INFORMATION TECHNOLOGY | AIBUY3A | *AI SOLUTION PROJECT*

AI SMARTHEALTH CONNECTION

GROUP MEMBERS:

STUDENT NUMBER	INITIALS AND SURNAME	SIGNATURE

CONTENTS

Information Technology | AIBUY3A | *AI Solution project* 1

Group Members: 1

1. Introduction 3

2. Business Objectives 4

Automating business processes: 4

2.2 Requirements: ..... **Error! Bookmark not defined.**

2.1 Business Success Criteria: 4

3. Machine Learning approach: 5

3.1 Decision-Tree model 4

REFERENCES 7

Libraries

In [453…

```
import pandas as pd
import numpy as np
import itertools
import matplotlib.pyplot as plt
import seaborn as sns
import datetime
%matplotlib inline


from sklearn.model_selection import train_test_split
from sklearn import ensemble
from sklearn.metrics import classification_report
from sklearn.model_selection import cross_val_score
from sklearn.metrics import confusion_matrix
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import roc_curve, auc
from sklearn.metrics import confusion_matrix
```

1. INTRODUCTION

In numerous developing communities, access to quality healthcare proves to be a major obstacle due to scarce resources, deficient infrastructure, and a shortage of medical expertise. Our proposed AI-based solution, aptly named "SmartHealth Connect," seamlessly aligns with the overarching theme of "an AI Solution for communities" by harnessing cutting-edge technologies and artificial intelligence to revolutionize healthcare delivery.

The program endeavors to tackle the issue of limited access to fundamental healthcare services in various regions across the globe by digitally transforming healthcare. We aim to foster a more just and fair society by creating solutions that enhance healthcare efficiency and outcomes.

1.1 Problem Definition:

The problem we address is the inadequate healthcare access and services in underserved communities. This includes limited medical facilities, healthcare professionals, and information dissemination, leading to preventable illnesses and higher mortality rates. Our solution aims to bridge this gap by employing AI-powered telemedicine, diagnostic tools, and health monitoring devices to provide timely and accurate healthcare services to residents.

By integrating AI technologies, such as natural language processing and machine learning algorithms, we can establish a virtual healthcare platform that enables remote consultations, symptom analysis, and even initial diagnostics. This solution also includes wearable devices for continuous health monitoring, allowing early detection of potential health issues. Furthermore, our AI system would utilize predictive analytics to identify health trends within the community, assisting healthcare authorities in resource allocation and planning.

```
In [550…
df['No-show'] = df['No-show'].replace('Yes', 1)
df['No-show'] = df['No-show'].replace('No', 0)

In [551…
df.AppointmentDay = df.AppointmentDay.apply(np.datetime64)

def calculateday(datetime):
    #year, month, day = (int(x) for x in date.split('-'))
    DayOfTheWeek = datetime.date().strftime("%A")
    return DayOfTheWeek

df['DayOfTheWeek'] = df['AppointmentDay'].apply(lambda x: calc
```

***1.2* System Benefits :**

- Error Reduction
- Instant Assistance
- Fast and Accurate
- Cost Reduction

2. Business Objectives

Automating business processes:

The main objective of "SmartHealth Connect" is to improve healthcare accessibility and outcomes within the chosen community. By leveraging AI, we aim to provide real-time medical advice, reduce the burden on physical healthcare facilities, enhance disease surveillance, and promote proactive healthcare management.

2.1 Business Success Criteria:

- ***Prioritise engineering over data science***
- ***Structure project with concrete milestones:***
  1. **Finished Prototype**: Promising 1 day—2 weeks
  2. **Offline tested system**: Tune the model and rigorously test it on existing data 2–4 weeks
  3. **Online tested system**: Finalise the model and test it live 2–4 weeks
  4. **Going live**: Automate data updates, model training, and code deployment 2–4 weeks
  5. **Continuous improvement**: (optional) 12 months

3.1 Tools and Techniques:

(1) feedforward neural networks,

(2) feedback neural networks, machine learning algorithms:

(3) Support Vector Machine (SVM),

(4) Decision Tree (DT),

(5) Logistic Regression (LR),

(8) Naïve Bayes (NB),

(9) Deep Learning (DL),

(10) Generalized Linear Model (GLM),

(11) Random Forest (RF) and

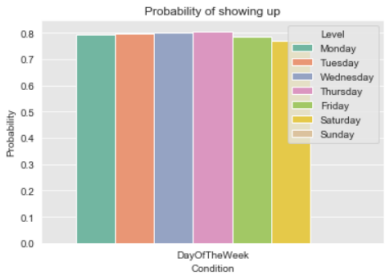
(12) K-Nearest Neighbours (KNN)

3.2 Time Series:

Time series processes will be speeded up by connecting pieces of images taken and data received since some of the data in the libraries is placed in order then time series will handle and slice data to get the required information

3. MACHINE LEARNING APPROACH:

- Our app improves how it performs diagnoses tasks based on how the machine performed in experience.
- It will also differ from performing the task always the same way because it has learned to do so,
- It will be able to do future predictions about the patient.
- The goal of this semi-supervised machine learning is to label data with a large amount of unlabelled data The machine will also be represented as a deep learning algorithm because it will concern neural network learning for modelling a large amount of unlabelled data and semi-structured data, this deep learning algorithm will allow us to represent data through the use of neural nets.



