostic suggestions to healthcare providers.
  Drawback: The major drawback of this tool is that it is not open to all. Infermedica provides APIs to hospitals that they can integrate into their systems for their patients to use. This makes them lack in terms of accessibility.

- **Buoy Health**(https://www.buoyhealth.com/): Buoy Health uses AI to provide a chatbot-based symptom checker. Users input their symptoms into the chatbot, and the AI algorithm provides personalized healthcare recommendations based on the user's symptoms and medical history. Drawback: Like Ada Health, recommending healthcare products to patients without due diligence by a medical professional. As a result other than being a good source of information the practicality of the product is reduced.

- **K Health**(https://www.khealth.ai/): K Health uses AI to provide a symptom checker and virtual consultations with healthcare professionals. Users input their symptoms into the app, and the AI algorithm provides a list of possible conditions that may be causing the symptoms. The app also allows users to book virtual consultations with healthcare professionals if needed.
  Drawback: Like Babylon health, the lack of face-to-face consulting with any doctors can create further problems due to a lack of due diligence. Face-to-face consultancy is very important in terms of the health sector.

- **Sensely**(https://www.sensely.com/): Sensely provides a virtual assistant that uses AI to provide personalized health assessments based on a user's symptoms. Users interact with the virtual assistant via text or voice commands, input their symptoms, and receive personalized healthcare recommendations based on the AI algorithm's analysis. Drawback: Again, having just an AI assistant to give virtually generated information without required diligence can be counterproductive to the purpose of having a healthcare system in place.

Concept Development

The service/product works in the following manner:

1. The patient accesses the service on any browser on either their computer or phone.
2. They enter their symptoms through a simple form asking them for their Age, gender, locality, and other symptoms such as body temperature, aches, and other details.
3. The algorithm analyzes the given details and makes a prediction on which kind of disease the user might have.
4. The output would also include some details about the disease so that the patient can have insights into it.
5. Finally based on the disease and symptoms the algorithm would recommend what kind of doctor the patient should see.
6. The algorithm based on the user's location would recommend a list of doctors around them and show the ratings and reviews that previous patients.
7. The user would be forwarded to an appointment page for the selected doctor.

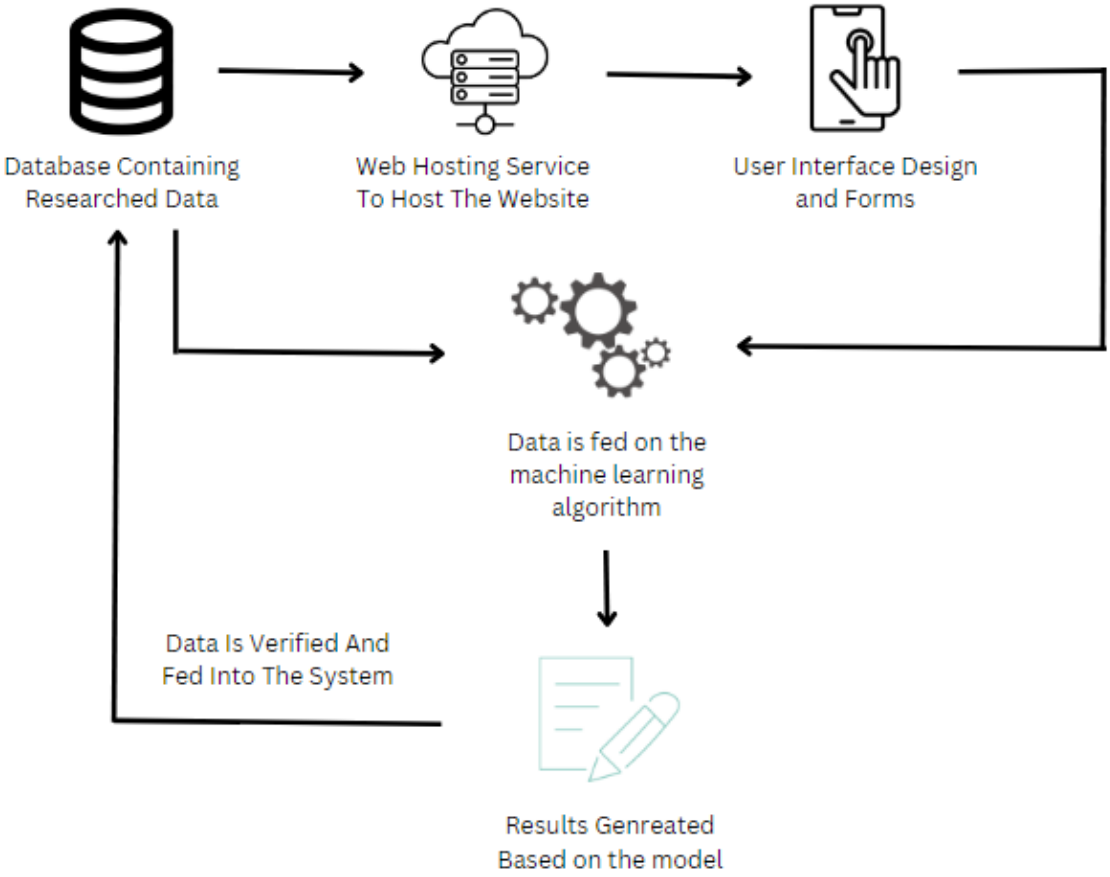
8. On the doctor's side, when an appointment is booked the doctor would have access to the details entered by the patient.
9. The doctor after analyzing the patient would decide on whether the analysis by the ML algorithm was right or wrong, and if wrong what actual disease was the patient suffering from.
10. This would help make the model better in the future.

