



# **Project Report on ML Based Medical Chatbot.**

***Submitted By Bishal Das and Sambhu Mahato.  
(Guided by Prof. Utpal Kumar Ray  
Jadavpur University.)***


# Contents

- What is Machine Learning?
  - Types of Learning
    - Supervised Learning
    - Unsupervised Learning
    - Reinforcement learning
  - Literature Survey
  - A Virtual Medical Chatbot
  - Disease Diagnosis
  - Implementation Details
  - Future Work
  - References
- 


# What is Machine Learning?

- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.
  - As intelligence requires knowledge, it is necessary for the computers to acquire knowledge.
- 

# Types of Learning

- **Supervised (inductive) learning**
    - Given: training data + desired outputs (labels)
  - **Unsupervised learning**
    - Given: training data (without desired outputs)
  - **Reinforcement learning**
    - Rewards from sequence of actions
- 

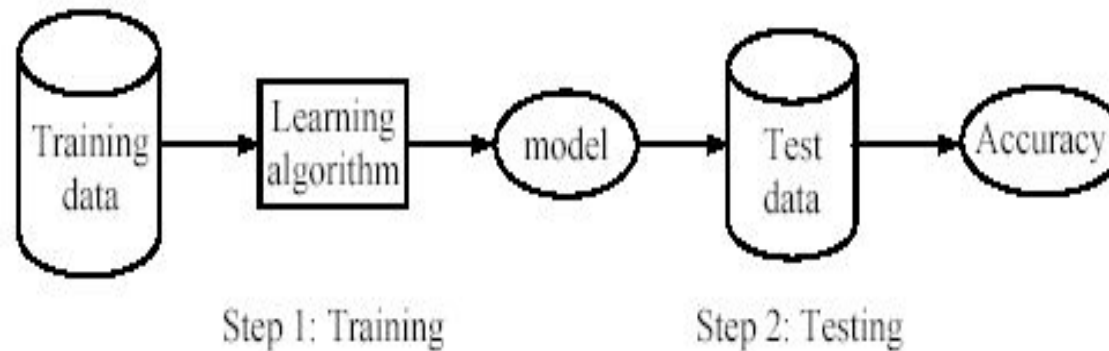
# Supervised Learning

- **Supervised learning:** classification is seen as supervised learning from examples.
  - **Supervision:** The data (observations, measurements, etc.) are labeled with pre-defined classes. It is like that a “teacher” gives the classes (**supervision**).
  - Test data are classified into these classes too.
- 

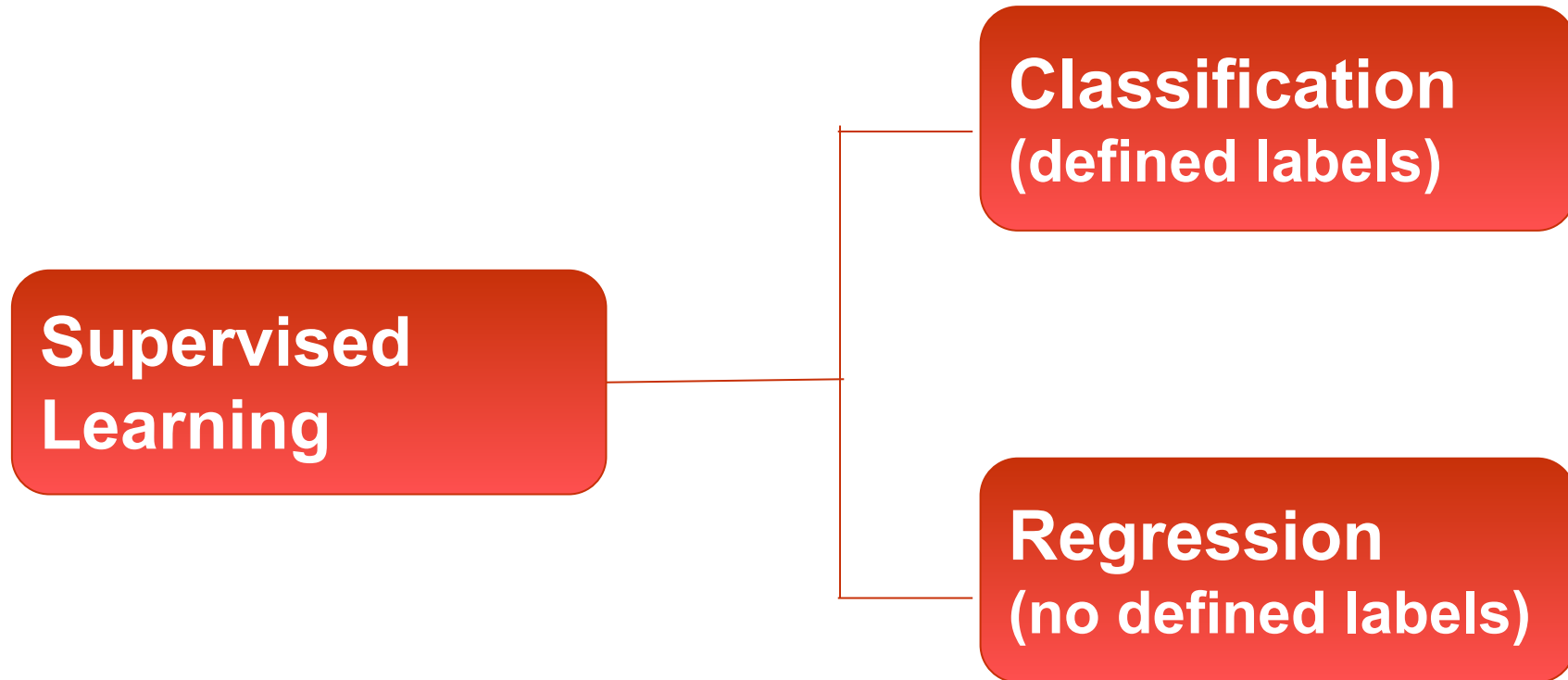
# Supervised learning process: two steps

