LIVER PATIENT ANALYSIS USING MACHINE LEARNING

PROJECT REPORT
CONCEPTS

1.INTRODUCTION

1.1OVERVIEW

Liver is the largest internal organ in the human body,it is essential for digesting food and releasing the toxic element of the body and plays a major role in metabolism and serving several vital functions. The liver is the largest glandular organ of the body. It weighs about 3 lb (1.36 kg) .The liver's main job is to strain the blood coming from the digestive tract, before passing it to the rest of the body. The liver also detoxifies chemicals and metabolizes drugs. As it does so, the liver hides bile that ends up back in the intestines. The liver also makes proteins important for blood clotting and other functions.The liver supports almost every organ in the body and is vital for our survival. Liver disease may not cause any symptoms at earlier stage or the symptoms may be vague, like weakness and loss of energy. Symptoms partly depend on the type and the extent of liver disease. Liver diseases are diagnosed based on the liver functional test.

A healthy liver, shiny pinkish-brown triangle tucked under the right rib cage, is what we expect for every human being. Liver plays a vital role of removing toxins, converting digested food to energy, storing vitamins and minerals and controlling how much fat and sugar is sent back out to the rest of the body. Malfunctioning of liver can disturb functioning of other body parts. Hence, it is of utmost importance to detect the status of liver functioning at an early stage. Further investigations can be carried out once demarcation is done between normal and affected liver. Liver functioning can be diagnosed after checking the levels of various parameters given in the Liver Function Tests (LFT) in the pathological report of a patient. An automatic decision support system (DSS) can be of a great help in such cases for the correct diagnosis of a patient

In this work, KNN is used as a classification algorithm for the demarcation of normal and diseased liver. Liver dataset by BUPA Medical Research Ltd., available on UCI repository, is used. KNN is used by implementing ten different kernel functions. Evaluation is done using performance parameters accuracy, mean square error (MSE), sensitivity and specificity. Best kernel function is selected after comparing performance parameters.

1.1 PURPOSE

Liver disease refers to any condition that **causes damage to the liver and may affect its function**. It is a serious condition that threatens human life and requires urgent medical attention. Early prediction of the disease using machine learning (ML) techniques will be the point of interest in this study.

This study uses **machine learning** techniques to enhance the detection of **liver illness** in an effort to solve this problem. The major goal of this study is to distinguish between liver patients and healthy people using classification algorithms.

Over the last few decades, liver diseases are one of the reasons for a large number of death cases worldwide and have emerged as a life-threatening disease also. By WHO report, around 59% of the mortality and 46% of global diseases are because of chronic diseases and around 35 millions of people worldwide die due to chronic diseases. Now, we are living in the era of information where millions of data are generating from various sources every day. We can use these data to improve our healthcare services or proper identification of diseases. We have collected patient data from open source platform and applied various kinds of data analysis techniques and machine learning (ML) approaches applied to see the pattern of the data sets. Then, a performance comparison between these models is made to get highly accurate model for predicting liver disease

2.PROBLEM DEFINITION & DESIGN THINKING

2.1 PROBLEM DEFINITION:

Liver disease is a broad term that encompasses a range of conditions that affect the liver's function and structure. Some common liver diseases include:

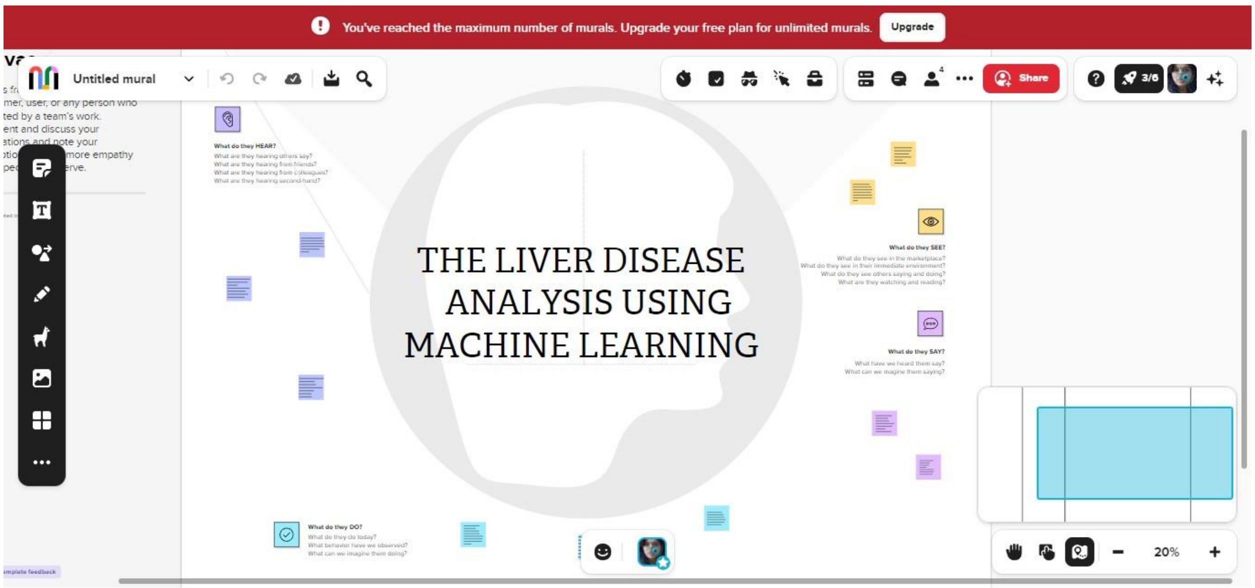
1. Hepatitis: Inflammation of the liver caused by a viral infection or exposure to toxins.
2. Cirrhosis: A chronic disease that results in scarring of the liver tissue, often caused by alcohol abuse or chronic viral hepatitis.

3. Fatty liver disease: A condition in which excess fat accumulates in the liver, often caused by obesity, high cholesterol, or diabetes.

