# A MODEL FOR LIFESTYLE DISEASE ALZHEIMER'S PREDICTION USING K-NEAREST NEIGHBORS

***T.RADHIKA, P.SHREYA, G. AKHILSAI***

***Dept. of Information Technology, TKR College of Engineering and Technology, Hyderabad, Telangana***

***radhikathatikonda08@gmail.com, shreyareddy2710@gmail.com, gopathiakhilsaimonty@gmail.com.***

## ABSTRACT

*If you decide to check out that offer for a vacation. You browse through the travel agency website and search for a hotel. When you look at a specific hotel, just below the hotel description there is a section titled "You might also like thesehotels". This is a common use case of Machine Learning called "Recommendation Engine". Again, many data points were used to train a model in order to predict what will be the best hotels to show you under that section, based on a lot of information theyalready know about you.So if you want your program to predict, for example, traffic patterns at a busy intersection (task T), you can run it through a machine learning algorithm with data about past traffic patterns (experience E) and, if it has successfully "learned", it will then do better at predicting future traffic patterns (performance measure P). The highlycomplex nature of many real-world problems, though, often means that inventing specialized algorithms that will solve them perfectly every time is impractical, if not impossible. Examples of machine learning problems include, "Is this cancer?", "Whichof these people are good friends with each other?", "Will this person like this movie?" such problems are excellent targets for Machine Learning, and in fact machine learninghas been applied such problems with great success.*

## 1. INTRODUCTION

The term Machine Learning was coined by Arthur Samuel in 1959, an Americanpioneer in the field of computer gaming and artificial intelligence and stated that "it gives computers the ability to learn without being explicitly programmed". And in 1997, Tom Mitchell gave a "well-posed" mathematical and relational definition that "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves withexperience E.

Machine Learning is a latest buzzword floating around. It deserves to, as it is one of the most interesting subfields of Computer Science. So, what does Machine Learning really mean?

Let's try to understand Machine Learning in layman terms. Consider you are trying to toss a paper to a dustbin. After first attempt, you realize that you have put toomuch force in it. After second attempt, you realize you are closer to target but you needto increase your throw angle. What is happening here is basically after every throw weare learning something and improving the end result. We are programmed to learn from our experience.

This implies that the tasks in which machine learning is concerned offers a fundamentally operational definition rather than defining the field in cognitive terms. This follows Alan Turing's proposal in his paper "Computing Machinery and Intelligence", in which the question "Can machines think?" is replaced with the question"Can machines do what we (as thinking entities) can do?" Within the field of data analytics, machine learning is used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "producereliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data set(input).

## 2. LITERATURE SURVEY

In July 2018, Mrunmayi Patil along with her colleagues described "A Proposed model for Lifestyle Disease Prediction Using Support Vector Machine". They proposed a model for predicting whether a person is having a lifestyle disease or not. So the main idea behind this is traditional method is traditional method is costly. They said that diseases are mainly caused by a combination of transformation, lifestyle selections, and surroundings. In addition, health risks in an individual's family is one of the most crucial things an individual can do to help his/her practitioner understand and diagnose hereditarily linked syndromes like cancer,

diabetes, and mental illness. Lifestyle diseases include atherosclerosis; heart disease and stroke, obesity and type II diabetes, and smoking and alcohol totalled diseases. They proposed a model using Supprt VectorMachine and used to predict lifestyle diseases that an individual might be susceptible to. The need for public awareness is not stressed enough, but lifestyle diseases are easyto prevent. Simply modifying an individual's lifestyle to reduce and eliminate risks can be interesting. Deoxyribonucleic acid (DNA) and genetic testing are creating a new expanse of personalized medicine. However, on an average, DNA testing may incur ₹ 10,000 to 20,000 [2], which is expensive. Though there are many receding diseases andtests, they are erratically tested because they are costly, and factual tests have not been developed yet. Our lifestyles are imperative in increasing or decreasing risks of various diseases. According to some research conducted in the discipline of epigenetics determines that an individual's lifestyle selections can modify his/her well-being at genetic level. This study discusses about a model that can predict the probabilities of an individual obtaining a lifestyle disease. Lifestyle diseases depend on factors like heaviness, workout, and food likings and thus have a strong association with the above-mentioned factors. In our simulated model, an actor will input his/her details like fatness, sleeping habits and will discover the likelihood of suffering from lifestyle diseases.

