

ABSTRACT

Health is the most important in every human's life. Weekly or monthly check up of one's health is most important for the prevention and also to stay healthy. Nowadays, the individual is not having that much time to go for health check-up. Recently, due to covid-19, no one are willing to go to hospital for health check-up due to the fear of spreading virus. In this situation, technology plays an important role. In this project, we have used Machine Learning. Machine Learning is the study of computer algorithms that improve automatically from the previous experience. It is widely used nowadays and it is the most efficient domain in health care. We will develop a GUI to get the symptoms from the user. The models used in this project are Naive Bayes and Decision Tree. The output is the disease, the accuracy of model, its definition and the treatment of the particular disease based on the symptoms given by the individual. As we all know the saying which tells that "Prevention of the disease at an early stage is much better than the cure which we take after we get affected by the disease". This project shows detailed explanation of how to find the diseases from symptoms, so that the individual can contact the respective doctor and stay healthy at an early stage.

TABLE OF CONTENTS

Chapter No.	TITLE	Page No.
	ABSTRACT	v
	LIST OF FIGURES	viii
	LIST OF ABBREVIATIONS	ix
1	INTRODUCTION	1
	1.1. OVERVIEW	2
	1.2. MACHINE LEARNING	3
	1.3 MACHINE LEARNING STRATEGIES	3
	1.3.1. SUPERVISED LEARNING	3
	1.3.2. UNSUPERVISED LEARNING	4
2	LITERATURE SURVEY	6
	2.1. RELATED WORK	6
3	METHODOLOGY	8
	3.1. EXISTING SYSTEM	8
	3.2. PROPOSED SYSTEM	8
	3.3. OBJECTIVE	8
	3.4. SOFTWARE AND HARDWARE REQUIREMENTS	9
	3.4.1. SOFTWARE REQUIREMENTS	9
	3.4.2. HARDWARE REQUIREMENTS	9
	3.4.3. LIBRARIES	9
	3.5. PROGRAMMING LANGUAGES	10
	3.5.1 PYTHON	10
	3.5.2. DOMAIN	10
	3.6. SYSTEM ARCHITECTURE	11
	3.7. ALGORITHMS USED	11

	3.7.1. NAIVE BAYES	11
	3.7.2. DECISION TREE	12
	3.8 MODULES	13
	3.8.1 DATASET COLLECTION	13
	3.8.2. TRAIN AND TEST THE MODELS	13
	3.8.3. DEPLOY THE MODELS	13
4	RESULTS AND DISCUSSION	14
	4.1. WORKING	14
5	CONCLUSION	15
	5.1. CONCLUSION	15
	REFERENCES	16
	APPENDICES	17
	A. SOURCE CODE	17
	B. SCREENSHOTS	46

LIST OF FIGURES

Figure No.	Figure Name	Page No.
1.1.	MACHINE LEARNING CLASSIFICATION	5
1.2.	MACHINE LEARNING TASK	5
3.1	SYSTEM ARCHITECTURE	11
3.2	DECISION TREE	12

LIST OF ABBREVIATIONS

ABBREVIATIONS	EXPANSION
ML	Machine Learning
NB	Naive Bayes
DT	Decision Tree
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

CHAPTER 1

INTRODUCTION

Health is the most important in every human's life. Weekly or monthly check up of one's health is most important for the prevention and also to stay healthy. Nowadays, the individual is not having that much time to go for health check-up. Recently, due to covid-19, no one are willing to go to hospital for health check-up due to the fear of spreading virus.

As we all know the saying which tells that "Prevention of the disease at an early stage is much better than the cure which we take after we get affected by the disease".

Healthcare is the most crucial parts of the human life. Nowadays, so many are not willing to go to hospital, due to work overload and negligence of their health. The doctors and nurses are putting up maximum efforts to save people's lives without even considering their own loves. There are also some villages which lack medical facilities.

We have designed a disease prediction system using ML algorithm (Naive Bayes), find the most accurate algorithm, and used it to find the disease and Tkinter for GUI. And also, we have created a chatbot using Decision Tree, which will help us in getting accurate predictions by taking into account the symptoms faced by an individual.

This project helps to get the idea about the disease of an individual based on the symptoms he/she have, and get the treatment easily by contacting the concern doctor.

A disease predictor can also be called as a virtual doctor, which can predict the disease based on symptoms. Recently due to covid-19 no one are willing to go outside. This disease predictor system can be a most useful as it identifies the disease without even contacting the individual.

1.1 OVERVIEW

A disease is a condition that affects the individual functioning of body totally. Diseases if neglected will lead to the death of an individual. Diseases can be identified by the symptoms of the body of an individual. Health is the most important in every human's life. Weekly or monthly check up of one's health is most important for the prevention and also to stay healthy.

