HEALTH MONITORING USING MACHINE LEARNING

Developed by:kanukuntla Pravalika,Chintakindi Harika,Hafsa Siddiqa,Chetla Akhil,Manchala Keerthana

Smart Bridge-Remote Summer Internship Program

1.INTRODUCTION

Health monitoring systems continue to evolve in this digital age. AI-powered health-tracking wearable devices today continuously monitor vital signs of a patient’s health and basis that generate a large number of observations per second. However, healthcare providers find it challenging to manage multiple data sets across

a large volume of patient data. To address this, machine learning applications in healthcare provide real-time activity recognition within clinical systems and further help them with betterdiagnosis and treatment.

In this Health monitoring system we will be detecting the level of ill health of the person and would recommend few medications that can be taken by him by which he can recover a bit than before and here we are seeing some parameters of the person which are blood pressure,pulse and temperature by these parameters we can analyze his health status from 0-2 where 0 is that he is

perfectly okay, 2 is that the person is very much ill that he need to consult doctor.machine learning lends itself to some processes better than others. Algorithms can provide immediate benefit to disciplines with processes that are reproducible or standardized. Also, those with large image datasets, such as radiology, cardiology, and pathology, are strong candidates. Machine learning can be trained to look at images, identify abnormalities, and point to areas that need attention, thus improving the accuracy of all these processes. Long term, machine learning will benefit the family practitioner or internist at the bedside. Machine learning can offer an objective opinion to improve efficiency,

reliability, and accuracy.

1.1.Overview

Our customer was seeking a reliable health tracking system for critical care patient diagnosis. As part of this, they wanted us to develop an ML-based solution to collect and monitor the

sequences of patient data recorded by connected devices. Further, the developed system would offer real-time insights from historical and current health data sets as and when needed. This system which is specifically developed to predict health status of the application users. It provides digital information of the patients which is centrally stored. This data can be used for designing an improved healthcare delivery system. The main

objective of Health monitoring systems is to predict whether a person is healthy or not, using parameters such as Body Mass Index (BMI), Temperature, Pressure, alcohol consumer and smoker and create website for interaction for the user to check his health status.

1.2.purpose

Our aim from the project is to make use of pandas,matplotlib&seaborn libraries from python to extract the

libraries for machine learning for health monitoring
Secondly,to learn how to hypertune the parameters using grid across validation for the random forest machine learning algorithm.
And in the end,to predict whether the person is healthy or not using ensemble techniques of combining the predictions from multiple machine learning algorithms and withdrawing the conclusions.

2.LITERATURE SURVEY

Huang Lin with Chi Zhang proposed a new framework for the wireless mobile health monitoring. Cloud based mobile health monitoring uses the concept of providing privacy over wireless communication channel so that the cost involves during the communication can be minimize [1]. This framework is designed so that the client's and user data involve can be made secure and private. Here in this paper a new way of providing decryption using private key and the concept of decryption reduces the computational complexity of the framework. This framework strongly enhances the security issues over assisted cloud based mobile and performance.

Although the technique implemented here provides high security to the user's data but further enhancements can be done for the escrow problem and proxy re-encryption problem so that the computational overhead can be reduces and chances from various attacks can be prevented [1].
A.Koteswaramma and S. Lakshmi Soujanya also provides a framework for mobile based health care system for the data security of patient's over wireless channel [2]. A MediNet is developed for the privacy of user's data that provides connectivity between mobile devices of the patient's and between server and mobile device.

2.1 PROPOSED SOLUTION

MACHINE LEARNING(RANDOM FOREST):

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

As the name suggests, ***"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."*** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

