



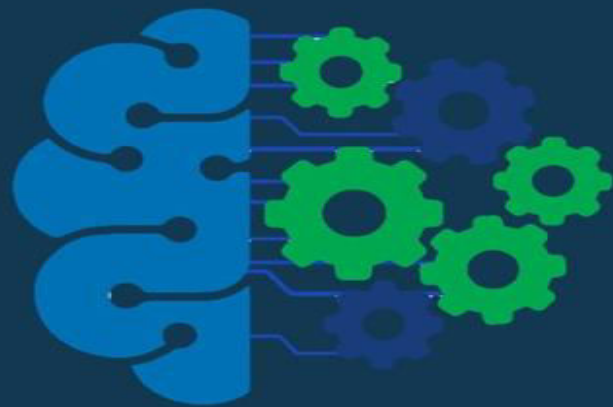
HEALTH ME UP!

SUBMITTED BY:

SRISHTY

ST-395

ROLL NO- 26



INDEX

SR.NO	PAGE NO.	CONTENT	TEACHER'S REMARK
1.	3	Acknowledgement	
2.	4	Introduction	
3.	5	Python	
4.	6	Data Science	
5.	7	Data Science with Python	
6.	9	Machine Learning	
7.	10	ML with Python	
8.	11	Decision Tree	
9.	14	Health Me Up	
10.	18	Conclusion	
11.	19	Reference	



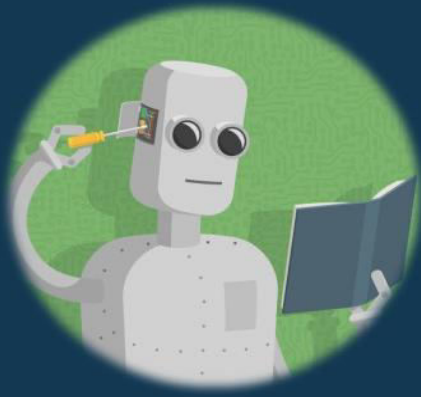
ACKNOWLEDGEMENT

I would like to express my sincere gratitude and special thanks to **Mrs. Anita Budhiraja** and **Mr. Sarwan Singh** who gave me a golden opportunity to learn under their vision and guidance.

It is with their help that I am able to learn and successfully make a project based on Machine Learning-**“Health Me Up”**.

Also a token of thanks to my parents and friends who helped me at every step to learn and grow.

Srishty



INTRODUCTION

Past few decades, we have seen a huge revolution in the Technical World. From room sized computer to pocket sized gadgets, from writing codes in low level languages to writing codes in high level languages and from waiting for hours for a single command to execute to executing the whole code in matter of a short while, everything has beautifully evolved.

With the advancement in Technology, one of the field that has bought a revolution to the Technical world is Artificial Intelligence and Data Science. The robots and gadgets are getting more and more intelligent and therefore making human life easier. Artificial Intelligence is a huge world in it itself. Its like a big umbrella and Machine Learning, Data Science are a part or subset of it.

The time to come will experience a huge revolution due to the further evolution of Artificial Intelligence. The lives of humans will be easier and will lead to evolution of many things advantageous to human lives.



pythonTM

Python is an interpreted, object-oriented programming language like PERL, that has gained popularity because of its clear syntax and readability. Python is said to be relatively easy to learn and portable, meaning its statements can be interpreted in several operating systems. Python was created by Guido van Rossum, a former resident of the Netherlands, whose favorite comedy group at the time was Monty Python's Flying Circus.

Python uses dynamic typing, and a combination of reference counting and a cycle-detecting garbage collector for memory management. It also features dynamic name resolution (late binding), which binds method and variable names during program execution.

Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python supports modules and packages, which encourages program modularity and code reuse.

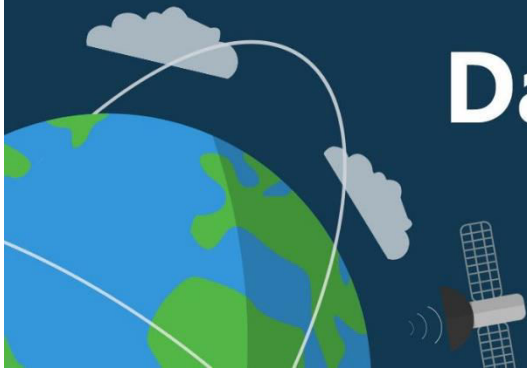


Data Science

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured

Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.

When Harvard Business Review called it "The Sexiest Job of the 21st Century", the term "data science" became a buzzword, and is now often applied to business analytics, business intelligence, predictive modeling, or any arbitrary use of data, or used as a glamorized term for statistics. In many cases, earlier approaches and solutions are now simply rebranded as "data science" to be more attractive, which can cause the term to become "dilute beyond usefulness."



Data Science & Analytics with Python

Python is often the choice for developers who need to apply **statistical techniques** or **data analysis** in their work, or for data scientists whose tasks need to be integrated with web apps or production environments. Python really shines in the field of machine learning. Its combination of machine learning libraries and flexibility makes Python uniquely well-suited to developing sophisticated models and prediction engines that plug directly into production systems.

