SYMPTECH

ETHICAL AI

Joseph Nguyen

November 5, 2024

TABLE OF CONTENTS

INTRODUCTION

OVERVIEW

SympTech is an innovative AI-driven project focused on transforming healthcare by developing an artificial intelligence system designed to diagnose a particular disease. The core objective of the project is to leverage AI's advanced capabilities to enhance the accuracy and speed of medical diagnoses. By doing so, SympTech aims to support healthcare professionals in making more precise and informed decisions, which can lead to improved patient outcomes.

SympTech underscores the responsible use of AI technology to advance healthcare systems, benefiting both patients and healthcare providers while ensuring that ethical considerations remain a priority.

The project highlights several key benefits:

***Reduction of Human Error**: By using AI, SympTech reduces the potential for diagnostic mistakes that may arise from human limitations, ensuring more consistent and reliable results.*

***Ethical Use of AI in Healthcare**: A significant focus of the project is demonstrating how AI can be ethically integrated into medical practice. SympTech aims to show that AI, when used responsibly, can benefit patients without compromising the core values of equity and quality care.*

***Improved Patient Outcomes**: With faster and more accurate diagnoses, patients can receive timely treatment, which can lead to better health outcomes and quicker recovery.*

***Equitable Access to Healthcare**: SympTech also emphasizes the role of AI in supporting broader access to high-quality medical care, particularly for underserved populations. By providing scalable diagnostic tools, the project can help bridge gaps in healthcare access and ensure more equitable treatment for all patients.*

PROPOSED SOLUTION

INITIAL THOUGHTS & PLANNING

Describe the design or concept I will test. Briefly summarize the research and brainstorming that led me to this solution.

Note: Think about the community involved in the solution. Ask yourself.

Whose voice is at the table?

Who are the key stakeholders?

Who is missing? Is the solution specific?

who are the people in power/influence?

Does the solution address a specific need?

How will you educate others or spread the word about your solution?

Who would need to know about this solution?

PROPOSAL

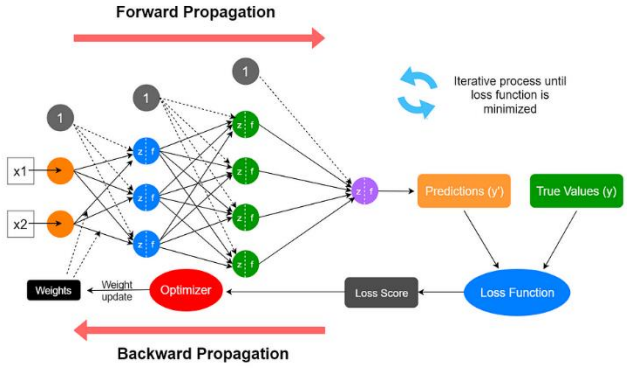
- *Inefficient scheduling in outpatient clinics leads to significant waste in the healthcare system, with billions of dollars lost each year due to inaccurate allocation of service time. This problem frustrates patients with long wait times and creates inefficiencies for clinicians, reducing overall productivity.*
- *To address this issue, I propose an intelligent appointment scheduling system powered by a machine learning model. By utilizing neural networks trained on historical data, this system aims to optimize appointment times, minimize wait times for patients and clinicians, and ultimately improve the efficiency and satisfaction of healthcare services.*
- *Through collaboration with healthcare providers, data scientists, and policymakers, this solution could bring meaningful change to the healthcare scheduling process.*

HYPOTHESIS

STATE THE HYPOTHESIS I PLAN TO TEST.

WHAT PART OF THE CHALLENGE WILL MY DESIGN OR CONCEPT SPECIFICALLY WORK TO SOLVE AND HOW?