- **Learning (training)**: Learn a model using the **training data**
- **Testing**: Test the model using **unseen test data** to assess the model accuracy,

$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$



# Types of Supervised Learning



# Learning Techniques: Decision tree

- **Decision tree learning** is one of the most widely used techniques for classification.
  - Its classification accuracy is competitive with other methods, and
  - it is very efficient.
- The classification model is a tree, called **decision tree**.

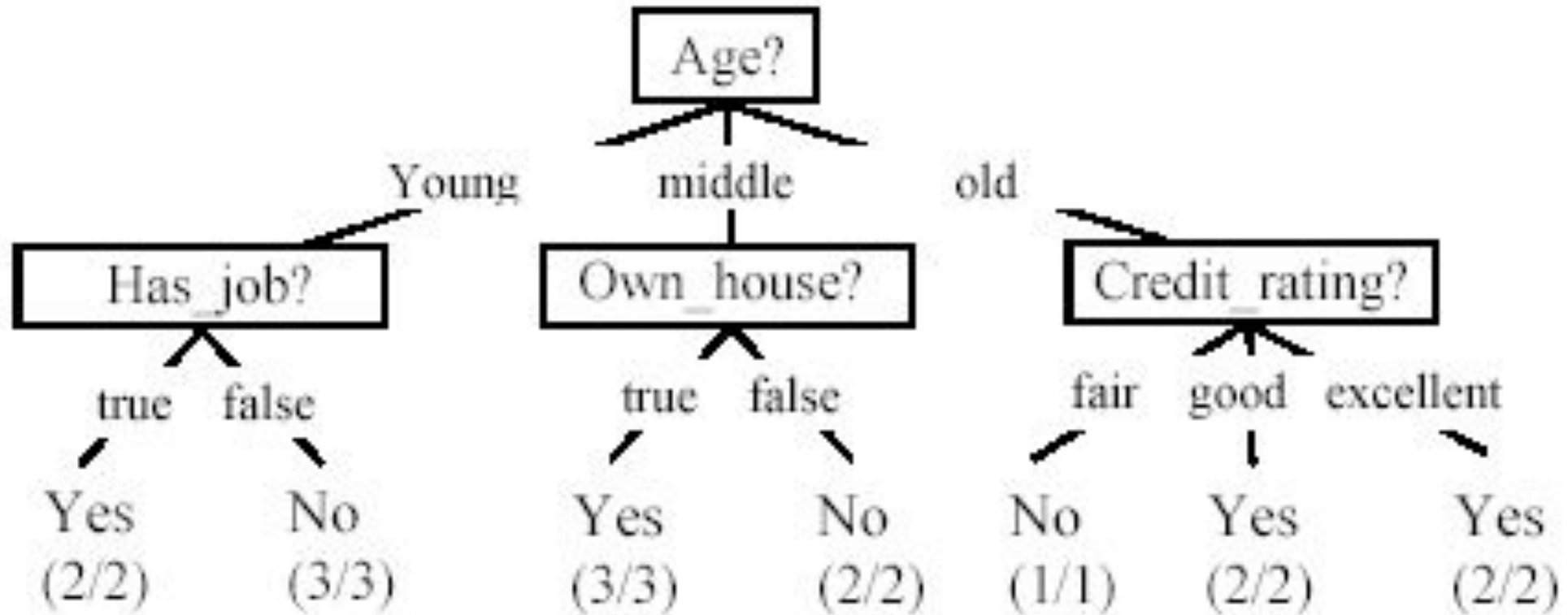


## Example:(Loan Data)

ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	young	false	false	fair	No
2	young	false	false	good	No
3	young	true	false	good	Yes
4	young	true	true	fair	Yes
5	young	false	false	fair	No
6	middle	false	false	fair	No
7	middle	false	false	good	No
8	middle	true	true	good	Yes
9	middle	false	true	excellent	Yes
10	middle	false	true	excellent	Yes
11	old	false	true	excellent	Yes
12	old	false	true	good	Yes
13	old	true	false	good	Yes
14	old	true	false	excellent	Yes
15	old	false	false	fair	No