3.THEORETICAL ANALYSIS

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results

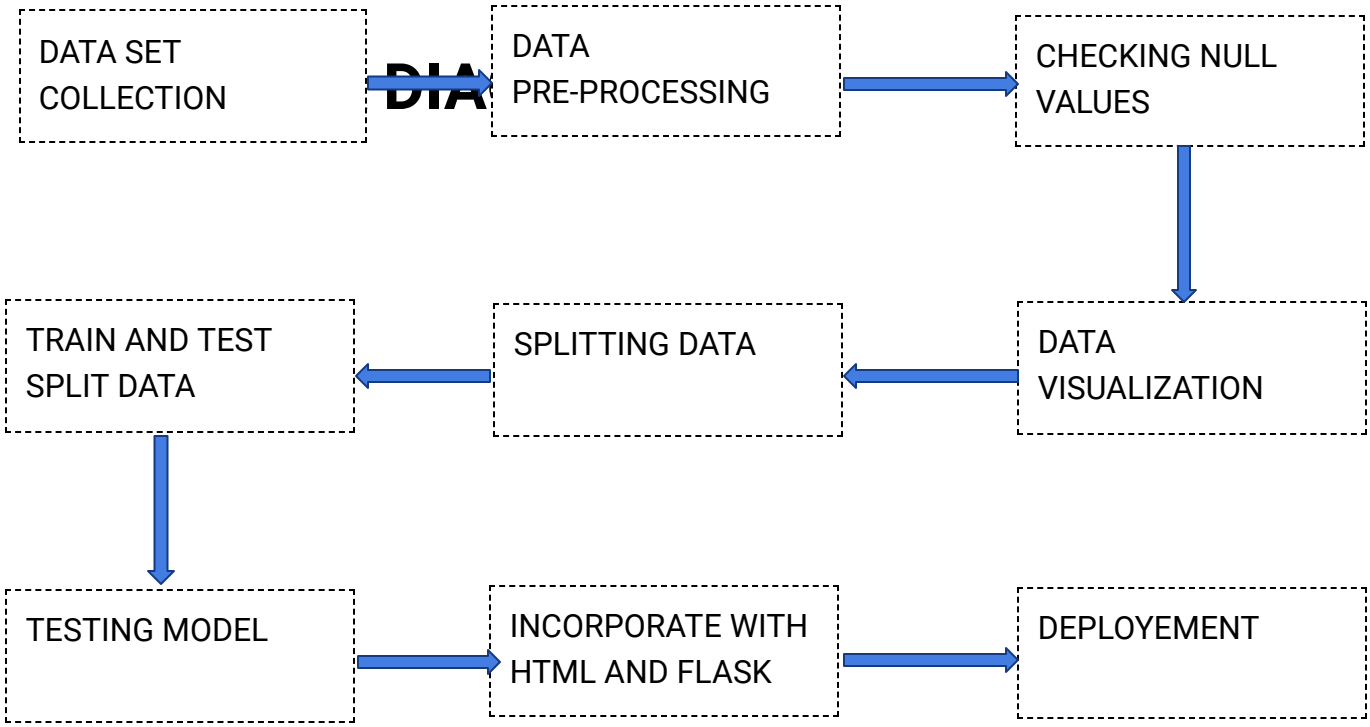
rather than a guessed result.

- The predictions from each tree must have very low correlations.

Below are some points that explain why we should use the Random Forest algorithm:

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.

- It can also maintain accuracy when a large proportion of data is missing.



3.2.SOFTWARE DESIGNING

- Jupytter Notebook Environment
- Spyder Ide
- Machine Learning Algorithms
- Python(Pandas,numpy,matplotlib,seaborn,sklearn)

- HTML
- Flask

We developed this health monitoring by using the pyton language which is a interpreted and high level programming language and using the machine learning algorithms.

For coding we used the jupyter notebook environment of the anaconda distributions and the spyder.It is an integrated scientific programming in the python language.

For creating an user interface for the prediction we used flask.It is a micro web framework written in python.It is classified as a microframework because it does not require particular tools or