Nowadays, the individual is not having that much time to go for health check-up. Recently, due to covid-19, no one are willing to go to hospital for health check-up due to the fear of spreading virus. As we all know the saying which tells that "Prevention of the disease at an early stage is much better than the cure which we take after we get affected by the disease".

Healthcare is the most crucial parts of the human life. Nowadays, so many are not willing to go to hospital, due to work overload and negligence of their health. The doctors and nurses are putting up maximum efforts to save people's lives without even considering their own loves. There are also some villages which lack medical facilities.

Accurate and on-time analysis of any health-related problem is important for the prevention and treatment of the illness. The traditional way of diagnosis may not be sufficient in the case of a serious ailment. In this situation, where everything has turned virtual, the doctors and nurses are putting up maximum efforts to save people's lives even if they have to danger their own.

There are also some remote villages which lack medical facilities. The dataset was processed in ML models Naive Bayes and Decision Tree. While processing the data, symptoms are given as input and the disease was received as an output. This project helps to get the idea about the disease of an individual based on the symptoms he/she have, and get the treatment easily by contacting the concern doctor.

1.2 MACHINE LEARNING

Machine learning could be a subfield of computer science (AI). The goal of machine learning typically is to know the structure information of knowledge of information and match that data into models which will be understood and used by folks. Although machine learning could be a field inside technology, it differs from ancient process approaches.

In ancient computing, algorithms are sets of expressly programmed directions employed by computers to calculate or downside solve. Machine learning algorithms instead give computers to coach on knowledge inputs and use applied math analysis so as to output values that fall inside a particular vary. thanks to this, machine learning facilitates computers in building models from sample knowledge so as to modify decision-making processes supported knowledge inputs.

1.3 Machine Learning Strategies

In machine learning, tasks square measure typically classified into broad classes. These classes square measure supported however learning is received or however feedback on the educational is given to the system developed. Two of the foremost wide adopted machine learning strategies square measure supervised learning that trains algorithms supported example input and output information that's tagged by humans, and unattended learning that provides the algorithmic program with no tagged information so as to permit it to search out structure at intervals its computer file.

1.3.1 *Supervised Learning*

In supervised learning, the pc is given example inputs that square measure labelled with their desired outputs. The aim of this technique is for the algorithmic program to be ready to “learn” by comparison its actual output with the “taught” outputs to search out errors, and modify the model consequently. Supervised learning thus uses patterns to predict label values on extra unlabelled information. For example, with supervised learning, an algorithm may be fed data with images of

sharks labelled as fish and images of oceans labelled as water. By being trained on this data, the supervised learning algorithm should be able to later identify unlabelled shark images as fish and unlabelled ocean images as water.

A common use case of supervised learning is to use historical information to predict statistically probably future events. It's going to use historical stock exchange info to anticipate approaching fluctuations, or be used to filter spam emails. In supervised learning, labeled photos of dogs are often used as input file to classify unlabeled photos of dogs.

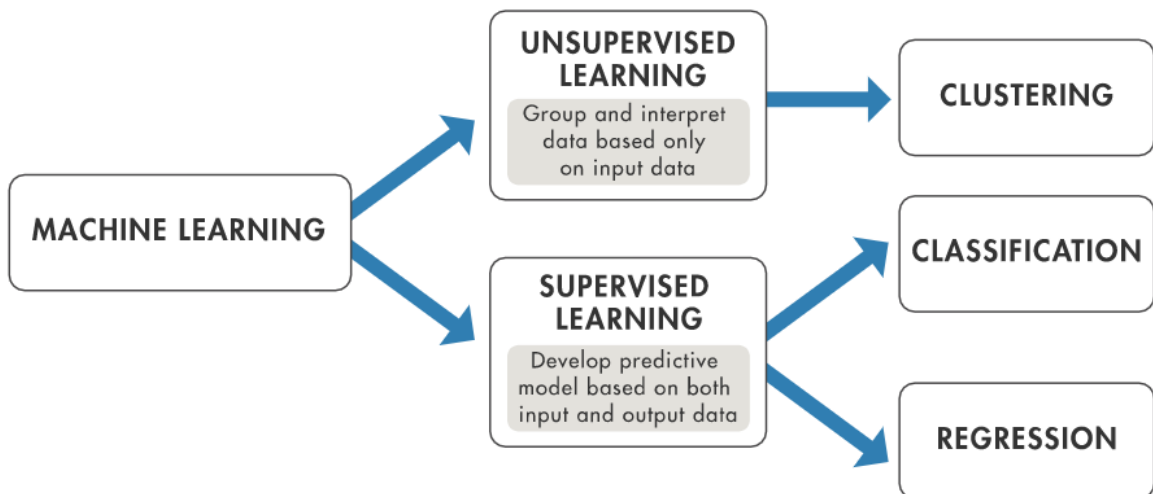
1.3.2 *Unattended Learning*

In unattended learning, information is unlabeled, that the learning rule is left to seek out commonalities among its input file. The goal of unattended learning is also as easy as discovering hidden patterns at intervals a dataset, however it should even have a goal of feature learning, that permits the procedure machine to mechanically discover the representations that square measure required to classify data.