4. Liver cancer: A type of cancer that originates in the liver cells.

.2.2 <u>DESIGN THINKING:</u>

2.2.1 EMPATHY MAP

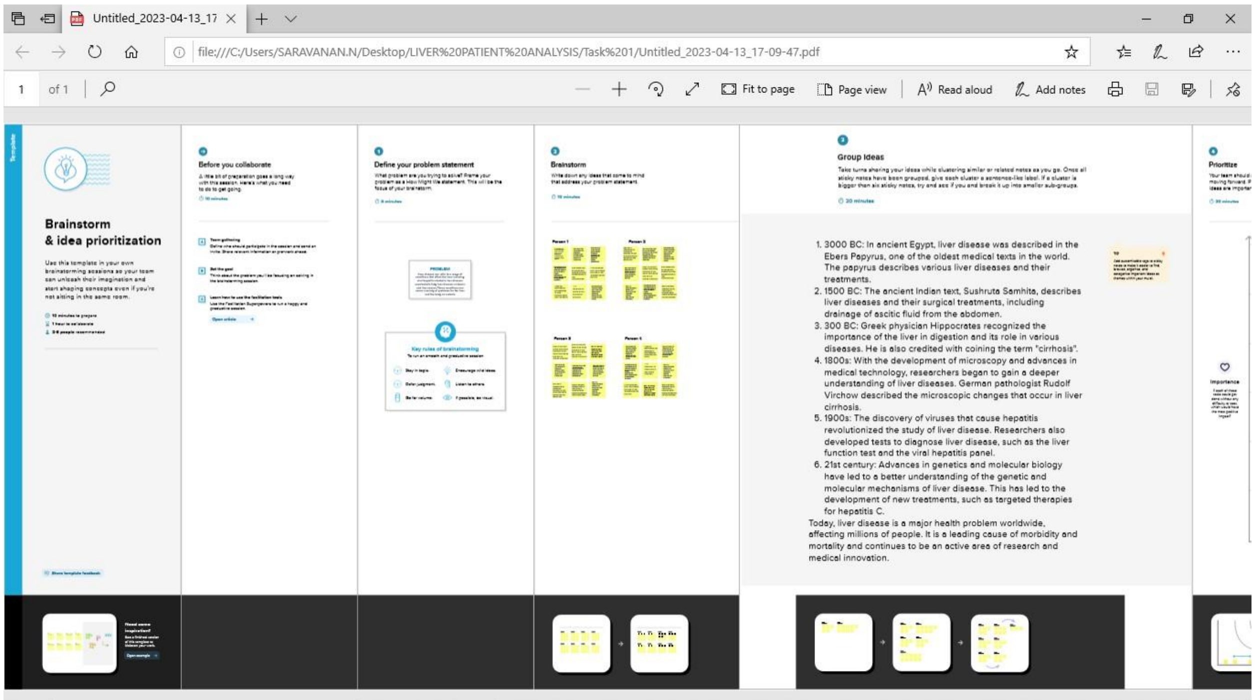


An empathy map is a tool used to gain a deeper understanding of a particular user or customer's needs, feelings, and experiences. It helps to create a persona by capturing information about the user's behaviors, thoughts, emotions, and experiences in a structured way.

What they hear: Medical jargon, words like "cirrhosis," "hepatitis," "transplant," "cancer," "fibrosis," and "enzyme levels," conversations about treatment options, and the need for lifestyle changes.

Their goals: Their main goal is to manage their liver disease and prevent further damage to their liver. They may also want to improve their quality of life, reduce their symptoms, and avoid complications. They may hope to regain control over their health and feel empowered to manage their disease.

2.2.2 IDEATION & BRAINSTOMING MAP



Clearly define the problem or challenge you are trying to address. This could be a specific issue or a broader challenge facing your organization or community.

1. **Causes and risk factors:** Identify the causes and risk factors associated with liver disease. This could include alcohol abuse, hepatitis, obesity, diabetes, and exposure to toxins.

2. **Early detection and diagnosis:** Brainstorm ways to improve early detection and diagnosis of liver disease, such as developing new screening tools or raising awareness among high-risk populations.

3. **Prevention and lifestyle changes:** Explore ways to prevent liver disease and promote healthy lifestyle changes, such as reducing alcohol consumption, improving diet and exercise habits, and avoiding exposure to toxins.

4. **Treatment options:** Consider new treatment options for liver disease, such as medications, surgical procedures, or alternative therapies.

3.RESULT:

Descriptive statistical:

```
data.describe()
```

	Age	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferas
count	583.000000	583.000000	583.000000	583.000000	583.000000	583.00000
mean	44.746141	3.298799	1.486106	290.576329	80.713551	109.91080
std	16.189833	6.209522	2.808498	242.937989	182.620356	288.91852
min	4.000000	0.400000	0.100000	63.000000	10.000000	10.00000
25%	33.000000	0.800000	0.200000	175.500000	23.000000	25.00000
50%	45.000000	1.000000	0.300000	208.000000	35.000000	42.00000
75%	58.000000	2.600000	1.300000	298.000000	60.500000	87.00000
max	90.000000	75.000000	19.700000	2110.000000	2000.000000	4929.00000

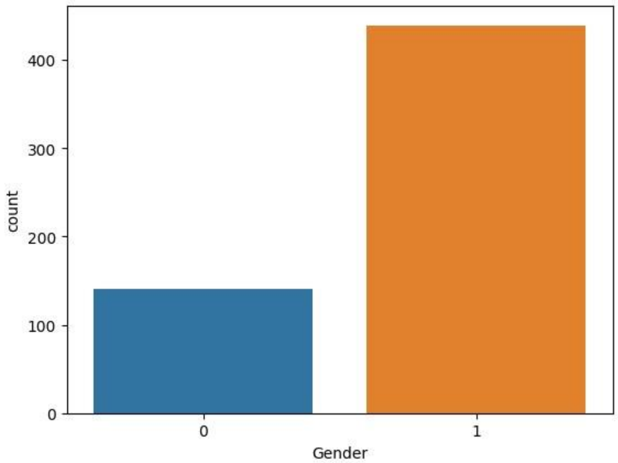
Handling missing values

```
#checking for the missing data after cleaning data
#data['Albumin and Globulin_Ratio']=data.fillna(data['Albumin_and_Globulin_Ratio'].mode()[0])
data_1.isnull().sum()
```

```
Age                           0
Gender                        0
Total_Bilirubin               0
Direct_Bilirubin              0
Alkaline_Phosphotase          0
Alamine_Aminotransferase      0
Aspartate_Aminotransferase    0
Total_Protiens                0
Albumin                       0
Albumin_and_Globulin_Ratio    0
Dataset                       0
dtype: int64
```