Business Model:

The monetization system would work by taking a cut from the doctor's consultation fee, and by taking a cut from the sale of other services mentioned in the future scope.

This method can be expanded to other methodologies and cover more aspects of the medical sector (refer to future scope section)

Concept Implementation

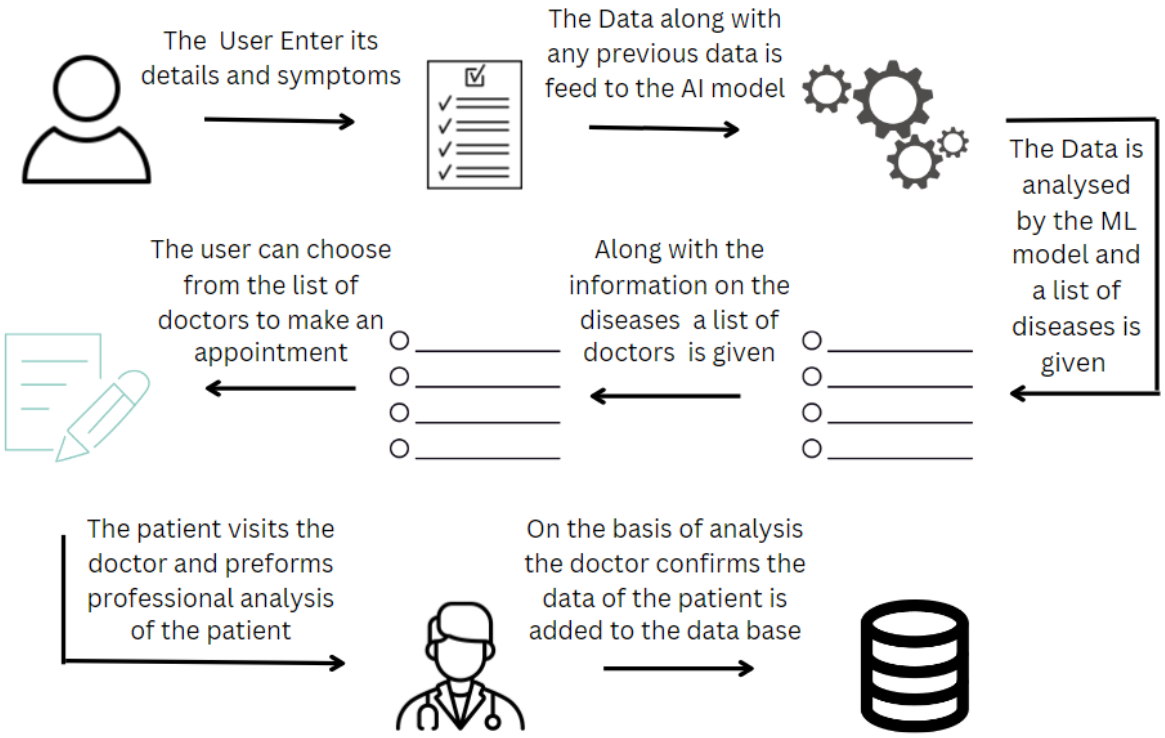


Product Prototype

- **How does it work:**

The product works on a logic regression model that classifies the data on the preexisting diseases in the database. When any patient enters their data, and it is verified by the doctor (or the data is verified by identifying the actual disease the patient is suffering from) the data is entered into the database which increases its efficiency even more over time.

The doctor recommendation system works simply by identifying the locality a patient is in, generating the list of the doctors that are in the area and showing the list of doctors specific for the diseases (for example. Orthopedic, pediatrician, etc.).



- **Database:**

The database includes the disease along with the Boolean value for applicable symptoms on whether the person has the symptom. The initial source of this is self-research using webmd.com. As the model is put in place self-reliant data is replaced with professionally verified data by actual doctors.

```
In [34]: dataset
```

Out[34]:

	Diseases	Joint_Ache	Fever	Fatigue	Irritability	Loss of apppetite	Stomach_Ache	Headache	Sore_Throat	Sneeezing	...	Bleeding	Painful Urination	Rapid Heart Rate	C in pr