Countless scholars have used ML- and data mining (DM)- based algorithms to predict diseases in health sciences. A few of them are explained below. Suzuki et al. [3] analyzed annual health checkup data from 1,546 employees. They concluded that 5% weight reduction with succeeding weight control and daily workouts would be helpful in treating nonalcoholic fatty liver disease after a systematic health check for 12 months to evaluate changes in lifestyle with a shift in serum alanine aminotransferase (ALT).

S. A. Pattekari and A. Parveen [4] recommended an intelligent system that uses a DM technique that retrieves unseen data from stockpiled database and acquires user answersfor predefined questions related to blood sugar, sex, age, height, etc. and compares them to stored database values, i.e., trained dataset. A. Anand and D. Shakti [5] conversed about relationship between diabetes risk probably to be developed from an individual'sdaily lifestyle activities such as his/her eating and sleeping routines, physical movementin addition to body mass index by acquiring data from questionnaires averring 75% accuracy.

There are a myriad number of lifestyle and genetic factors that contribute toward the

participation of lifestyle diseases. Numerous exorbitant tests are performed and their results are analyzed to recognize lifestyle diseases. Doctors scrutinize these results based on patients' previous results with similar conditions. Analyzing them is difficult and requires specialized doctors. The burgeoning field of artificial intelligence, especially ML, has made it easier to adjudicate a patient's lifestyle disease. Based on former patients' lifestyle, we can predict if a person is suffering from a lifestyle disease or not. Computational techniques are accurate and fast, and since there is a need for spreading lifestyle disease awareness, we were motivated to design a lifestyle disease prediction model. The literature survey carried out advocates that numerous research papers each directing a specific disease have been presented. Lifestyle and inheritances go concurrently and are interrelated.

Nevertheless, a healthy lifestyle surpasses congenital risk. Diseases are mainly caused by a combination of transformation, lifestyle selections, and surroundings; hence, we propose a lifestyle disease prediction model that mutually forecasts lifestyledisorders that an individual might be susceptible to.

So, here we are trying to predict for Alzheimer's disease- it is the most common cause of dementia — a continuous decline in thinking, behavioural and social skills that disrupts a person's ability to function independently. using KNN algorithm where are taking a dataset from Kaggle repository. We are dividing the dataset into two parts- 80% to training data and 20% to testing data. And in KNN algorithm the accuracy of the algorithm depends on the k value we choose. For this, we tried the different values for k ranging from 1 to 25. After checking for all the values of k we chose k as 6 as it is giving better accuracy than other values. Then, we are creating a web page from whichwe will we are taking input from the user. Using flask framework, we are writing backend code where we passing the values of the web page to the dataset, and are tryingto predict whether a person is suffering from Alzheimer's or not.

## 3. PROPOSED SYSTEM

The aim of our study was to develop a novel classification system based on Machine learning which provides easy and efficient methods to simplify complex problems by reducing the measure of complexity and also very determined in providing easier and simpler ways which are adaptable in our present modern world where everything is bigon technology. We will develop a system for detecting the Lifestyle disease Alzheimer's disease which is the most common cause of dementia — a continuous