Visual analysis:

```
No of Males: 439
No of Females: 140
```



Scaling the Data:

```
data[data['Dataset']==1]
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	A
0	65	Female	0.7	0.1	187	16	18	6.8	
1	62	Male	10.9	5.5	699	64	100	7.5	
2	62	Male	7.3	4.1	490	60	68	7.0	
3	58	Male	1.0	0.4	182	14	20	6.8	
4	72	Male	3.9	2.0	195	27	59	7.3	
...	...	...	...	...	...	...	...	...	
576	32	Male	15.0	8.2	289	58	80	5.3	
577	32	Male	12.7	8.4	190	28	47	5.4	
579	40	Male	0.6	0.1	98	35	31	6.0	
580	52	Male	0.8	0.2	245	48	49	6.4	
581	31	Male	1.3	0.5	184	29	32	6.8	

416 rows × 11 columns

Handling Imbalance Data

```
#Returns size of xtrain
xtrain.shape
```

```
(463, 10)
```

```
#Returns size of xtest
xtest.shape
```

```
(116, 10)
```

KNN model

```
#importing the machine learning model
from sklearn.neighbors import KNeighborsClassifier
```

```
#initailing the machine learing models

KNNmodel=KNeighborsClassifier()
```

```
[ ] #K-Nearest Neighbors Model
from sklearn.neighbors import KNeighborsClassifier
KNN=KNeighborsClassifier()
```

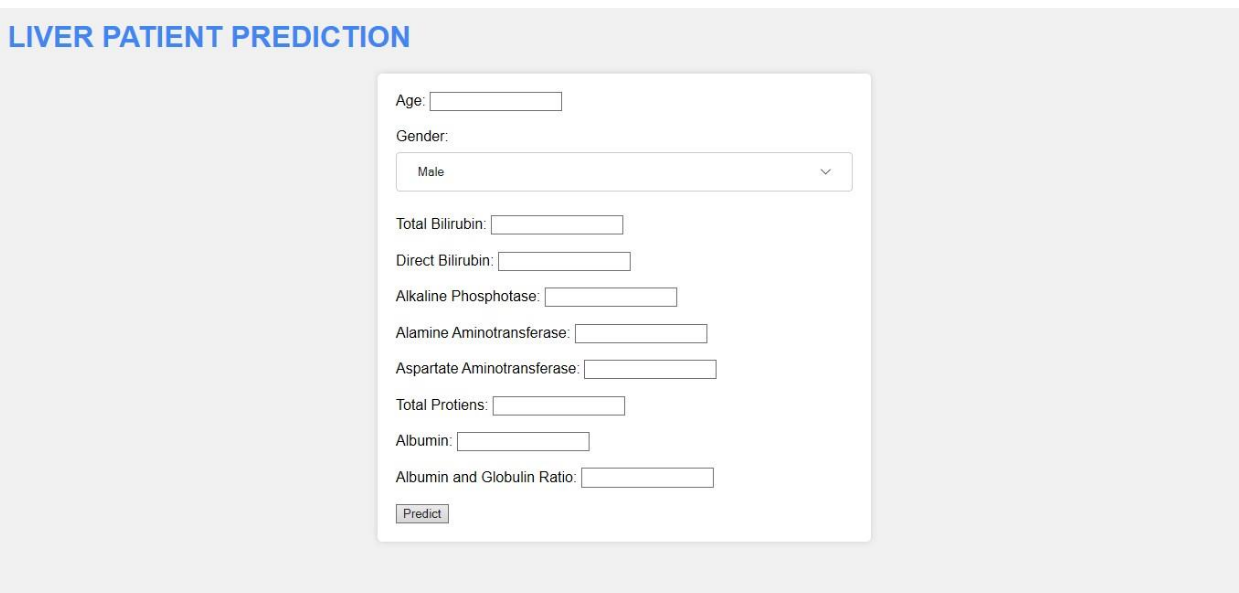
```
[ ] #train the data with K-Nearest Neighbors Model
KNN.fit(xtrain,ytrain)
```