0	Chickenpox	0	1	1	1	1	0	0	0	0	...	0	0	0	
1	Cysticercosis	0	1	1	0	0	0	0	0	0	...	0	0	0	
2	Hantavirus Infections	0	1	1	0	0	0	0	0	0	...	0	0	0	
3	Human Immunodeficiency Virus	0	1	1	0	0	0	0	0	0	...	0	0	0	
4	Novel Coronavirus	0	1	1	0	0	0	0	0	0	...	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
7537	West Nile Virus	0	1	0	0	0	0	1	0	0	...	0	0	0	
7538	Whooping cough	0	1	0	0	0	0	0	0	0	...	0	0	0	
7539	Yellow Fever	0	1	1	0	0	0	0	0	0	...	0	0	0	
7540	Yersiniosis	0	1	0	0	0	1	0	0	0	...	0	0	0	
7541	Zika Virus	1	1	1	0	0	0	1	0	0	...	0	0	0	

7542 rows × 39 columns

- **Machine Learning Model:**

The machine learning model used in this case is random forest. A couple of reasons to opt for this model.

a) This model can never be overfitted to the data. Random forest can always be either under fitted or perfectly fitted to the given data.
b) Regularization techniques are not required.
c) One of the most efficient and best classification techniques.

- **Required Frameworks:**

1) A server system such as amazon AWS or Google Cloud.
2) Google map API for accessing user and doctor locations.
3) A website domain for hosting the application.
4) A web developer, a marketing team, a team of data scientists and machine learning engineers and medical researchers.
5) Also, it would require mailing or in-person asking doctors to sign up for the service either by giving them access to the platform free of cost for a brief period.

Future Scope:

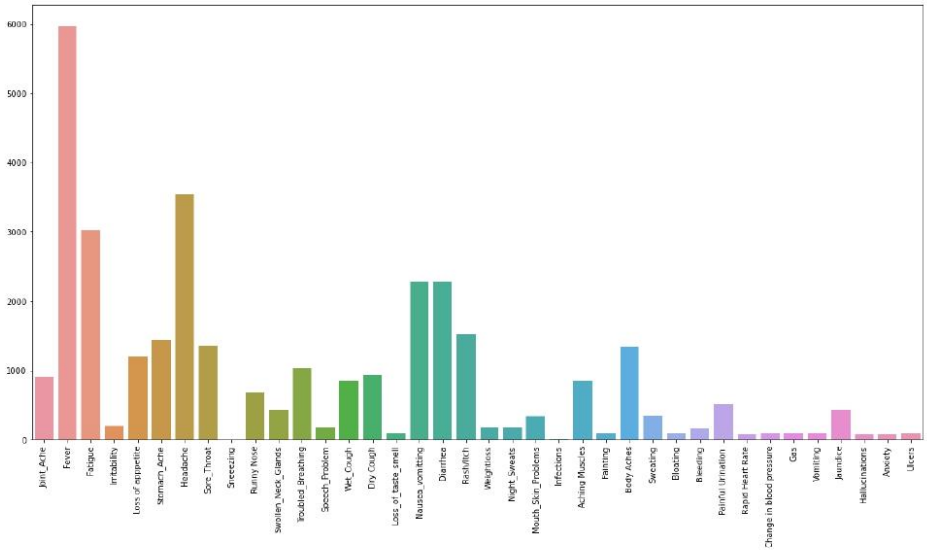
This project right now covers just the part of initial analysis and recommending doctors in the vicinity having their expertise in the said disease. In future this can be developed to cover every single requirement of the medical sector and become world's first virtual hospital. Virtual assistance by doctors, online test booking and online ordering of medicines and other requirements. This would become the first stop for any patient looking get treatment to their ailments.

Code Implementation:

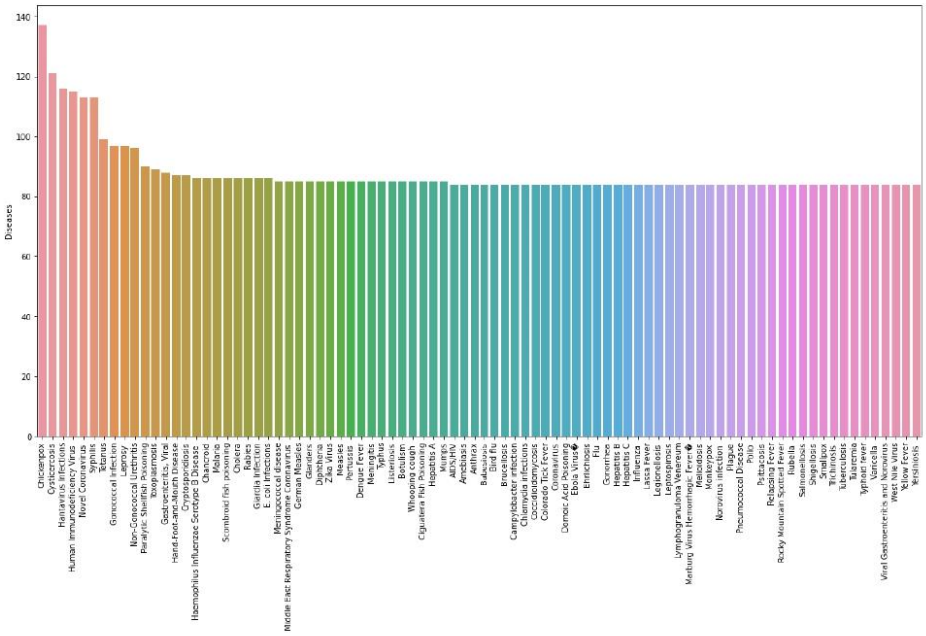
In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df = pd.read_csv("./symptoms_machine.csv")
```

In [3]:
```
figure =sns.barplot(x=df.columns[1:],y=df.sum(axis=0)[1:])
plt.xticks(range(len(df.columns[1:])), df.columns[1:], rotation='vertical')
figure.figure.set_size_inches(20,10)
```



In [4]:
```
figure =sns.barplot(x=df['Diseases'].unique(),y=df['Diseases'].value_counts())
plt.xticks(range(len(df['Diseases'].unique())), df['Diseases'].unique(), rotation='v
figure.figure.set_size_inches(20,10)
```



In [5]:
```
X = df.drop('Diseases',axis=1)
y = df['Diseases']
```

In [6]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)
```

