LUNGS CANCER PREDICTION

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

by

A.SAI JASHWANTH	**2103A52001**
A. VIVEK VARDHAN	**2103A52071**
K. SRINATH	**2103A52084**

Under the guidance of

Mr. D. Ramesh

Assistant Professor, Department of CSE.

SR UNIVERSITY

Department of Computer Science and Artificial Intelligence

Department of Computer Science and Artificial Intelligence

CERTIFICATE

This is to certify that project entitled **"LUNGS CANCER PREDICTION**" is the bonafied work carried out by **K. SRINATH, A. VIVEK VARDHAN, A. SAI JASHWANTH** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** during the academic year 2022-2023 under our guidance and Supervision.

Mr. D Ramesh	**Dr. M.Sheshikala**
Asst. Professor,	Assoc. Prof.& HOD(CSE),
S R University,	S R University,
Ananthasagar, Warangal.	Ananthasagar, Warangal.

ACKNOWLEDGEMENT

We express our thanks to Course co-coordinator **Mr. D. Ramesh, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude to Head of the department CS&AI, **Dr. M.Sheshikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

ABSTRACT

Lung cancer is a serious and potentially life-threatening disease that can be difficult to detect in its early stages. Advances in medical technology have made it possible to use various imaging techniques, such as X-rays, computed tomography (CT) scans, and magnetic resonance imaging (MRI), to detect lung cancer at an earlier stage.

However, accurate detection and diagnosis of lung cancer still present significant challenges, and ongoing research is focused on improving these methods. Overall, detecting lung cancer early is crucial for successful treatment and management of the disease, and our project is focused on developing more accurate and effective methods for early detection and diagnosis

Table of Contents

CHAPTER-1

INTRODUCTION

1.1 Overview

Lung cancer prediction involves the use of various techniques to identify individuals who are at high risk of developing lung cancer. The goal of these techniques is to detect lung cancer early, when it is most treatable. Using our model we can predict at early stage

1.2 Problem Statement

The problem statement for lung cancer prediction using AI and ML is to develop accurate and reliable methods for identifying individuals who are at high risk of developing lung cancer. Early detection of lung cancer is critical for improving patient outcomes, and AI and ML algorithms can help improve the accuracy and efficiency of lung cancer screening and prediction.

1.3 Existing system

One such system is the translational lung cancer research System (TLCRS), developed by researchers at the University of California, Los Angeles. TLCRS uses data on nodule classification and lung cancer prediction from CT imaging data, Risk factor analysis.

1.4 Proposed system

A proposed system for predicting the occurrence of lungs cancer could be developed using machine learning algorithms and data analytics. The system would rely on historical data on lungs cancer outbreaks, Risk factor analysis, Imaging analysis, Genetic analysis.

1.5 Objectives

Early detection is One of the primary objectives of lung cancer prediction is to detect the disease at an early stage when it is more treatable and has a better chance of being cured. Another objective of lung cancer prediction is to assess an individual's risk of developing the disease based on various factors such as age, smoking history, family history, and other risk factors. Lung cancer prediction can also play a role in prevention by identifying individuals at high risk and implementing measures to reduce their risk of developing the disease, such as smoking cessation programs and regular lung cancer screenings

1.6 ARCHITECTURE

In this at first data set to be read after that the null values in the data set to be replaced by 0's. The data must be per-processed and cleaned. The data must be splitted into training and testing. A suitable model must be selected to evaluate it and to predict values and predicted values must be compared to original values to find accuracy of the model

CHAPTER-2

LITERATURE SURVEY

2.1 DOCUMENT THE SURVEY DONE BY YOU:

Machine learning is helpful for a variety of situations. The prediction of dependent variable values from independent variables is one of the uses of this methodology. The field of LUNGS CANCER prediction is a data mining field since it contains extensive data resources that are challenging to manually handle. We will have data and with the help of that data inputs that we have we predict the LUNGS CANCER . Machine learning method scan none the less be helpful in solving this challenge and in anticipating danger early. Some of the techniques used for such prediction problems are multiple linear regression algorithms. The existing research have less inputs. We predict the LUNGS CANCER by taking few more configurations and we used different methods to improve the accuracy in the prediction of lungs cancer but they are not much accurate than our proposed system with the comparison accuracy rates when comparative to existing system to the proposed system

s.no	Year	Mean Square Error
1.	2013	0.652756677329679
2.	2012	0.21017827976428802
3.	2011	0.56134344842135901