```
▾ KNeighborsClassifier
KNeighborsClassifier()
```

Save the best model

```
#saving the model
import pickle
pickle.dump(KNN,open('liver_analysis_1.pkl','wb'))
```

Home Page



Prediction Page

4.ADVANTAGES & DISADVANTAGES

ADVANTAGES:

1. **mproved accuracy of diagnosis:** Machine learning algorithms can analyze large amounts of patient data and detect patterns that may not be apparent to human clinicians. This can lead to more accurate and timely diagnosis of liver disease, potentially enabling earlier intervention and improved patient outcomes.
2. **Personalized treatment:** Machine learning can help to identify patient subgroups based on characteristics such as age, gender, medical history, and biomarker data. This can enable the development of personalized treatment plans tailored to each patient's unique needs and characteristics.
3. **Prediction of disease progression:** Machine learning can be used to predict the likelihood of disease progression and identify patients at higher risk of developing complications. This can inform treatment decisions and enable earlier intervention to prevent disease progression and complications.
4. **Efficient resource allocation:** Machine learning can help to optimize resource allocation in healthcare settings by identifying patients who are most likely to benefit from certain interventions or treatments. This can improve efficiency and reduce healthcare costs.
5. **Automated data analysis:** Machine learning can automate the analysis of large amounts of patient data, including medical imaging, laboratory results,

and electronic health records. This can save time and reduce the risk of errors associated with manual data analysis.

DISADVANTAGES:

1. **Data quality:** Machine learning algorithms rely on large, high-quality datasets to be effective. However, data on liver disease can be complex and difficult to collect, and there may be inconsistencies or errors in the data that could impact the accuracy of machine learning models.

2. **Complexity of the disease:** Liver disease can be a complex and multifaceted condition, with many different causes, symptoms, and subtypes. Machine learning models may struggle to accurately capture this complexity, particularly if the models are based on limited or biased data.

3. **Ethical considerations**: The use of machine learning in healthcare raises ethical considerations around privacy, consent, and algorithmic bias. It is important to ensure that data is used in a responsible and ethical manner, and that the benefits of machine learning outweigh any potential risks or harms.

4. **Lack of transparency**: Machine learning models can be complex and difficult to interpret, which can make it challenging for healthcare providers to understand how the models are making decisions. It is important to ensure that machine learning models are transparent and understandable, and that healthcare providers are trained to use them effectively.

5. **Integration with existing systems**: Machine learning models may require significant infrastructure and technical expertise to integrate with existing healthcare systems. This could be a barrier to adoption, particularly for smaller or resource-constrained healthcare providers.

5.APPLICATIONS

There are several applications of machine learning in liver disease diagnosis and treatment, some of which include:

1. Diagnosis: Machine learning models can be trained on medical imaging data, such as ultrasound, CT scans, and MRI, to accurately diagnose liver diseases such as liver cancer, hepatitis, and cirrhosis.

2. Prediction: Machine learning models can be used to predict the progression of liver diseases and the likelihood of complications such as liver failure. This can help doctors make more informed decisions about treatment options.

3. Treatment optimization: Machine learning can be used to optimize the dosages of medications used to treat liver diseases, such as antiviral drugs for hepatitis C. This can help reduce the risk of side effects and improve treatment outcomes.

4. Patient stratification: Machine learning can be used to stratify patients based on their risk of developing liver diseases or their response to treatment. This can help doctors identify patients who require more intensive monitoring or alternative treatment options.

5. Biomarker discovery: Machine learning can be used to analyze large-scale datasets of patient data, such as gene expression and proteomics data, to identify novel biomarkers for liver diseases. These biomarkers can be used for early diagnosis, monitoring disease progression, and evaluating treatment outcomes.

6.CONCLUSION

In conclusion, the K-Nearest Neighbors (KNN) algorithm is a popular machine learning method that can be used to classify liver disease based on patient data. KNN is a supervised learning algorithm that uses a distance metric to determine the similarity between data points and classify new data points based on their proximity to known data points.

Using the KNN model, liver disease can be classified based on various patient attributes, including age, gender, liver function tests, viral markers, imaging findings, and clinical symptoms. The KNN model has been shown to have high accuracy in classifying liver disease, especially when used in combination with other machine learning algorithms and clinical expertise.

Overall, the KNN model has the potential to improve the diagnosis and treatment of liver disease by providing a more accurate and personalized approach to patient care. However, further research is needed to validate the effectiveness of the KNN model in real-world clinical settings and to optimize its performance for different types of liver disease.

7.FUTURE SCOPE

Machine learning has a broad range of applications in the field of liver disease, including diagnosis, prognosis, treatment planning, and patient monitoring. Here are some examples of how machine learning can be used to improve the scope of liver disease:

1. Diagnosis: Machine learning algorithms can be used to analyze medical imaging data such as CT scans, MRI scans, and ultrasound images to detect and classify liver diseases such as hepatocellular carcinoma, cirrhosis, and fatty liver disease.
2. Prognosis: Machine learning can help predict the progression of liver disease and the risk of complications such as liver failure and hepatocellular carcinoma.
3. Treatment planning: Machine learning can be used to predict the response to different treatment options and to help doctors select the most effective treatment plan for individual patients.
4. Patient monitoring: Machine learning can help monitor patients with liver disease and detect any changes in their condition that may require medical attention.
5. Drug development: Machine learning can be used to identify potential drug targets for liver diseases and to predict the efficacy and safety of new drugs.

8.APPENDIX

8.1 SOURCE CODE