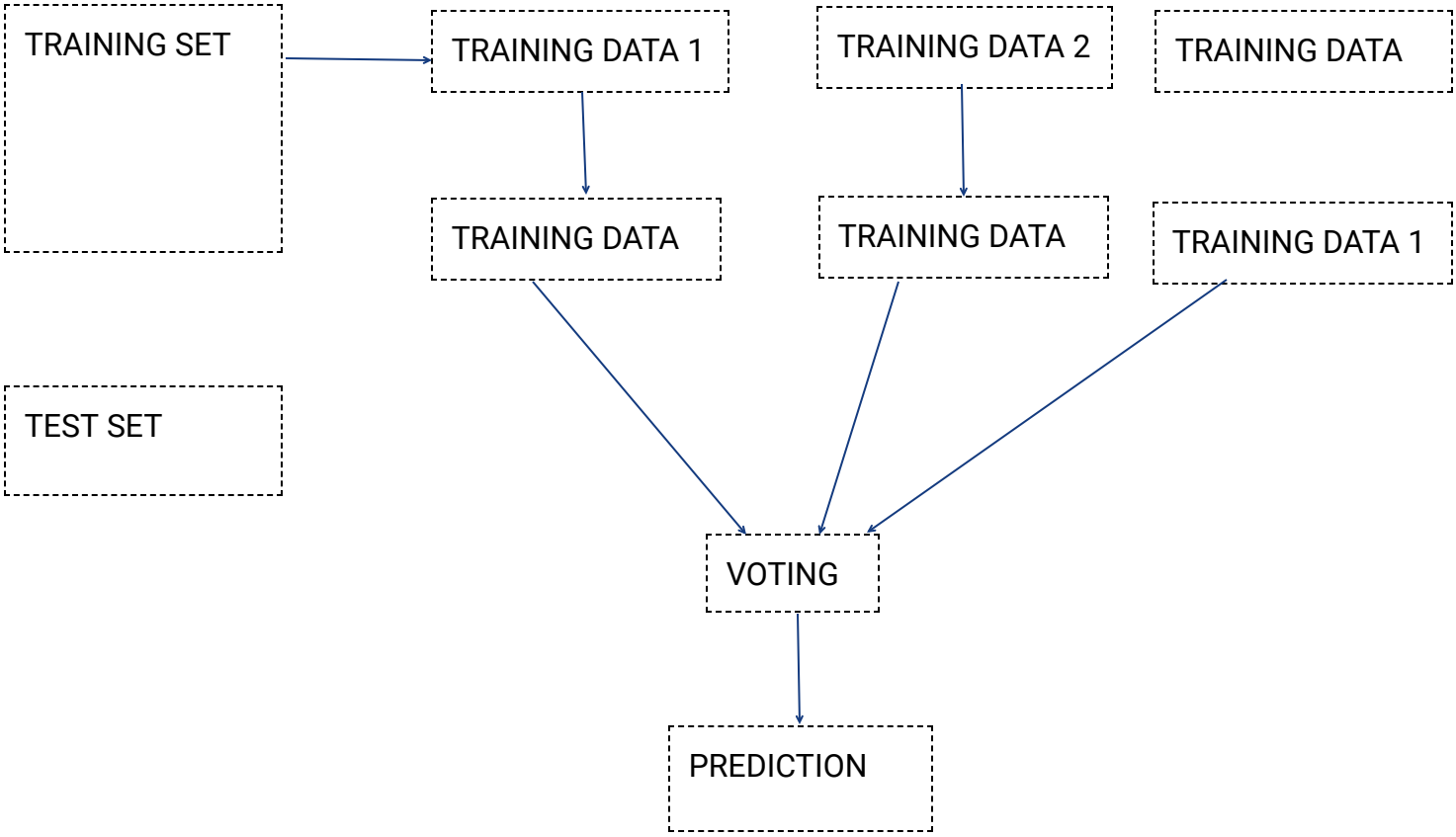
libraries.It has no database abstraction layer,form validation,or any other components where pre-existing third-party libraries provide common functions,and a scripting language to create a webpage in HTML by creating the templates to use in functions of the Flask andHTML.

4.EXPERIMENTAL INVESTIGATION

In this paper,the data set we uesd contains more than 100 people with 10 attributes,After that,the missing values are filled and the duplicate or meaning less attributes are deleted,finally we have retained to 5 attributes.Those attributes were shown below in the screenshot of the data set we used.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	SBP	DBP	Pulse	smoker	Temperatur	Level													
2	120	80	80	no	98	0													
3	140	83	75	no	100	1													
4	155	100	92	yes	104	2													
5	115	82	79	no	97	0													
6	119	79	85	no	102	1													
7	95	65	75	yes	101	2													
8	116	75	70	no	99	0													
9	100	60	79	no	98	2													
10	119	78	68	no	100	0													
11	110	69	65	yes	102	2													
12	117	73	76	no	99	0													
13	130	88	84	yes	101	1													
14	145	90	82	yes	104	2													
15	118	75	75	no	99	0													
16	114	76	69	yes	98	0													
17	105	65	60	yes	102	2													
18	132	89	76	no	101	1													
19	135	87	76	no	100	1													
20	140	91	80	yes	102	2													
21	119	78	70	yes	99	0													
22	116	77	74	no	98	0													
23	106	68	79	yes	103	2													
24	129	89	82	no	100	1													
25	126	90	80	yes	101	1													
26	147	95	84	yes	104	2													
27	116	78	70	no	98	0													
28	118	76	68	no	99	0													
29	128	87	80	no	100	1													
30	130	85	82	no	101	1													