In [554…

```
sns.barplot(data = probStatusCategorical(['Gender']),
            x = 'Condition', y = 'Probability', hue = 'Level', palette = 'Set2')
plt.title('Probability of showing up')
plt.ylabel('Probability')
plt.show()
```

3.3 NATURAL LANGUAGE PROCESSING:

Our machine included NLP which will enable it to understand human communication for extracting different information. This will enable our machine to analyse texts, human speech, and documents. This option or algorithm will then enhance the security by adding a voice security check "your voice is your password" but it will focus more on processing speech than processing text. For information extraction, our machine will extract goal (input text), unstructured text (emails and documents), and also structured text (database table).

3.4 DATA:

Data received will be documents grouped by ID number as a primary key, data will be updated every 5 seconds, and other data will be unlabelled data like images

MODEL:

1- Statistical model will represent the relationship or mapping between a set of inputs and outputs

- New data will predict the output

-The system will be fed with data indicating that the rate of heart disease is higher when the weather is hot and around 7 pm to 10 pm, even rising sharply during higher degrees temperature, this Statistical model will predict the rate of this year's heart diseases to previous year's rate based on weather condition

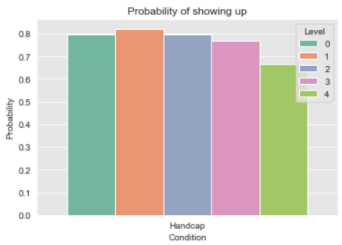
Heart disease per day = Average temperature x 2

THE FOLLOWING STEPS WILL BE TAKEN FOR OUR DATA CLEANING**/ DATA SCRUBBING PROCESS:**

Step 1: collecting the data after establishing our objectives, we created a strategy for collecting and aggregating the appropriate data, already we have determined which data we need, is numeric such as ID numbers, descriptive data, and user reviews such data is directly collected from users including data like images and fingerprints it will be collected in the form of information from the application this first-party data is usually structured and organized in a clear and defined way.

Step 2: Cleaning data (using pandas python library) once we have collected the data the machine will do an analysis and clean it to make sure that we are working with high-quality data. We will clean it by removing duplicates, major errors, and unwanted data points and also bringing structure to the data.

Step 3: Regression analysis will be implemented after cleaning the data.




```
In [556…
sns.barplot(data = probStatusCategorical(['SMS_received']),
            x = 'Condition', y = 'Probability', hue = 'Level', palette = 'Set2'
plt.title('Probability of showing up')
plt.ylabel('Probability')
plt.show()
```

```
In [557…
#convert gender to numerical values
df['Gender'] = df['Gender'].replace('F', 0)
df['Gender'] = df['Gender'].replace('M', 1)
```

Multicollinearity Analysis

```
In [473…
X.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110527 entries, 0 to 110526
Data columns (total 8 columns):
Gender          110527 non-null int64
Age             110527 non-null int64
Scholarship     110527 non-null int64
Hipertension    110527 non-null int64
Diabetes        110527 non-null int64
Alcoholism      110527 non-null int64
Handcap         110527 non-null int64
SMS_received    110527 non-null int64
dtypes: int64(8)
memory usage: 6.7 MB
```

```
In [474…
X.corr()
```

DEEP LEARNING:

The highest average sensitivity is achieved also when the DL algorithm was used.

This algorithm will be used in a more increasing number of collaborations, on which conclusions are drawn, and important information for clinical practice is obtained. Taking into account all possible risk factors that lead to the occurrence of a cardiovascular incident will lead to a comprehensive analysis of the patient, and the collection of a large amount of data.

The stated path in the treatment of the patient will have to be based on the conclusions from the analysis, which will enable the use of DL algorithms.

REFERENCES

1. HTTPS://BOKSBURGADVERTISER.CO.ZA/411838/LONG-QUEUE-AT-THE-BOKSBURG-NORTH-SAPS-FIREARM-OFFICE-CAUSES-HEADACHE-FOR-RESIDENT/AMP

2. HTTPS://WWW.DATAREVENUE.COM/EN-BLOG/AI-PROJECT-SUCCESS-FACTORS

3. HTTPS://WWW.SPICEWORKS.COM/TECH/ARTIFICIAL-INTELLIGENCE/ARTICLES/BUSINESS-APPLICATIONS-OF-MACHINE-LEARNING/

4. HTTPS://WWW.TABLEAU.COM/LEARN/ARTICLES/WHAT-IS-DATA-CLEANING