Some of Python Libraries for Data Science and Data Analytics are:

1. **NumPy** is the foundational library for scientific computing in Python, and many of the libraries on this list use NumPy arrays as their basic inputs and outputs. In short, NumPy introduces objects for multidimensional arrays and matrices, as well as routines that allow developers to perform advanced mathematical and statistical functions on those arrays with as little code as possible.
2. **SciPy** builds on NumPy by adding a collection of algorithms and high-level commands for manipulating and visualizing data. This package includes functions for computing



Data Science & Analytics with Python

integrals numerically, solving differential equations, optimization, and more.

3. **Pandas** adds data structures and tools that are designed for practical data analysis in finance, statistics, social sciences, and engineering. Pandas works well with incomplete, messy, and unlabeled data (i.e., the kind of data you're likely to encounter in the real world), and provides tools for shaping, merging, reshaping, and slicing datasets.
4. **IPython** extends the functionality of Python's interactive interpreter with a souped-up interactive shell that adds introspection, rich media, shell syntax, tab completion, and command history retrieval. It also acts as an embeddable interpreter for your programs that can be really useful for debugging.
5. **matplotlib** is the standard Python library for creating 2D plots and graphs. It's pretty low-level, meaning it requires more commands to generate nice-looking graphs and figures than with some more advanced libraries. However, the flip side of that is flexibility. With enough commands, you can make just about any kind of graph you want with matplotlib.

MACHINE LEARNING



Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed. Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning can also be unsupervised and be used to learn and establish baseline behavioral profiles for various entities and then used to find meaningful anomalies.

Within the field of data analytics, machine learning is a method 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 "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.



MACHINE LEARNING WITH PYTHON

Machine_learning sits at the intersection of Artificial Intelligence and statistical analysis. By training computers with sets of real-world data, we're able to create algorithms that make more accurate and sophisticated predictions, whether we're talking about getting better driving directions or building computers that can identify landmarks just from looking at pictures. The following libraries give Python the ability to tackle several machine learning tasks, from performing basic regressions to training complex neural networks:-

1. **Scikit-learn** builds on NumPy and SciPy by adding a set of algorithms for common machine learning and data mining tasks, including clustering, regression, and classification. As a library, scikit-learn has a lot going for it. Its tools are well-documented, and its contributors include many machine learning experts.
2. **TensorFlow** is another high-profile entrant into machine learning, developed by Google as an open-source successor to DistBelief, their previous framework for training neural networks. TensorFlow uses a system of multi-layered nodes that allow you to quickly set up, train, and deploy artificial neural networks with large datasets. It's what allows Google to identify objects in photos or understand spoken words in its voice-recognition app.
3. **NLTK** is a set of libraries designed for Natural Language Processing (NLP). NLTK's basic functions allow you to tag text, identify named entities, and display parse trees, which are like sentence diagrams that reveal parts of speech and dependencies. We can do sentiment analysis with it.



DECISION TREE

A **decision tree** is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

Decision Trees (DTs) are a non-parametric supervised learning method used for [classification](#) and [regression](#). The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal but are also a popular tool in machine learning.



DECISION TREE

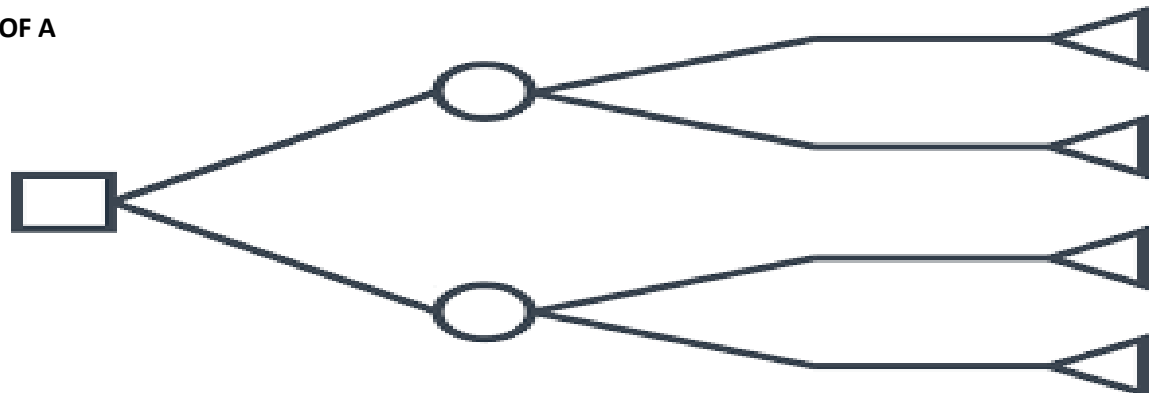
A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules.

In decision analysis, a decision tree and the closely related influence diagram are used as a visual and analytical decision support tool, where the expected values(or expected utility) of competing alternatives are calculated.