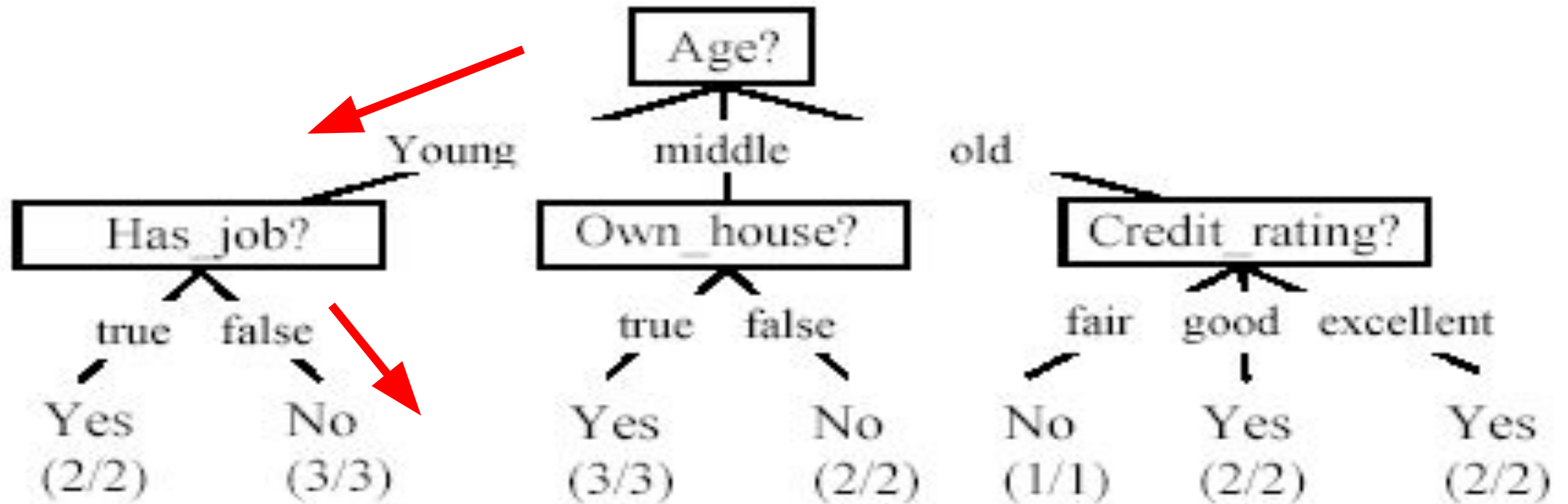
# A decision tree from the loan data

Decision nodes and leaf nodes (classes)