In [7]:
```
from sklearn.ensemble import RandomForestClassifier
```

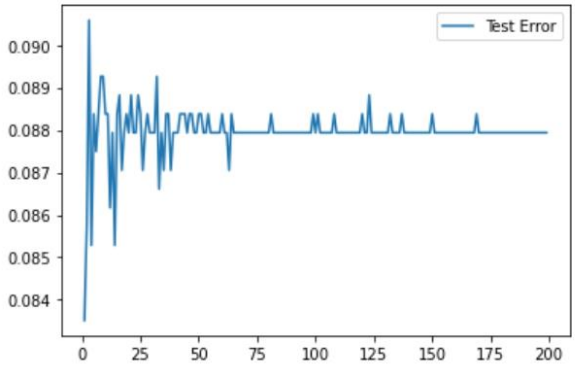
In [8]:
```
from sklearn.metrics import f1_score,classification_report,accuracy_score
```

In [9]:
```
test_error = []

for n in range(1,200):
    # Use n random trees
    model = RandomForestClassifier(n_estimators=n,max_features='auto')
    model.fit(X_train,y_train)
    test_preds = model.predict(X_test)
    test_error.append(1-accuracy_score(test_preds,y_test))
```

In [10]:
```
plt.plot(range(1,200),test_error,label='Test Error')
plt.legend()
```

Out[10]:
```
<matplotlib.legend.Legend at 0x1c35c9639a0>
```



In [11]:
```
model = RandomForestClassifier(n_estimators=64,max_features='auto')
```

In [12]:
```
model.fit(X_train,y_train)
```

Out[12]:
```
RandomForestClassifier(n_estimators=64)
```

In [13]:
```
preds = model.predict(X_test)
```

In [14]:
```
f1_score(y_test, preds, average='macro')
```

Out[14]:
```
0.9021226107683803
```

In [15]:
```
print(classification_report(y_test,preds))
```

```
                                    precision    recall  f1-score   suppo
rt

                        AIDS/HIV       1.00      1.00      1.00
28
                       Amebiasis       1.00      1.00      1.00
21
                         Anthrax       1.00      1.00      1.00
28
                       Babesiosis       1.00      1.00      1.00
21
                         Bird flu       1.00      1.00      1.00
27
                         Botulism       1.00      1.00      1.00
21
                      Brucellosis       1.00      1.00      1.00
23
          Campylobacter infection       1.00      1.00      1.00
21
                        Chancroid       1.00      0.96      0.98
28
                       Chickenpox       1.00      1.00      1.00
19
           Chlamydia infections       1.00      1.00      1.00
29
                          Cholera       1.00      0.97      0.99
36
        Ciguatera Fish Poisoning       1.00      1.00      1.00
28
```

	Coccidioidomycosis	1.00	1.00	1.00
26				
	Colorado Tick Fever	1.00	1.00	1.00
24				
	Coronavirus	1.00	1.00	1.00
25				
	Cryptosporidiosis	1.00	1.00	1.00
30				
	Cysticercosis	0.74	0.67	0.70
42				
	Dengue Fever	0.47	1.00	0.64
25				
	Diphtheria	1.00	1.00	1.00
16				
	Domoic Acid Poisoning	1.00	1.00	1.00
30				
	E. coli Infections	1.00	1.00	1.00
31				
	Ebola Virus|	1.00	1.00	1.00
32				
	Ehrlichiosis	1.00	1.00	1.00
18				
	Flu	1.00	1.00	1.00
27				
	Gastroenteritis, Viral	1.00	1.00	1.00
27				
	German Measles	1.00	0.95	0.97
20				
	Giardia Infection	1.00	1.00	1.00
19				
	Glanders	0.93	1.00	0.96
27				
	Gonococcal Infection	0.46	0.72	0.56
29				
	Gonorrhea	1.00	1.00	1.00
20				
	Haemophilus Influenzae Serotype B Disease	1.00	1.00	1.00
18				
	Hand-Foot-and-Mouth Disease	1.00	1.00	1.00
30				
	Hantavirus Infections	0.84	0.90	0.87
30				
	Hepatitis A	1.00	1.00	1.00
25				
	Hepatitis B	0.45	1.00	0.62
19				
	Hepatitis C	0.00	0.00	0.00
23				
	Human Immunodeficiency Virus	1.00	0.81	0.90
32				
	Influenza	1.00	1.00	1.00
23				
	Lassa Fever	1.00	1.00	1.00
26				
	Legionellosis	1.00	1.00	1.00
24				
	Leprosy	0.96	0.68	0.80
38				
	Leptospirosis	1.00	1.00	1.00
33				
	Listeriosis	1.00	1.00	1.00
21				
	Lymphogranuloma Venereum	1.00	1.00	1.00
23				
	Malaria	1.00	0.96	0.98
28				
	Marburg Virus Hemorrhagic Fever|	1.00	1.00	1.00
32				
	Measles	1.00	1.00	1.00