Lung Cancer Detection Using Artificial Neural Network

Artificial Neural Network (ANN) for detect the absence or presence of lung cancer in human body. Symptoms were used to diagnose the lung cancer, these symptoms such as Yellow fingers, Anxiety, Chronic Disease, Fatigue, Allergy, Wheezing, Coughing, Shortness of Breath, Swallowing Difficulty and Chest pain. They were used and other information about the person as input variables for our ANN. Our ANN established, trained, and validated using data set, which its title is “survey lung cancer”. Model evaluation showed that the ANN model is able to detect the absence or presence of lung cancer with 96.67 % accuracy

Lungs cancer detection using digital image processing techniques:

CV (Computer Vision) play vital role to prevent lung cancer. Since image processing is necessary for computer vision, further in medical image processing there are many technical steps which are necessary to improve the performance of medical diagnostic machines. If preprocessing and segmentation process have some ambiguity than ultimately it effects on classification process.

Lung Cancer Detection and Classification with 3D Convolutional Neural Network:

A computer-aided diagnosis (CAD) system for lung cancer classification of CT scans with unmarked nodules, a dataset from the Kaggle Data Science. Thresholding was used as an initial segmentation approach to segment out lung tissue from the rest of the CT scan. Thresholding produced the next best lung segmentation. The initial approach was to directly feed the segmented CT scans into 3D CNNs for classification, but this proved to be inadequate. Instead, a modified U-Net trained on LUNA16 data (CT scans with labeled nodules) was used to first detect nodule candidates in the Kaggle CT scans.

Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images :

Lung Cancer is one of the leading life taking cancer worldwide. Early detection and treatment are crucial for patient recovery. Medical professionals use histopathological images of biopsied tissue from potentially infected areas of lungs for diagnosis. Most of the time, the diagnosis regarding the types of lung cancer are error-prone and time-consuming. Convolutional Neural networks can identify and classify lung cancer types with greater accuracy in a shorter period, which is crucial for determining patients' right treatment procedure and their survival rate. Benign tissue, Adenocarcinoma, and squamous cell carcinoma are considered in this research work. The CNN model training and validation accuracy of 96.11 and 97.2 percentage are obtained

CHAPTER-3
DATA PRE-PROCESSING

3.1 DATA PRE-PROCESSING

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data and while doing any operation with data. It is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task. In this particular section we re-label and convert some categorical features into numeric values. This is crucial for training machine learning models since machine learning models accepts the numeric values

3.2 DATASET DESCRIPTION

Different parameters like

1. Patient Id
2. Age
3. Gender
4. Air Pollution
5. Alcohol use

6 Dust Allergy
7 OccuPational Hazards
8 Genetic Risk
9 chronic Lung Disease
10 Balanced Diet
11 Obesity
12 Smoking
13 Passive Smoker
14 Chest Pain
15 Coughing of Blood
16 Fatigue
17 Weight Loss
18 Shortness of Breath
19 Wheezing
20 Swallowing Difficulty
21 Clubbing of Finger Nails
22 Frequent Cold
23 Dry Cough
24 Snoring
25 Level