decline in thinking, behavioural and social skills that disrupts a person's ability to function independently. This can be developed by a simple algorithm called the k-nearest neighbours (KNN) algorithm which is a simple, easy-to-implement supervised machinelearning algorithm that can be used to solve both classification and regression problems.In this we take a dataset related to Alzheimer's disease and divide it into two sets. One isused as training data and the other as testing data .The training data is used tomake sure the machine recognizes patterns in the data, the cross-validation data is used to ensure better accuracy and efficiency of the algorithm used to train the machine, andthe test data is used to see how well the machine can predict new answers based on its training. In this dataset we use various terms in order to acquire information regarding the disease. We use certain attributes like Socioeconomic status(SES),The Mini-Mental State Examination (MMSE),Clinical Dementia Rating (CDR),Estimated Total Intracranial Volume(eTIV), etc, which determine the severity of the condition. We willdesign a user interface where we acquire the data from patient and determine whether he is suffering from dementia or not by using KNN algorithm. The result we obtain willbe accurate and basing on the result the patient can go the appropriate treatment suggested by their respective doctor.

Flask is a web framework. This means flask provides you with tools, libraries and technologies that allow you to build a web application. This web application can be some web pages, a blog, a wiki or go as big as a web-based calendar application or a commercial website. Flask is part of the categories of the micro-framework. Micro-framework is normally framework with little to no dependencies to external libraries. This has pros and cons. Pros would be that the framework is light, there are little dependency to update and watch for security bugs, cons is that some time you will haveto do more work by yourself or increase yourself the list of dependencies by adding plugins. In the case of Flask, its dependencies are: Werkzeug a WSGI utility library and jinja2 which is its template engine.

WSGI is basically a protocol defined so that Python application can communicatewith a web-server and thus be used as web-application outside of CGI. Have you ever built a website? Did you face the problem that to keep the style of the website consistent, you have had to write multiple times the same text? Did you ever tried to change the style of such website? If your website contains only few pages,changing its style will take you some time but is doable. However, if you have a lot of pages

(for example the list of items you sell in your store), this task become overwhelming. Using templates you are able to set a basic layout for your pages and mention which element will change. This way you can define your header once and keep it consistent over all the pages of your website, and if you need to change your header, you will only have to update it in one place. Using a template engine will save you a lot of time when creating your application but also when updating and maintainingit.

***A Hello Flask Application***

```
import flask app=flask.Flask(_name_)
@app.route('/')
def index():
#Displays the index page accessible at'/'
return flask.render_template('index.html')
if name == 'main':
app.debug=trueapp.run
```

***Create the template index.html***

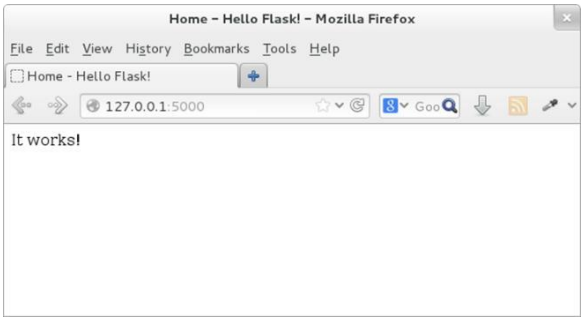
Put the following code in this file

```
<!DOCTYPE html>
<html lang="en">
<head>
<title>Hello World!</title>
<link rel="stylesheet" type="text/css"
href="{{url_for('static',filename=hello.css')
}}">
</head>
<body>
It works!
</body>
</html>
```

***Run the flask application***

python hello_flask.py

Access http://127.0.0.1:5000/ this should simply show you in black on white the text "Itworks!" (see Figure below).




***TEST CASE 1:***

The following test case test the component "Demented". This component classifies a particular row into "Demented" group. If we give the values that results into a demented group which is in the dataset, then we will get result as "You are suffering from dementia".

**Table 6.2.1.** Description of "demented" group

| Step | Action | Expected System Response | Pass/Fail |
|---|---|---|---|
| 1 | Fill the first form | Should open second web page | Pass |
| 2 | Fill the second form | Should open third web page | Pass |
| 3 | Fill the third form | - | Pass |
| 4 | Click submitbutton | Should display- "You are suffering from dementia" | Pass |

***TEST CASE 2:***

The following test case test the component "Non-Demented". This component classifies a particular row into "Non-Demented" group. If we give the values that results into a demented group which is in the dataset, then we will get result as "You are suffering from dementia".