# Use the decision tree

Age	Has_Job	Own_house	Credit-Rating	Class
young	false	false	good	? No



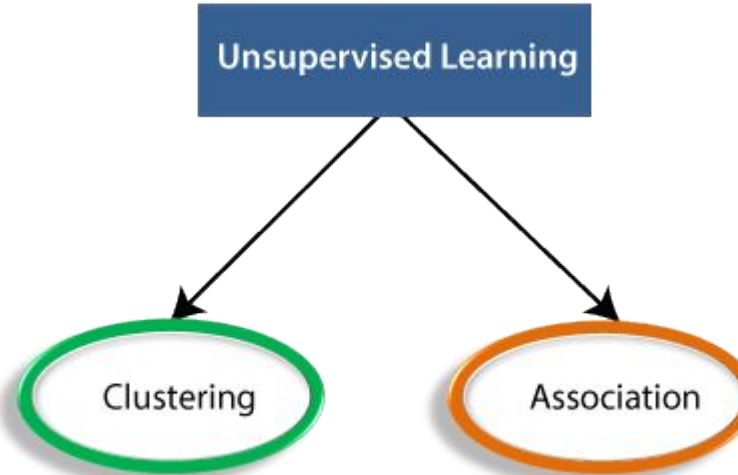
# Unsupervised Learning

- Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data.
- 

# Classification of unsupervised Learning

Unsupervised learning classified into two categories

- Clustering
- Association



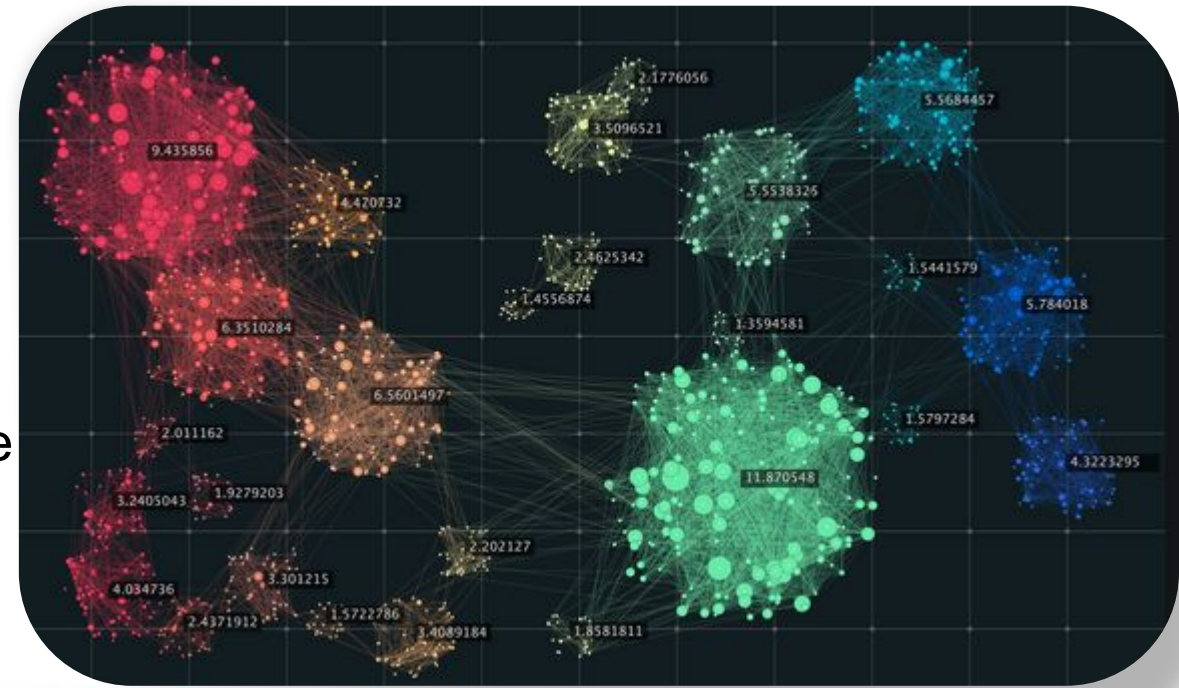


# Clustering and its types

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

## Two Types of Clustering

- ❑ Clustering algorithms: Construct various partitions and then evaluate them by some criterion
- ❑ Hierarchical algorithms: Grouping data into a tree of clusters. Hierarchical clustering begins by treating every data points as a separate cluster



# Clustering and its types

## 1. Clustering Algorithms:

### K-means clustering:

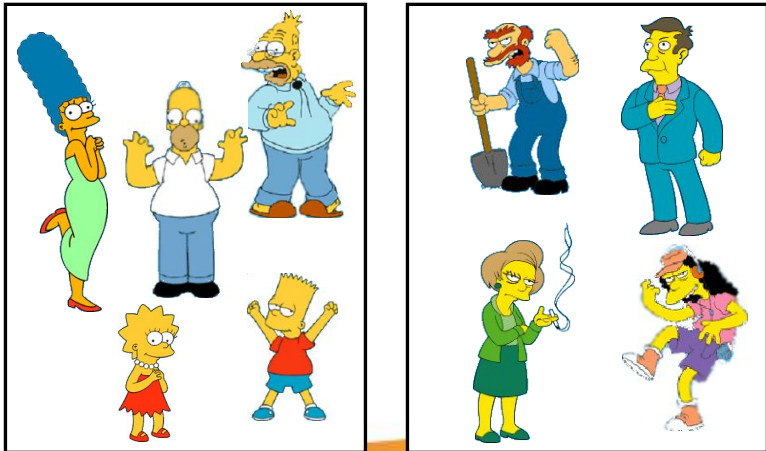
The objective of K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number ( $k$ ) of clusters in a dataset.

## 2. Hierarchical Clustering : The basic method to generate hierarchical clustering are

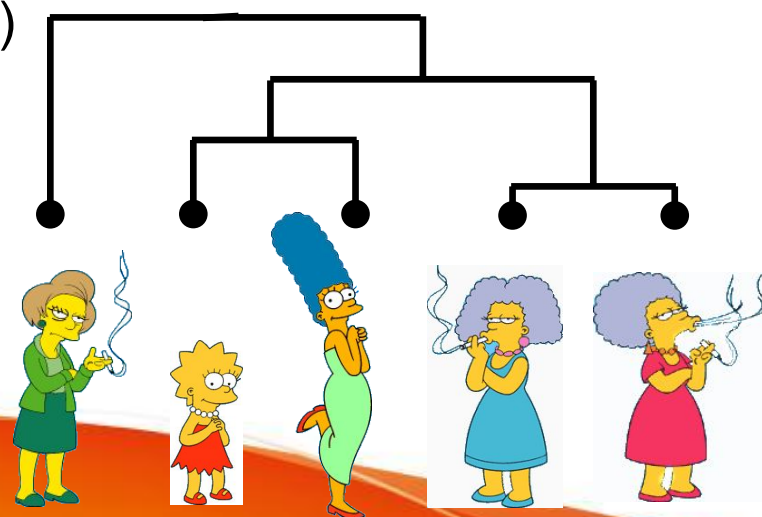
1) Agglomerative (Bottom-up)

2) Divisive (Top-down)

1)



2)



# Association

Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently a itemset occurs in a transaction. A typical example is Market Based Analysis.