3.3 DATA CLEANING

index	Patient Id	Age	Gender	Air Polluti	Alcohol us	Dust Aller	OccuPatio	Genetic Ri	chronic Lu	Balanced I	Obesity	Smoking	Passive Sn	Chest Pain	Coughing	Fatigue	Weight Lo	Shortness	W
0	P1	33	1	2	4	5	4	3	2	2	4	3	2	2	4	3	4	2	
1	P10	17	1	3	1	5	3	4	2	2	2	2	4	2	3	1	3	7	
2	P100	35	1	4	5	6	5	5	4	6	7	2	3	4	8	8	7	9	
3	P1000	37	1	7	7	7	7	6	7	7	7	7	7	7	8	4	2	3	
4	P101	46	1	6	8	7	7	7	6	7	7	8	7	7	9	3	2	4	
5	P102	35	1	4	5	6	5	5	4	6	7	2	3	4	8	8	7	9	
6	P103	52	2	2	4	5	4	3	2	2	4	3	2	2	4	3	4	2	
7	P104	28	2	3	1	4	3	2	3	4	3	1	4	3	1	3	2	2	
8	P105	35	2	4	5	6	5	6	5	5	5	6	6	6	5	1	4	3	
9	P106	46	1	2	3	4	2	4	3	3	3	2	3	4	4	1	2	4	
10	P107	44	1	6	7	7	7	7	6	7	7	7	8	7	7	5	3	2	
11	P108	64	2	6	8	7	7	7	6	7	7	7	8	7	7	9	6	5	
12	P109	39	2	4	5	6	6	5	4	6	6	6	6	6	6	5	3	2	
13	P11	34	1	6	7	7	7	6	7	7	7	7	7	7	8	4	2	3	
14	P110	27	2	3	1	4	2	3	2	3	3	2	2	4	2	2	2	3	
15	P111	73	1	5	6	6	5	6	5	6	5	8	5	5	5	4	3	6	
16	P112	17	1	3	1	5	3	4	2	2	2	2	4	2	3	1	3	7	
17	P113	34	1	6	7	7	7	6	7	7	7	7	7	7	8	4	2	3	
18	P114	36	1	6	7	7	7	7	7	6	7	7	7	7	7	8	5	7	

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabelled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time.

Steps involved in data cleaning are:

1. Removal of unwanted substances
2. Fixing structural errors
3. Managing unwanted outliers
4. Handling missing data

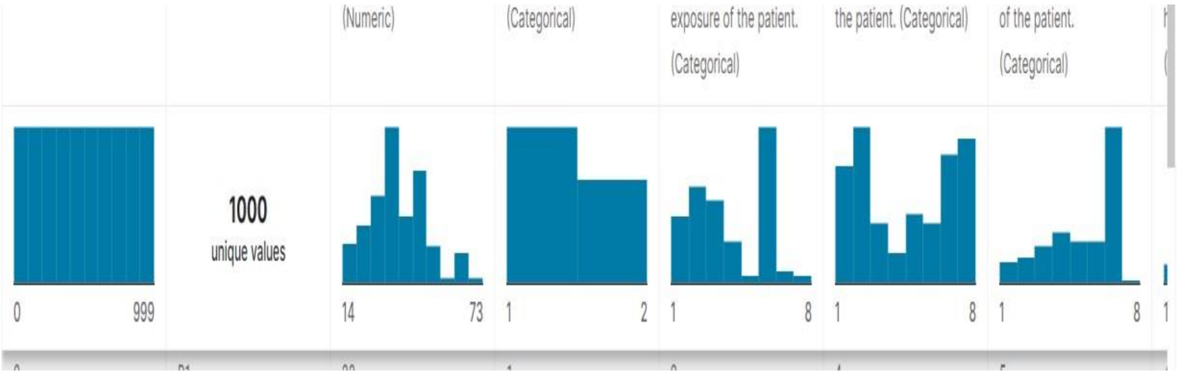
3.4 DATA AUGUMENTATION

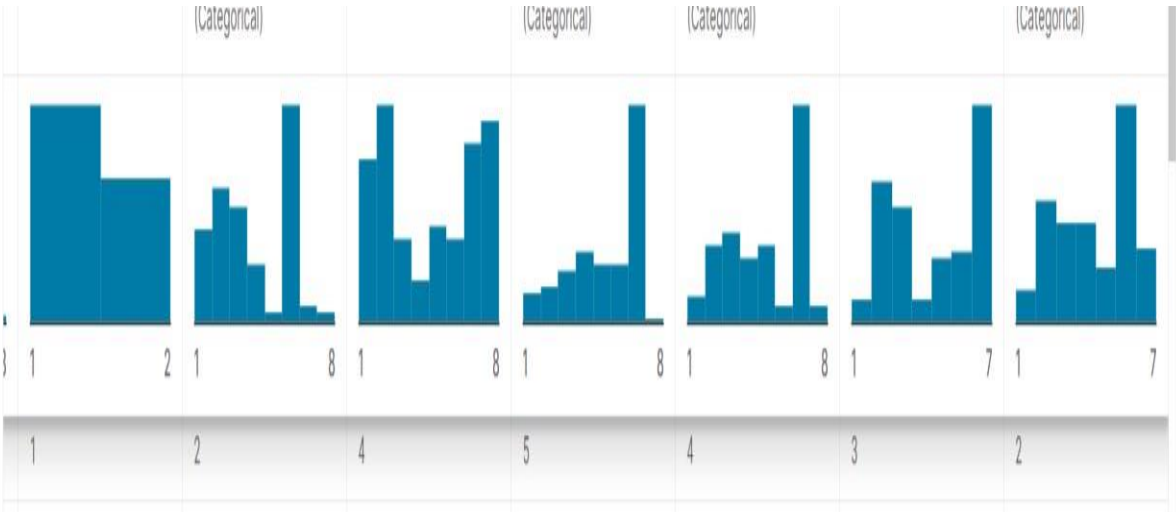
Data augmentation is a set of techniques to artificially increase the amount of data by generating new data points from existing data. This includes making small changes to data or using deep learning models to generate new data points. Machine learning applications especially in the deep learning domain continue to diversify and increase rapidly. Data-centric approaches to model development such as data augmentation techniques can be a good tool against challenges which the artificial intelligence world faces.Data augmentation is useful to improve performance and outcomes of machine learning models by forming new and different examples to train datasets. If the dataset in a machine learning model is rich and sufficient, the model performs better and more accurately.For machine learning models, collecting and labeling of data can be exhausting and costly processes. Transformations in datasets by using data augmentation techniques allow companies to reduce these operational costs.

In our data set the independent variables like area type consist categorical values like super area, carpet area, built area. To make this simple to model we have made the categorical values super area, carpet area, built area as dependent variables and made categorical values for them as 0 and 1. If it belongs to super area then value will be of 1 or else it is 0

3.5 Data Visualization

Data visualization is the graphical representation of information and data in a pictorial or graphical format(Example: charts, graphs, and maps). Data visualization tools provide accessible way to see and understand trends, patterns in data, and outliers. Data visualization tools and technologies are essential to analyzing massive amounts of information and making data- driven decisions. The concept of using pictures is to understand data that has been used for centuries. General types of data visualization are Charts, Tables, Graphs, Maps, Dashboards.


(Numeric)
(Categorical)
exposure of the patient. (Categorical)
the patient. (Categorical)
of the patient. (Categorical)
1000
unique values
0
999
14
73
1
2
1
8
1
8
1
8


(Categorical)
(Categorical)
(Categorical)
(Categorical)
1
2
1
8
1
8
1
8
1
8
1
7
1
7
1
2
4
5
4
3
2

CHAPTER 4

METHODOLOGY

4.1 Procedure to solve the given problem

In our project, the lungs cancer Prediction dataset is imported from Kaggle in Comm SeparatedValues (csv) format. The dataset is analyzed with the help of pandas, numpy and scikit-learn. Tableau is used as a data visualization tool. After drawing insights from the dataset with the help of Tableau, we identify the important factors i.e. factors majorly affecting the change in result.The factors adding insignificant values to the overall result are omitted. The dataset is divided into two parts - training set and testing set. The various machine learning models are trained with the help of the training set. The testing set is then used to check the performance of all the machine learning models. Accuracy score is calculated

4.1.1 LOGISTIC REGRESSION

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

4.1.2 K-NEAREST NEIGHBORS:

K-Nearest Neighbors (KNN) is a non-parametric algorithm that works by finding the k closest points in the training dataset to a new data point and using their labels to predict the label of the new point. The value of k is usually chosen by cross-validation or other techniques. The main advantage of KNN is its simplicity and ease of implementation. It is a lazy learning algorithm, which means that it does not require any training, and the model is built directly on the training data. Additionally, KNN is versatile and can be used for both classification and regression tasks.