- *The hypothesis we plan to test is that implementing a machine learning-based appointment scheduling system will significantly reduce patient wait times and clinician idle time in outpatient clinics, thereby increasing overall efficiency and satisfaction for both patients and healthcare providers.*
- *Our design specifically addresses the challenge of inefficient time allocation during patient appointments, which leads to wasted resources and dissatisfaction.*
- *By using a neural network created from a specified dataset, the model will predict appointment durations more accurately, the system will dynamically allocate service times based on individual patient needs, ensuring that clinicians are neither overbooked nor underutilized.*
- *This optimized scheduling will directly solve the problem of mismatched time allocation, reducing unnecessary delays and improving the flow of clinical operations.*



RESEARCH PLAN

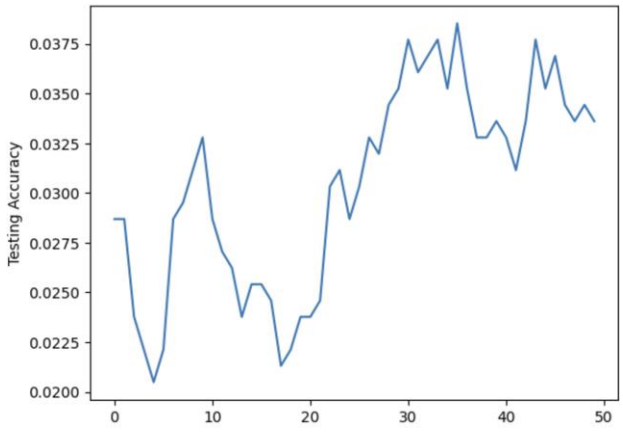
OUTLINE THE METHODS THAT WILL BE USED TO TEST MY HYPOTHESIS OR INVESTIGATE THE VIABILITY OF MY SOLUTION.

- *For our research, we will focus on quantitative methods such as surveys and statistical analysis to gather data on consumer times needed for their hospital visits and as many characteristics and different cases of why they needed to come to the hospital.*
- *We are looking into existing data depending on online libraries and datasets with specified information to complete this task, and also through documentation of other attempts at this problem to see if there are other ways to go about a solution.*

DATA / FEEDBACK

SUMMARIZE THE DATA YOU OBTAINED FROM EXPERIMENTS AND/OR FEEDBACK I HAVE GATHERED.

- *To begin, a preliminary round of research was conducted to gain a comprehensive understanding of the issue, elongated patient wait time, its causes, and current solutions.*
- *The research concentrated on the level of impacts certain factors, such as the presence of chronic illness and type, age, and impairments have on the time required by the individual to maximize a satisfactory visit.*
- *Thus, each category was first researched briefly to see its relative impact on patient time and to consider if it is viable to be included in the data.*
- *Decreasing the number of unnecessary variables further decreases the model's prediction of false negatives or positives.*
- *The data once collected was cleaned to remove outliers, which can shift the average based on a random chance.*
- *Incomplete data or empty cells were also removed to reduce gaps and increase efficiency.*



LESSONS LEARNED

SUMMARIZE THE MOST IMPORTANT FINDINGS FROM YOUR TESTING.

WHERE COULD YOU HAVE TESTED MORE? WHAT WERE SOME OF YOUR LIMITATIONS?

- *The first important lesson learned working on SympTech was rigorous data cleaning. I had a lot of inconsistencies in records and missing values for various patient demographics, symptoms, and details about previous visits in the beginning.*
- *Noisy or incomplete data would lead to inaccuracies in our model; thus, cleaning the data was very important. This was followed by the handling of missing values, standardizing categorical variables, and scaling numeric features into consistent data.*
- *Also, I first started working with the "sklearn" library in Python for preprocessing and machine learning modeling. However, there were significant limitations to this.*
- *While "sklearn" works extremely effectively on smaller, less complex datasets, it could not support the scale and complexity of my data, mainly considering feature representation and handling multi-layer dependencies among patient attributes.*
- *Likewise, these models from "sklearn" showed insufficient flexibility to handle certain neural network architectures I wanted to implement; this limited me from enhancing model performance in the manner intended.*
- *I started working on my introduction and research and addressed the issues regarding the inefficient schedules in outpatient clinics. I learned to elaborate on my data and knowledge by using my sources to support my findings. By using additional sources, I have learned to create a reference page with APA citations at the end.*

NEXT STEPS

DESCRIBE HOW YOU PLAN TO REFINE YOUR SOLUTION BASED ON WHAT YOU LEARNED AND CONDUCT FURTHER TESTING.

WILL YOU MODIFY YOUR TESTING METHODS?