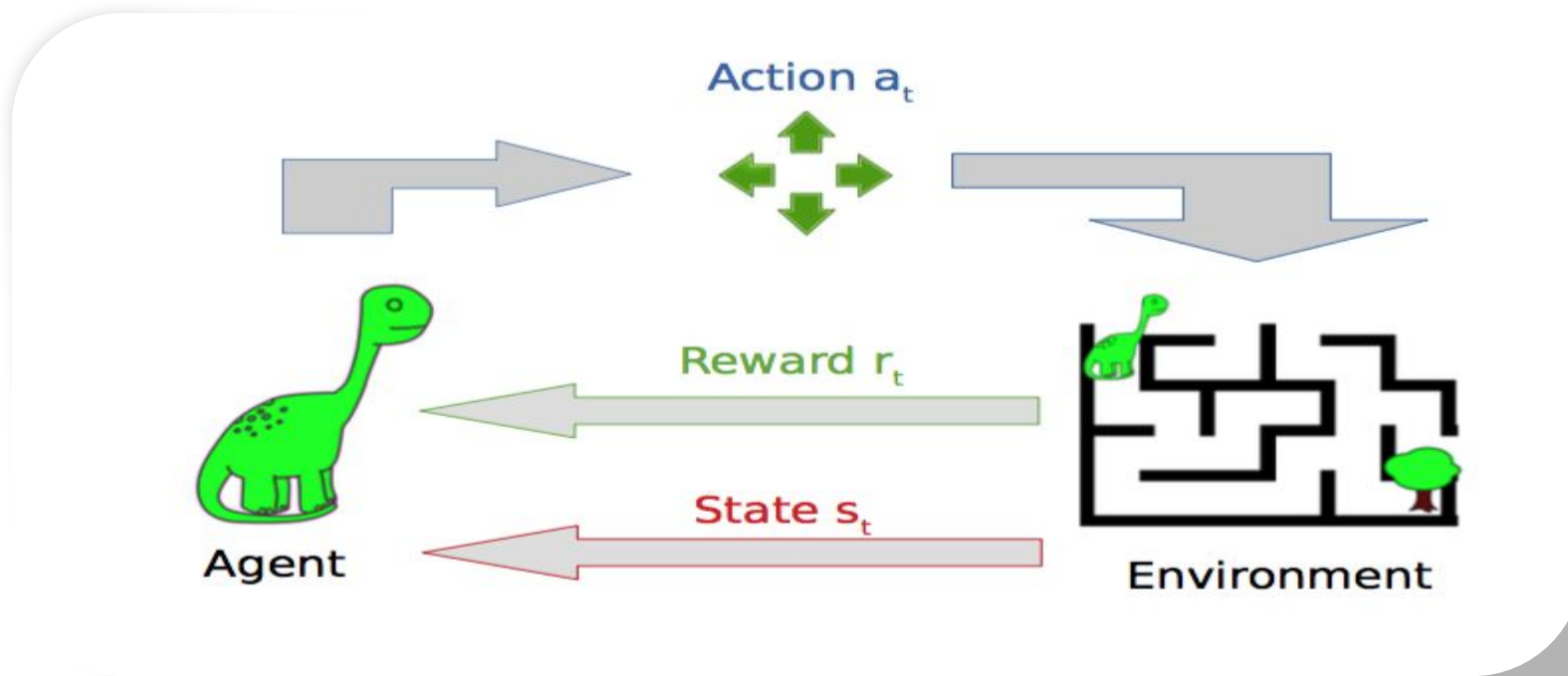
Given a set of transactions, we can find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction.

TID	ITEMS
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke



# Reinforcement learning

In this learning the problem faced by an agent that learns behavior through trial-error interactions with a dynamic environment with a dynamic environment.



**TABLE ON LITERATURE SURVEY**

<b>S. NO</b>	<b>PROJECT</b>	<b>JOURNAL</b>	<b>TECHNIQUES</b>	<b>RESULT</b>	<b>ISSUES</b>	<b>YEAR</b>
1.	Application of synonous text-based dialogue system in mental health interventions	<i>Medical Internet Research , volume: 19</i>	Web-based chat technologies are compared.	The review provides an evaluation of individual synchronous Web-based chat technologies as a mode of psychological intervention and support.	The issue of whether these technologies are cost effective in clinical practice remains a consideration for future research studies.	2017
2	Dr.Vdoc: A medical chatbot that acts as a virtual doctor	<i>Medical Science and Technology, Volume: 6</i>	Natural language processing and pattern matching algorithm.	80% of accuracy	The full consultation is given at a free of cost, which is not a good practice.	2017
3	A Novel approach for medical assistant using trained chatbot	<i>International Conference on Inventive Communication and Computational Technologies</i>	Natural language processing and pattern matching algorithm.	It is possible to predict any possible problem even before they start to cause any damage to the body.	Provides the prescription and composition of medicine without consultation of a doctor.	2017