**Table 6.2.2.** Description of "non-demented" group

| Step | Action | Expected System Response | Pass/Fail |
|---|---|---|---|
| 1 | Fill the first form | Should open second web page | Pass |
| 2 | Fill the second form | Should open third web page | Pass |
| 3 | Fill the third form | - | Pass |
| 4 | Click submitbutton | Should display-"You are not suffering from dementia" | Pass |

***TEST CASE 3:***

The following test case test the component "Converted". This component classifies a particular row into "Converted" group. If we give the values that results into a demented group which is in the dataset, then we will get result as "You are suffering from dementia".

**Table 6.2.3.** Description of "converted" group

| Step | Action | Expected System Response | Pass/Fail |
|---|---|---|---|
| 1 | Fill the first form | Should open second web page | Pass |
| 2 | Fill the second form | Should open third web page | Pass |
| 3 | Fill the third form | - | Pass |
| 4 | Click submitbutton | Will display either you are suffering from dementia or not suffering from dementia | Fail |

**KNN.ipynb:Out[2]:**

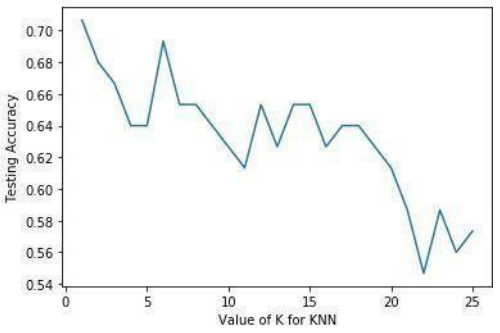
| | Subject ID | MRI ID | Group | Visit | MR Delay | M/F | Hand | Age | EDUC | SES | MMSE | CDR | eTIV | nWBV | ASF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | OAS2_0001 | OAS2_0001_MR1 | Nondemented | 1 | 0 | M | R | 87 | 14 | 2.0 | 27.0 | 0.0 | 1987 | 0.696 | 0.883 |
| 1 | OAS2_0001 | OAS2_0001_MR2 | Nondemented | 2 | 457 | M | R | 88 | 14 | 2.0 | 30.0 | 0.0 | 2004 | 0.681 | 0.876 |
| 2 | OAS2_0002 | OAS2_0002_MR1 | Demented | 1 | 0 | M | R | 75 | 12 | NaN | 23.0 | 0.5 | 1678 | 0.736 | 1.046 |
| 3 | OAS2_0002 | OAS2_0002_MR2 | Demented | 2 | 560 | M | R | 76 | 12 | NaN | 28.0 | 0.5 | 1738 | 0.713 | 1.010 |
| 4 | OAS2_0002 | OAS2_0002_MR3 | Demented | 3 | 1895 | M | R | 80 | 12 | NaN | 22.0 | 0.5 | 1698 | 0.701 | 1.034 |

**Out[4]:**

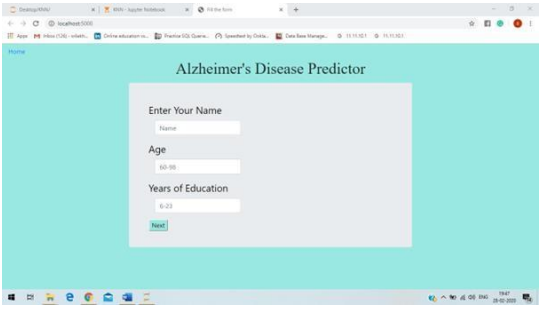
| | Group | M/F | Age | EDUC | SES | MMSE | CDR | eTIV | nWBV | ASF |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 87 | 14 | 2.0 | 27.0 | 0.0 | 1987 | 0.696 | 0.883 |
| 1 | 0 | 0 | 88 | 14 | 2.0 | 30.0 | 0.0 | 2004 | 0.681 | 0.876 |
| 2 | 1 | 0 | 75 | 12 | 2.0 | 23.0 | 0.5 | 1678 | 0.736 | 1.046 |
| 3 | 1 | 0 | 76 | 12 | 2.0 | 28.0 | 0.5 | 1738 | 0.713 | 1.010 |
| 4 | 1 | 0 | 80 | 12 | 2.0 | 22.0 | 0.5 | 1698 | 0.701 | 1.034 |