4.1.3 NAIVE BAYES:

Naive Bayes is a probabilistic supervised machine learning algorithm used for classification tasks. It is based on the Bayes theorem, which is a fundamental theorem in probability theory. Naive Bayes is called "naive" because it makes the assumption that the features in the input data are independent of each other.This assumption allows the algorithm to simplify the probability calculations and make them computationally efficient.

4.1.4 SUPPORT VECTOR MACHINE:

Support Vector Machine (SVM) is a popular supervised machine learning algorithm used for classification and regression tasks. SVM is particularly useful for classification tasks where the data points are not linearly separable in the input space. SVM can be used for both binary and multiclass classification tasks. In the case of binary classification, SVM tries to find a hyperplane that separates the two classes.while in the case of multiclass classification, SVM constructs multiple hyperplanes to separate the classes.

4.1.5 DECISION TREE:

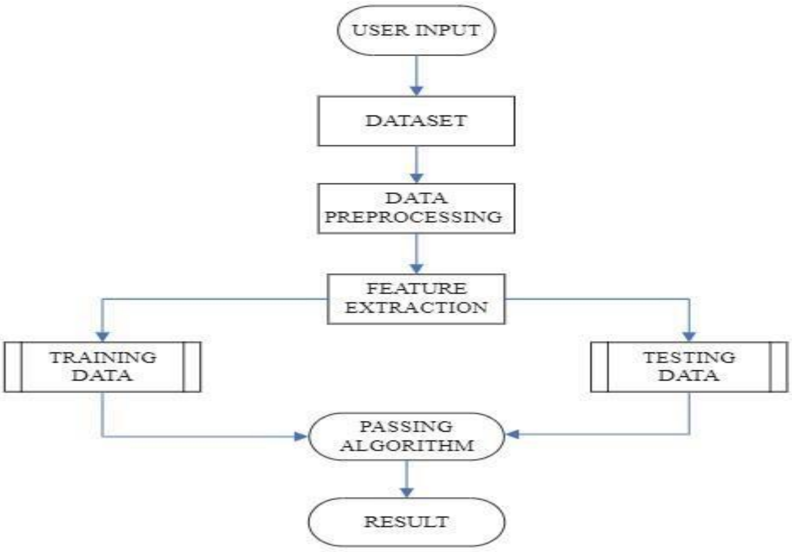
A decision tree is a type of supervised learning algorithm used in machine learning. It is used for both classification and regression tasks. A decision tree is a graphical representation of all the possible outcomes of a decision based on certain conditions. It consists of nodes and branches. One of the key advantages of decision trees is that they can handle missing values and outliers in the data, and can also be used to identify and remove irrelevant features. However, decision trees can be prone to overfitting, especially when dealing with noisy or complex data. To mitigate this problem, techniques such as pruning and ensemble methods like Random Forest can be used.

4.2 Model architecture

Machine Learning is a field of Artificial Intelligence that enables PC frameworks to learn and improve in execution with the assistance of information. It is used to study the construction of algorithms that make predictions on data. Machine learning is used to perform a lot of computing tasks. It is also used to make predictions with the use of computers. Machine learning is sometimes also used to devise complex models. The principle point of machine learning is to permit the PCs to learn things naturally without the assistance of people. Machine learning is very useful and is widely used around the whole world. The process of machine learning involves providing data and then training the computers by building machine learning models with the help of various algorithms. Machine learning can be used to make various applications such as face detection applications, etc. Machine Learning is a field in software engineering that has changed the way of examining information colossally.

A. System Architecture

1. In this system we take data as text input from the user and then we pre-process data of the user
2. Next we extract the required information from the data and then it is sent for classification.
3. In classification data is classified using train data set available in the system and using various algorithm price is predicted.