**TABLE ON LITERATURE SURVEY**

S. NO	ARTICLE TITLE	JOURNAL	TECHNIQUES & ALGORITHMS	RESULT	ISSUES	YEAR OF PUBLICATION
4	Medchatbot: An UMLS based chatbot for medical students	<i>International Journal of Computer Applications Volume 55– No.17</i>	AIML pattern technique for Pattern matching.	47% of accuracy	The system has not been specially designed for the task of supporting natural dialog in chatbots	2016
5	Medical predictions system	<i>International Journal of Engineering Research and Applications , ISSN: 2248-9622 , Vol. 2, Issue 3</i>	Natural language processing.	These systems are created only to suggest a solution.	Won't be able to replace the most important factor in decision making.	2015
6	Designing a Chatbot for diabetics patients	<i>ACM Transactions on Management Information Systems (TMIS), Volume 4, Issue 2</i>	Sequence Words Deleted (SWD) technique. Pattern matching algorithm.	The conversation is to be controlled by chatbot rather than by user.	Keep repeating the previous question until the keywords is detected.	2015


**TABLE ON LITERATURE SURVEY**

<b>S. NO</b>	<b>ARTICLE TITLE</b>	<b>JOURNAL</b>	<b>TECHNIQUES &amp; ALGORITHMS</b>	<b>RESULT</b>	<b>ISSUES</b>	<b>YEAR OF PUBLICATION</b>
7	Conditional entropy based Retrivel model in Patient-Carer Conersational cases	<i>IEEE 30th International conference on Computer-Based Medical System</i>	Natural Language Processing.	The results show that even in small training datasets, this method outperforms up to date methods for automated communication.	The issue here is,it requires a compact Adjacency Matrix based on the dialogues.	2017
8	Towards a chatbot for digital counselling	<i>Medical Internet Research, 4(1), pp. e3.</i>	Natural language processing and pattern matching algorithm.	60% of accuracy	The issue here is to maintain the ethical considerations.	2017


# A Virtual Medical Chatbot: Purpose

- The basic idea is to develop a **user-friendly** virtual assistant that would help every person even in the last mile.
  - The user can converse with the Chatbot using a **regular language**.
  - The therapeutic Chatbots analyze the user's requests, recollect them and answer related queries.
  - This medical Chatbot **mimic** a Physician to stimulate a conversation that resembles a human.
- 

# Disease Diagnosis based on Patient's Symptoms

- Diseases are found based on the patients' response to the query.
  - The exact infection is recognized and communicated to the patient.
  - The Chatbot categorizes the recognized infection as a major or minor issue based on the conditions in the Chatbot. If it is a major issue, it refers an expert to the patient by sending the Physician details.
  - In case it is a minor issue, the Chatbot determines the illness and gives information about home remedies or first aids and advises the patients to visit a specialist immediately.
- 

# Implementation Details

- Python 3.x is used as the backend for co-operating with UI tools.
  - By utilizing NLP strategies and Amazon Polly, change-over is applied to the local language to convert it into International English.
  - Python modules such as sklearn, pandas has been used
  - It uses NLP to recognize text and interprets the user's queries using algorithms.
- 

# Screenshot of Dataset used

← → ↻ 🔒 kaggle.com/itachi9604/disease-symptom-description-dataset?select=dataset.csv

Apps YouTube Inbox (2,698) - bish... WhatsApp linkedin

☆ ⚙️ 👤 ⋮

📖 Reading list

Data Explorer

633.83 KB

📄 Symptom-severity.csv

📄 dataset.csv

📄 symptom\_Description.csv

📄 symptom\_precaution.csv

< dataset.csv (617.38 KB)

⬇️ ⚙️

Detail Compact Column

10 of 18 columns ▾

About this file

Disease with its symptoms.


▲ Disease	▲ Symptom_1	▲ Symptom_2	▲ Symptom_3	▲ Symptom_4	▲ Symptom_5	▲ Symptom_6
Diseases that may be present	the symptoms experienced during the disease	the symptoms experienced during the disease	the symptoms experienced during the disease	the symptoms experienced during the disease	the symptoms experienced during the disease	the symptoms experienced during the disease
41 unique values	vomiting fatigue Other (3408) 17% 14% 69%	vomiting fatigue Other (3648) 18% 8% 74%	fatigue high_fever Other (3870) 15% 7% 79%	high_fever [null] Other (4194) 8% 7% 85%	[null] headache Other (3366) 25% 7% 68%	[null] nausea Other (2544)
Fungal infection	itching	skin_rash	nodal_skin_eruptions	dischromic _patches		
Fungal infection	skin_rash	nodal_skin_eruptions	dischromic _patches			
Fungal infection	itching	nodal_skin_eruptions	dischromic _patches			
Fungal infection	itching	skin_rash	dischromic _patches			
Fungal infection	itching	skin_rash	nodal_skin_eruptions			
Fungal infection	skin_rash	nodal_skin_eruptions	dischromic _patches			
Fungal infection	itching	nodal_skin_eruptions	dischromic _patches			
Fungal infection	itching	skin_rash	dischromic _patches			