34	Melioidosis	1.00	1.00	1.00
27	Meningitis	0.00	0.00	0.00
28	Meningococcal disease	0.96	1.00	0.98
22	Middle East Respiratory Syndrome Coronavirus	1.00	1.00	1.00
19	Monkeypox	1.00	1.00	1.00
26	Mumps	1.00	1.00	1.00
24	Non-Gonococcal Urethritis	0.00	0.00	0.00
41	Norovirus infection	1.00	1.00	1.00
27	Novel Coronavirus	0.84	0.94	0.89
34	Paralytic Shellfish Poisoning	0.97	0.79	0.87
38	Pertussis	1.00	1.00	1.00
24	Plague	1.00	1.00	1.00
28	Pneumococcal Disease	1.00	1.00	1.00
25	Polio	1.00	1.00	1.00
28	Psittacosis	1.00	1.00	1.00
26	Rabies	0.96	1.00	0.98
26	Relapsing Fever	1.00	1.00	1.00
23	Rocky Mountain Spotted Fever	1.00	1.00	1.00
26	Rubella	1.00	1.00	1.00
27	Salmonellosis	0.00	0.00	0.00
23	Scombroid fish poisoning	0.96	1.00	0.98
22	Shigellosis	0.00	0.00	0.00
21	Smallpox	1.00	1.00	1.00
25	Syphilis	0.97	0.89	0.93
36	Tetanus	1.00	1.00	1.00
28	Toxoplasmosis	0.40	0.97	0.57
32	Trichinosis	1.00	1.00	1.00
36	Tuberculosis	1.00	1.00	1.00
23	Tularemia	1.00	1.00	1.00
26	Typhoid fever	1.00	1.00	1.00
24	Typhus	1.00	0.96	0.98
28	Varicella	1.00	1.00	1.00
26	Viral Gastroenteritis and Norovirus	1.00	1.00	1.00
24	West Nile Virus	1.00	1.00	1.00
19				

```
                    Whooping cough       1.00      1.00      1.00
25
                      Yellow Fever       1.00      1.00      1.00
23
                       Yersiniosis       0.32      1.00      0.49
21
                        Zika Virus       0.92      0.96      0.94
25

                          accuracy                           0.91      22
63
                         macro avg       0.90      0.92      0.90      22
63
                      weighted avg       0.89      0.91      0.90      22
63

C:\Users\Lenovo\anaconda3\lib\site-packages\sklearn\metrics\_classification.py:1245:
UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 i
n labels with no predicted samples. Use `zero_division` parameter to control this be
havior.
  _warn_prf(average, modifier, msg_start, len(result))
C:\Users\Lenovo\anaconda3\lib\site-packages\sklearn\metrics\_classification.py:1245:
UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 i
n labels with no predicted samples. Use `zero_division` parameter to control this be
havior.
  _warn_prf(average, modifier, msg_start, len(result))
C:\Users\Lenovo\anaconda3\lib\site-packages\sklearn\metrics\_classification.py:1245:
UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 i
n labels with no predicted samples. Use `zero_division` parameter to control this be
havior.
  _warn_prf(average, modifier, msg_start, len(result))
```

In [35]:

```
symptoms = {}
```

In [ ]:

```
for i in df.columns[1:]:
    temp = int(input("Do u have "+i+" (Enter Boolean value 0 or 1):"))
    symptoms[i]=[temp]

print((pd.DataFrame(symptoms)))
```

In [52]:

```
pred = model.predict(pd.DataFrame(symptoms))
```

In [ ]:

```
print(pred)
```

In [ ]:

```
pip install nbconvert
```

In [ ]:

Github link: https://github.com/IAmTheBlueDragon/Symptom_Matcher