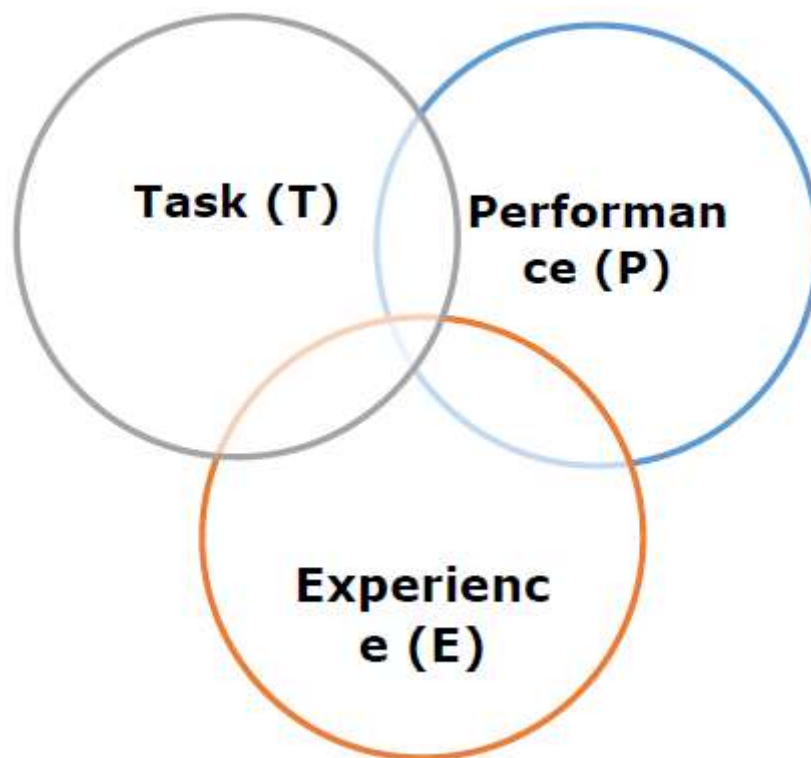
Unsupervised learning is usually used for transactional information. You will have an oversized dataset of consumers and their purchases, however as a person's you'll probably not be able to add up of what similar attributes will be drawn from client profiles and their styles of purchases.

With this information fed into Associate in Nursing unattended learning rule, it should be determined that ladies of a definite age vary UN agency obtain unscented soaps square measure probably to be pregnant, and so a promoting campaign

associated with physiological condition and baby will be merchandised.



1.1 MACHINE LEARNING CLASSIFICATION



1.2 MACHINE LEARNING TASK

CHAPTER 2

LITERATURE SURVEY

2.1 RELATED WORK

- Tarigoppula V.S Sriram et al. in [1] collects the voice dataset from UCI Machine Learning repository and train four algorithms on that dataset. The result is the prediction of Parkinson Disease by considering the most accurate algorithm.
- Shubham Bind et al. in [2] studies about all the available researches in literature to predict the Parkinson diseases.
- K. Gomathi, D. Shanmuga Priya in [3] used different data mining techniques to predict Heart disease, Breast Cancer, Diabetes. The models used are Decision Tree and Naive Bayes Classifier. Performance of both the models was compared and the best classifier is used to predict the above diseases.
- Isha Pandya et al. in [4] used two supervised machine learning algorithms Decision Tree, accuracy 91% and Naïve Bayes classifier, accuracy 87%. Here, they used the combination of both to get the best accuracy. Naïve Bayes Classifier accuracy should be improved.
- Akash C. Jamgade, Prof. S. D. Zade in [5] paper determined the most danger diseases which occur in a person in a locality and community. But, the data collection is difficult.
- Siddhika Arunachalam in [6] six classification algorithms are used after analyzing 14 attributes in the dataset. But, we may get confused which algorithm to use.
- Ionela-Catalina ZAMFIR, Ana-Maria Mihaela IORDACHE methodologies in [7] used are Support Vector Machines, Artificial Neural Networks, K-Means Algorithm, Decision Trees, Logistic Regression and predicted diseases are breast cancer, lung cancer, heart diseases, diabetes, thyroid or kidney diseases.