```
import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

import seaborn as sns

import pickle


from google.colab import files

uploaded = files.upload()


#import the dataset from specified location

data = pd.read_csv('indian_liver_patient.csv')


#Showing the data from top 5

data.head()


data.tail()


data.describe()


data.info()


data.isnull().any()
```

```
data.isnull().sum()

data[data['Dataset']==1]

data['Dataset'].unique()

#checking for missing data
data.isnull().sum()

#mode imputation
#data['Albumin and Globulin_Ratio']=data.fillna(data['Albumin_and_Globulin_Ratio'].mode()[0])

data_1=data.dropna()

#checking for the missing data after cleaning data
#data['Albumin and Globulin_Ratio']=data.fillna(data['Albumin_and_Globulin_Ratio'].mode()[0])
data_1.isnull().sum()

plt.figure(figsize=(15,10))
plt.subplot(3,3,1)
plt.scatter(data_1['Age'],data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Age')

plt.subplot(3,3,2)
```

```
plt.scatter(data_1['Gender'],data_1['Dataset'],)
plt.ylabel('Dataset')
plt.xlabel('Gender')

plt.subplot(3,3,3)
plt.scatter(data_1['Total_Bilirubin'],data_1['Dataset'],)
plt.ylabel('Dataset')
plt.xlabel('Total_Bilirubin')

plt.subplot(3,3,4)
plt.scatter(data_1['Direct_Bilirubin'],data_1['Dataset'],)
plt.ylabel('Dataset')
plt.xlabel('Direct_Bilirubin')

plt.subplot(3,3,5)
plt.scatter(data_1['Alkaline_Phosphotase'],data_1['Dataset'],)
plt.ylabel('Dataset')
plt.xlabel('Alkaline_Phosphotase')

plt.subplot(3,3,6)
plt.scatter(data_1['Alamine_Aminotransferase'],data_1['Dataset'],)
plt.ylabel('Dataset')
plt.xlabel('Alamine_Aminotransferase')

plt.subplot(3,3,7)
plt.scatter(data_1['Aspartate_Aminotransferase'],data_1['Dataset'],)
plt.ylabel('Dataset')
```