**Out[7]:**

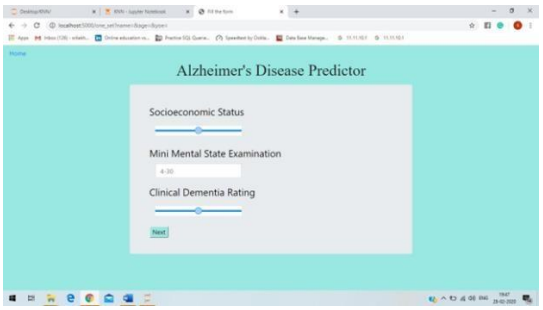
```
Text(0, 0.5, 'Testing Accuracy')
```



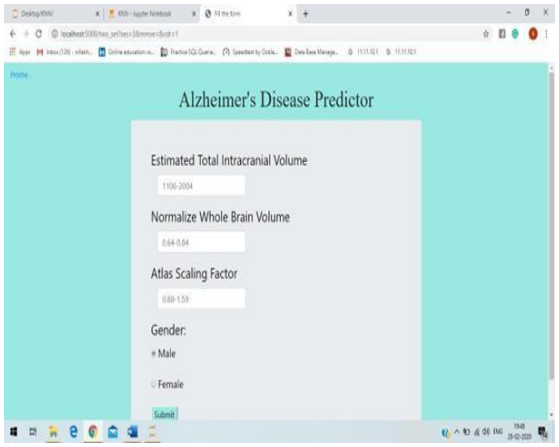

**One.html:**



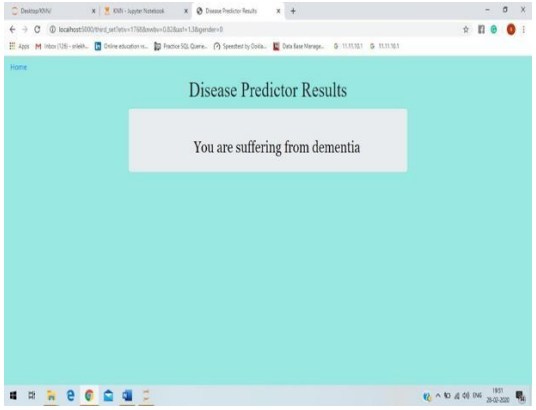
**Two.html:**




**Three.html:**




**Result1.html:**




## 4. CONCLUSION

In this project, we build a system for determining a Lifestyle disease known as Alzheimer's which is a type of dementia that causes problems with memory, thinking and behavior. Symptoms usually develop slowly and get worse over time, becoming severe enough to interfere with daily tasks. We use a machine learning algorithm known as k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. We build a system using various modules of python and an user interface, which intakes data from the respective patient. The data consists of all attribute which are required to determine whether the patient is suffering from dementia or not. This system helps the doctors to easily determine the disease and therefore, they can suggest the treatment the patient is supposed to go through. This system helps in decreasing the level of complexity by giving accurate and faster result.

## REFERENCES

1. https://www.ritchieng.com/machine-learning-k-nearest-neighbors-knn/
2. https://www.youtube.com/watch?v=6kZ-OPLNcgE
3. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447
4. http://profsite.um.ac.ir/~monsefi/machine-learning/pdf/Machine-Learning-Tom-Mitchell.pdf
5. https://www.kaggle.com/jboysen/mri-and-alzheimers/metadata