Summary



# Implementation Details(code snippet of modules)

```
import pandas as pd
import pytsx3
from sklearn import preprocessing
from sklearn.tree import DecisionTreeClassifier, _tree
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score
from sklearn.svm import SVC
import csv
import warnings
```



# Implementation Details(code snippet)

**This part of the code collects data and analyses them:**

```
training = pd.read_csv('Training.csv')
testing = pd.read_csv('Testing.csv')
cols = training.columns
cols = cols[:-1]
x = training[cols]
y = training['prognosis']
y1 = y

reduced_data = training.groupby(training['prognosis']).max()

# mapping strings to numbers
le = preprocessing.LabelEncoder()
le.fit(y)
y = le.transform(y)
```

# Implementation Details(code snippet)

**Training is done at this part and Model is created:**

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.33, random_state=42)
testx = testing[cols]
testy = testing['prognosis']
testy = le.transform(testy)
clf1 = DecisionTreeClassifier()
clf = clf1.fit(x_train, y_train)
scores = cross_val_score(clf, x_test, y_test, cv=3)
model = SVC()
model.fit(x_train, y_train)
print('Model is Ready!!!!!!!!!!!!!!')
importances = clf.feature_importances_
indices = np.argsort(importances)[::-1]
features = cols
```

# Implementation Details(code snippet)

The below function is called to give the actual output:

```
tree_to_code(clf,cols)
```

# Result

**SCREENSHOT 1: User is giving what he/she is feeling, as input and as a result some symptoms are listed.**

```
PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL
Model is Ready!!!!!!!!!!!!!!
Your Name
                                -> Hulk
hello Hulk
Enter the symptom you are experiencing
                                -> pain
searches related to input:
0 ) joint_pain
1 ) stomach_pain
2 ) pain_behind_the_eyes
3 ) back_pain
4 ) abdominal_pain
5 ) chest_pain
6 ) pain_during_bowel_movements
7 ) pain_in_anal_region
8 ) neck_pain
9 ) knee_pain
10 ) hip_joint_pain
11 ) muscle_pain
12 ) belly_pain
13 ) painful_walking
Select the one you meant (0 - {num}):
                                -> █
```

2: Python

# Result

**SCREENSHOT 2: More questions are being asked by the chatbot to give an appropriate solution to the user.**

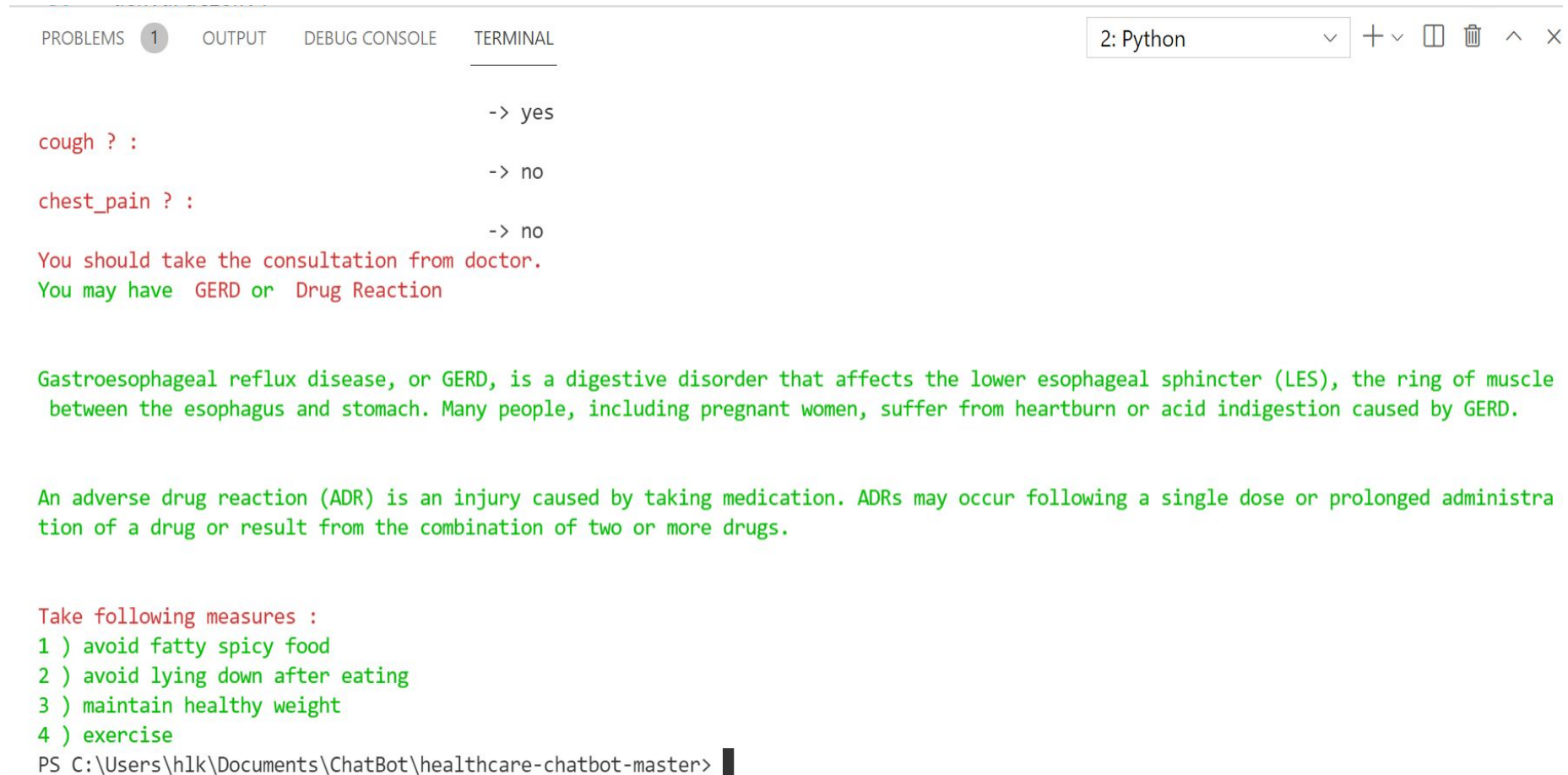
PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL

2: Python

Select the one you meant (0 - {num}):  
-> 1  
Okay. From how many days ? :  
-> 5  
Are you experiencing any  
stomach\_pain ? :  
-> yes  
acidity ? :  
-> yes  
ulcers\_on\_tongue ? :  
-> no  
vomiting ? :  
-> yes  
cough ? :  
-> no  
chest\_pain ? :  
-> no  
You should take the consultation from doctor.  
You may have GERD or Drug Reaction  
  
Gastroesophageal reflux disease, or GERD, is a digestive disorder that affects the lower esophageal sphincter (LES), the ring of muscle between the esophagus and stomach. Many people, including pregnant women, suffer from heartburn or acid indigestion caused by GERD.

# Result

**SCREENSHOT 3 : If the chatbot finds an appropriate solution of the disease then feedback is given the user otherwise if suggest to visit to doctor.**