- *Seeing these limitations, we will move to PyTorch. The flexibility of PyTorch will give us, will construct and tailor neural networks, that would fit the complexity of our problem.*
- *Unlike sklearn, PyTorch can support deeper network configurations intended for feature learning from data using more sophisticated patterns.*
- *Therefore, instead of using only the sklearn library in Python, we will use Pytorch for the neural networks.*

***TIPS**: This is an important time to think about my results and how they will impact communities, people, and the environment. Think through how those results could positively and negatively affect those areas and include them in my findings. Remember, telling the whole story is important and a well-rounded solution includes both positive and negative aspects.*

CODE IN PYTHON

In [50]:
```
import numpy as np
import pandas as pd
```

In [51]:
```
path = "/kaggle/input/symptech-dataset/Appointment_TimeData.xlsx"
```

In [52]:
```
df = pd.read_excel(path, engine="openpyxl")

df.head()
```

Out[52]:

	Provider ID	Patient ID	service_time	Visit Type_n	Patient Type	Gender	Age	Refill/Referral	Vision	Annual Exam	...	Chronic_Count	C_Asthma	C_Al
0	7018	118782558	27	PC VISIT	0	F	28	0	0	0	...	0	0	
1	7018	163281546	24	PC SAME DAY	0	M	20	0	0	0	...	0	0	
2	6945	81916344	23	PC VISIT	0	F	44	0	0	0	...	0	0	
3	6945	79547760	36	PC VISIT	0	M	30	1	0	0	...	1	0	
4	6945	81286812	40	PC VISIT	0	M	27	0	0	0	...	0	0	

5 rows × 31 columns

In [53]:
```
y = df["service_time"]
x1 = df.drop("service_time", axis = "columns")
x2 = x1.drop("Gender", axis = "columns")
x3 = x2.drop("#ofvaccsin_n", axis = "columns")
x = x3.drop("Visit Type_n", axis = "columns")
```

In [54]:
```
from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, shuffle = True)
```

In [55]:
```
print(len(x_train))
print(len(x_test))
```

```
4876
1220
```

In [56]:
```
neighbors = np.arange(1,51)
print(neighbors)
```

```
[ 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
 49 50]
```

In [57]:
```
testing_accuracies = np.empty(len(neighbors))
```

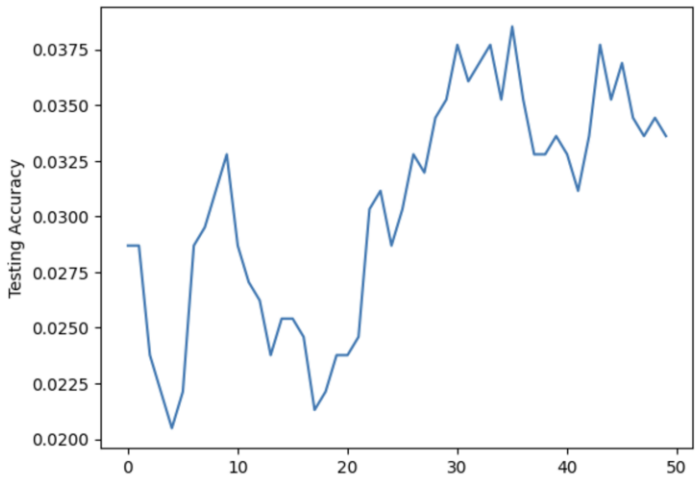
In [58]:
```
from sklearn.neighbors import KNeighborsClassifier

for index, neighbor_count in enumerate(neighbors):
    model = KNeighborsClassifier(n_neighbors = neighbor_count)
    model.fit(x_train, y_train)
    testing_accuracies[index] = model.score(x_test, y_test)
```

In [59]:
```
import matplotlib.pyplot as plt

plt.plot(testing_accuracies)
plt.ylabel("Number of Neighbors")
plt.ylabel("Testing Accuracy")
```

Out[59]: Text(0, 0.5, 'Testing Accuracy')



In [60]:
```
final_model = KNeighborsClassifier(n_neighbors = 28)
final_model.fit(x_train, y_train)
print(final_model.score(x_test, y_test) * 100)
```

3.1967213114754096