A decision tree consists of three types of nodes:

1. Decision nodes – typically represented by squares
2. End nodes – typically represented by triangles
3. Chance nodes – typically represented by circles

REPRESENTATION OF A
DECISION TREE:





DECISION TREE

The decision tree can be linearized into **decision rules**, where the outcome is the contents of the leaf node, and the conditions along the path form a conjunction in the if clause. In general, the rules have the form:

if condition1 and condition2 and condition3 then outcome

Among decision support tools, decision trees (and influence diagrams) have several advantages:

- Are simple to understand and interpret. People are able to understand decision tree models after a brief explanation.
- Have value even with little hard data. Important insights can be generated based on experts describing a situation (its alternatives, probabilities, and costs) and their preferences for outcomes.
- Allow the addition of new possible scenarios.
- Help determine worst, best and expected values for different scenarios.
- Can be combined with other decision techniques.



Health Me Up!

HEALTH ME UP is a Health Prediction Application based on Machine Learning.

It is coded in Python and its GUI is coded in Tkinter-a python library for designing effective and catchy GUI. The application is coded on PyCharm IDE.

The Data Set used for training was fetched from Kaggle

<https://www.kaggle.com/yersever/500-person-gender-height-weight-bodymassindex/version/2>

The file was a .csv file. The data was used for training the model. The file contained approximately 500 entries.

The various in columns included “Gender”, “Height (in m)”, “Weight (in kg)”, “Body Mass Index” and “Index”.

From the above data set, the model predicts about the health status of an individual.



Health Me Up!

The reason for choosing “Health me up” as the project was that in the present time the old quote comes to action that “Health is Wealth”.

Health is defined as a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.

Every person in today’s time is conscious about his or her Health and the one who values it rises ahead of others.

The application is very simple to use for people of all age groups. Its interface is very simple and easy to use.

Just we have to input our Gender, Height and Weight. That’s all!

It will predict the Health Status and we can then take preventive measures to ensure that we fall under the Normal Category.



Health Me Up!

The application is trained with Decision Tree Classifier Algorithm in Machine Learning.

The data set is having 500 different entries.

```
In [2]: #Fetching the Data
data=pd.read_csv("BMI_Data.csv")
data_dataframe=pd.DataFrame(data)
data_dataframe.shape
```

Out[2]: (500, 5)

The Columns in the Data are:

```
In [11]: data_dataframe.head()
```

Out[11]:

	Gender	Weight(in Kg)	Height(in m)	BMI	Index
0	0	96	1.74	31.708284	Obesity
1	0	87	1.89	24.355421	Normal
2	1	110	1.85	32.140248	Obesity
3	1	104	1.95	27.350427	Overweight
4	0	61	1.49	27.476240	Overweight



Health Me Up!

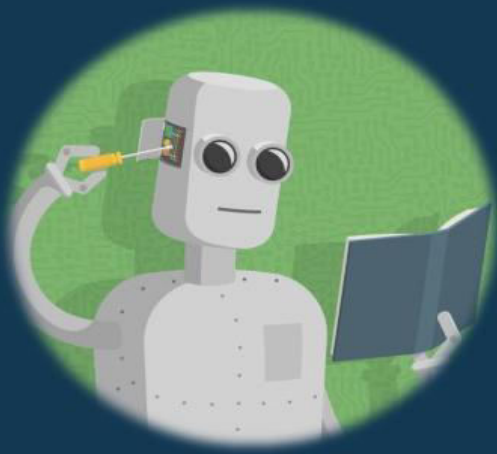
The validation size was set to 20% i.e. the data was randomly split into 80% for Training Data and 20% for Validation Data .

```
In [14]: # Splitting-out validation dataset
array = data_dataframe.values
X = array[:,0:4]
Y = array[:,4]
validation_size = 0.20
seed = 7
X_train, X_validation, Y_train, Y_validation = model_selection.train_test_split(X, Y, test_size=validation_size,
                                                                                random_state=seed)
```

The outcoming accuracy in case of Decision Tree Classifier comes out to be nearly 95.2 %

```
In [19]: print("1) Using Decision Tree Prediction ",model_selection.cross_val_score(clf,X,Y,scoring=scoring).mean()*100)
```

```
1) Using Decision Tree Prediction 95.18150781357454
```



CONCLUSION

Because of new computing technologies, machine learning today is not like machine learning of the past. It was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks; researchers interested in artificial intelligence wanted to see if computers could learn from data. The iterative aspect of machine learning is important because as models are exposed to new data, they can independently adapt. They learn from previous computations to produce reliable, repeatable decisions and results. It's a science that's not new – but one that has gained fresh momentum.

With the rapid increase in this field and increasing number of contributions to this field from all over the Globe, Health Me Up is a small contribution to the world of Machine Learning.



REFERENCES

- https://www.sas.com/en_in/insights/analytics/machine-learning.html
- <https://www.python.org/>
- <https://www.upwork.com/hiring/data/15-python-libraries-data-science/>
- <https://www.kaggle.com/yersever/500-person-gender-height-weight-bodymassindex/version/2>
- <http://scikit-learn.org/stable/modules/tree.html>



**THANK
YOU**