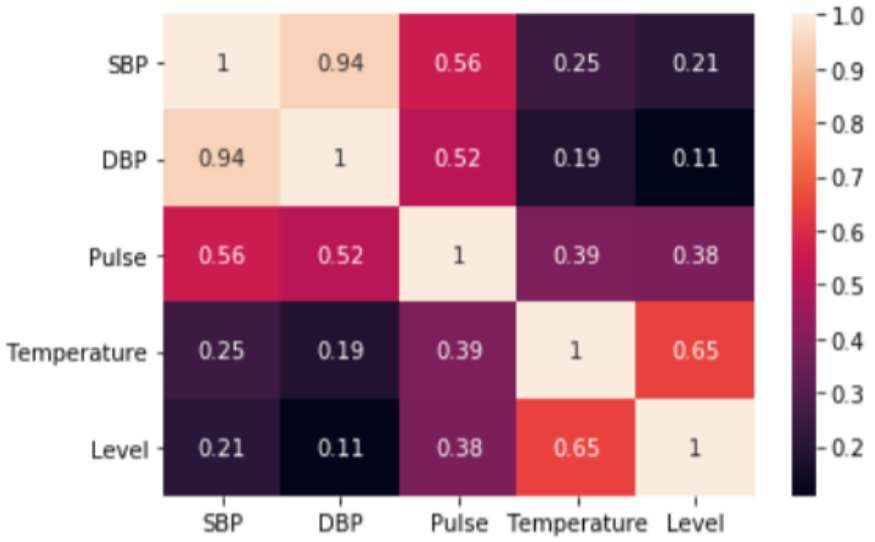
5.FLOWCHART


TRAINING SET
TRAINING DATA 1
TRAINING DATA 2
TRAINING DATA
TRAINING DATA
TRAINING DATA
TRAINING DATA 1
TEST SET
VOTING
PREDICTION

RESULT

In this paper,the random forest algorithm is used to predict the person is healthy or not.Random forest is a supervised learning algorithm.It can be used both for classification and regression.Iis also the most flexible and easy to use algorithm.It technically is an ensemble method(based on the divide-and-conquer approach). Random forests also offers a good feature selection indicator.scikit-learn provides an extra variable with the model,which shows the relative importance or contribution of each feature in the training phase.Then it scales the relevance

down so that the sum of all score is 1.
this score helps to choose the most important feature and drop the least important ones for model building.
Random forest uses gini importance or mean decrease in impurity(MDI)to calculate the importance of each feature.Gini importance is also known as the total decrease in node impurity.This is how much the model fit or accuracy decreases when you drop a variable.The gini index can describe the overall explanatory power of the variables.



7.ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

- Random forests is considered as a highly accurate and robust method because of the number of decision tress participating in the process.
- It does not suffer from the overfitting problem.The main reason is that it takes the average of all the predictions,which cancels out the biases.
- The algorithm can be used in both classification and regression problems.
- Random forests can also handle missing values.There are two ways to handle these:using median values to replace

continuous variables,and computing the proximity-weighted averagae of missing values.

- You can get the relative feature importance,which helps in selecting the most contributing features for the classifer.

DISADVANTAGES:

- Random forests is slow in generating predictions because it has multiple decision tress.Whenever it makes a predictiion,all the trees in the forest have to make a prediction for the same given input and then perform voting on it.This whole process is time-consuming.

- The mode is difficult to interpret compared to a decision tree,where you can easily make a decision by following the path in the tree.

APPLICATIONS

- Banking:Banking sector mostly uses this algorithm for the identification of loan risk.
- Medicine:With the help of this algorithm,disease trends and risks of the disease can be identified.
- Land use:We can identify the areas of similar land use by this algorithm.

- Marketing:Marketing trends can be identified using this algorithm.

9.CONCLUSION

Random forest isa great algorithm to train early in the model development process,to see how it performs.Its simplicity makes building a "bad"random forest a tough proposition.
The algorithm is also a great choice for anyone who needs to develop a model quickly.on the top of that,it provides a pretty good indicator of the importance it assigns to your features.

Random forests are also very hard yo beat performance wise.Of corse ,you can probably always find a model that can perform better,like a neural network for example,but these usually take more time to develop,though they can handle a lot of different feature types,like binary,categorical and numerical.
Overall,random forest is a (mostly)fast,simple and flexible tool,but not without some limitations.

FUTURE SCOPE

In future random forest can be applied on other data sets available for health care.A rigorous analysis of other machine

learning algorithms other than these six can also be done in future to investigate the power of machine learning algorithms for health care.In further study,we will try to conduct experiments on larger data sets or try to tune the model so as to achieve the state of art performance of the model and a great UI system making it complete web application model.

APPENDIX

HTML:

```
<html>
<body bgcolor="green">
```

```
<form action="/login"method="post">
<h1>HEALTH MONITORING</h1>
<p>SBP</p>
<p><input type="number" name="SBP"/></p>
<p>DBP</p>
<p><input type="number" name="DBP"/></p>
<p><pulse</p>
<p><input type="number" name="pulse"/></p>
<p>Temperatures</p>
<p><input type="number" name="Temperature"/></p>
<p><input type="submit" value="click"/></p>
```

{{y}}
</form>
</body>
</html>

APP.PY:

```
from flask import Flask,render_template,request
import pickle
model=pickle.load(open('level.pkl','rb'))
app=Flask(__name__)
```

```
@app.route('/')
def hello_world():
return render_template("index.html")
@app.route('/login',methods=["POST"])
def func2():
SBP=request.form['SBP']
DBP=request.form['DBP]
pulse=request.form['Temperature']

data=[[int(SBP),int(DBP),int(pulse),int(Temperature)]]
```

```
pred=model.predict(data)
print(pred[0])
return render_template("index.html".y=("The
level:"+str(pred[0])))
if __name__='__main__';
app.run(debug=True)
```