```
PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL 2: Python + v [ ] [ ] ^ X

-> yes
cough ? :
-> no
chest_pain ? :
-> no
You should take the consultation from doctor.
You may have GERD or Drug Reaction


Gastroesophageal reflux disease, or GERD, is a digestive disorder that affects the lower esophageal sphincter (LES), the ring of muscle between the esophagus and stomach. Many people, including pregnant women, suffer from heartburn or acid indigestion caused by GERD.

An adverse drug reaction (ADR) is an injury caused by taking medication. ADRs may occur following a single dose or prolonged administration of a drug or result from the combination of two or more drugs.

Take following measures :
1 ) avoid fatty spicy food
2 ) avoid lying down after eating
3 ) maintain healthy weight
4 ) exercise
PS C:\Users\h1k\Documents\ChatBot\healthcare-chatbot-master>
```


# Conclusion

The review concludes that the usage of Chatbot is user friendly and can be used by any person who knows how to type in their own language in mobile app or desktop version. A medical chatbot provides personalized diagnoses based on symptoms. In the future, the bot's symptom recognition and diagnosis performance could be greatly improved by adding support for more medical features, such as location, duration, and intensity of symptoms, and more detailed symptom description.





# FUTURE WORK

- The smartness and intelligence of the chatbot can be increased by conducting more study and increasing the database so that Chabot could answer all type of question about every type of disease.
  - Audio system can also be included in this system to make this Chabot more interactive
- 

# References

- Saurav Kumar Mishra, Dharendra Bharti, Nidhi Mishra,” Dr. Vdoc: A Medical Chatbot that Acts as a Virtual Doctor”, Journal of Medical Science and Technology, Volume: 6, Issue 3, 2017
- <https://dzone.com/articles/python-chatbot-project-build-your-first-python-pro>
- <https://www.slideshare.net/embeddedvision/an-introduction-to-machine-learning-and-how-to-teach-machines-to-see-a-presentation-from-tryolabs>
- <https://towardsdatascience.com/machine-learning-basics-part-1-a36d38c7916>
- <https://www.kaggle.com/itachi9604/disease-symptom-description-dataset?select=Symptom-severity.csv>

Thank You...