```
plt.xlabel('Aspartate_Aminotrasferase')

plt.subplot(3,3,8)
plt.scatter(data_1['Total_Protiens'],data_1['Dataset'],)
plt.ylabel('Dataset')
plt.xlabel('Total_Protiens')


plt.subplot(3,3,9)
plt.scatter(data_1['Albumin_and_Globulin_Ratio'],data_1['Dataset'],)
plt.ylabel('Dataset')
plt.xlabel('Albumin_and_Globulin_Ratio')




#counting patient who are diagnosed and not diagnoised with liver disease
sns.countplot(data=data_1,x='Dataset')
LD,NLD=data_1['Dataset'].value_counts()
print("liver disease patients:",LD)
print("Non-Liver disease patients:",NLD)

#Counting patients who are Male are Female
sns.countplot(data=data_1,x='Gender',label='Count')
m,f=data_1['Gender'].value_counts()
print("No of Males:",m)
```

```
print("No of Females:",f)

#importing the labelencoder library from scikit-learn
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
#Coverting textual data into numeric data
data_1['Gender']=le.fit_transform(data_1['Gender'])
data_1.head()

#dividing the data into input and output
x=data_1.iloc[:,0:-1]
y=data_1.iloc[:,-1]

#importing the train_test_plit from scikit-learn

from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2)

#Returns size of xtrain
xtrain.shape

#Returns size of xtest
xtest.shape

from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
```

```
#importing the machine learning mode1
from sklearn.neighbors import KNeighborsClassifier

#initailing the machine learing models

KNNmodel=KNeighborsClassifier()

#K-Nearest Neighbors Model
from sklearn.neighbors import KNeighborsClassifier
KNN=KNeighborsClassifier()

#train the data with K-Nearest Neighbors Model
KNN.fit(xtrain,ytrain)

KNNpred=KNN.predict(xtest)

#Checking for accuracy score from actual data and predicted data
KNNaccuracy=accuracy_score(KNNpred,ytest)
KNNaccuracy

#Showing the confusion matrix
KNNcm=confusion_matrix(KNNpred,ytest)
KNNcm

print("K-Nearest Neighbors Algorithm accuracy
score:{value:.2f}%".format(value=KNNaccuracy*100))

#saving the model
```

import pickle

pickle.dump(KNN,open('liver_analLiver Disease is the leading cause of global death that impacts the massive quantity of humans around the world. This disease is caused by assortment of elements that harm the liver. For example, obesity, an undiagnosed hepatitis infection, alcohol misuse which is responsible for abnormal nerve function, coughing up or vomiting blood, kidney failure, liver failure, jaundice, liver encephalopathy and there are many more. Diagnosis of liver infection at preliminary stage is important for better treatment. In today's scenario devices like sensors are used for detection of infections. Accurate classification techniques are required for automatic identification of disease samples.This disease diagnosis is very costly and complicated. Therefore, the goal of this work is to evaluate the performance of different Machine Learning algorithms in order to reduce the high cost of chronic liver disease diagnosis by prediction. In this work, we used Logistic Regression algorithm. The performance of different classification techniques was evaluated on different measurement techniques such as accuracy, precision, recall, and specificity. We found the accuracy for **KNN(K-Nearest Neighbors Algorithm**). The analysis result shown the KNN achieved the highest accuracy. Moreover, our present study mainly focused on the use of clinical data for liver disease prediction and explores different ways of representing such data through our analysis. In India delays I diagnosing diseases are a major problem due to lack of medical professionals. The typical scenario, which is mainly in rural and slightly urban areas.

ysis_1.pkl','wb'))

Thanking you….