4.3 Software description

GOOGLE COLAB – Colaboratory or "Colab" for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education

```
import pandas as pd
import numpy as np
data=pd.read_csv('lung_cancer_examples
(1).csv')
print(data)
data.head
data.info
```

CHECKING NULL VALUES

```
data.isnull.sum
```

SCATTER PLOT:

```
import matplotlib.pyplot
as plt
```

```
x1=data['Age']
y=data['Result']
plt.scatter(x1,y)
x2=data['Alkhol']
y=data['Result']
plt.scatter(x2,y)
```

BOX PLOT USING SEABORN:

```
 import pandas as pd import seaborn as
sns sns.boxplot(data=datas)
sns.boxplot(x='Result',y='AreaQ',data=
datas)
sns.boxplot(x='Result',y='Smokes',data
=datas)
sns.boxplot(x='Result',y='Alkhol',data=
datas)
sns.boxplot(x='Result',y='Age',data=dat
as) datas.describe()
```

CORRELATIONAL MATRIX

```
corr_matrix=data.corr()
print(corr_matrix) import
matplotlib.pyplot as plt import
seaborn as sns
sns.heatmap(corr_matrix,annot=
True) plt.show()
```

COVARIENCE MATRIX HEATMAP

```
cov_matrix=pd.DataFrame.cov(data)
sns.heatmap(cov_matrix,annot=True,fmt='g')
plt.show
```

LOGISTIC REGRESSION

import numpy as np import pandas as pd

from sklearn.model_selection import

train_test_split from sklearn.linear_model

import LogisticRegression from

sklearn.metrics import accuracy_score

x=data.drop(columns='Result',axis=1)

y=data['Result'] print(x) print(y)

x_train,x_test,y_train,y_test=train_test_split(

x,y,test_size=0.2,stratify=y,random_state=2)

model=LogisticRegression()

model.fit(x_train,y_train)

x_train_prediction=model.predict(x_train)

training_data_accuracy=accuracy_score(x_tr

ain_prediction,y_train) print("THE

ACCURACY SCORE OF OUR MODEL

IS:",training_data_accuracy)

x_test_prediction=model.predict(x_test)

test_data_accuracy=accuracy_score(x_test_p

rediction,y_test) print("THE TESTING

ACCURACY OF OUR MODEL

IS:",test_data_accuracy) from

sklearn.metrics import confusion_matrix cm

= confusion_matrix(ytest, ypred) print

("Confusion Matrix ", cm)

DEVELOPMENT PHASE OF LOGISTIC REGRESSION:

```
#buliding prediction
#convert data to numpyarray
#reshape because aiml thinks that it will predict all of the values thats why it changes to (1,-1)
input_data=(35,3,5,4)
data_as_numpy=np.asarray(input_data)
data_reshaped=data_as_numpy.reshape(1,-1)
pred=model.predict(data_reshaped)
print(pred)
```

```
[1]
```

KNN ALGORITHM

import numpy as np from sklearn.neighbors import KNeighborsClassifier from sklearn.metrics import accuracy_score from sklearn.model_selection import train_test_split x=data.drop(columns='Result',axis=1)
y=data['Result']
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,stratify=y,random_state=2) model=KNeighborsClassifier(n_neighbors=10)
model.fit(x_train,y_train) x_train_prediction=model.predict(x_train)
training_data_accuracy=accuracy_score(x_train_prediction,y_train)
print("THE ACCURACY SCORE OF OUR MODEL IS:",training_data_accuracy) model=KNeighborsClassifier(n_neighbors=3)
model.fit(x_train,y_train) x_train_prediction=model.predict(x_train)
training_data_accuracy=accuracy_score(x_train_prediction,y_train)
print("THE ACCURACY SCORE OF OUR MODEL IS:",training_data_accuracy) x_test_prediction=model.predict(x_test)
test_data_accuracy=accuracy_score(x_test_prediction,y_test)
print("THE TESTING ACCURACY OF OUR MODEL IS:",test_data_accuracy)

DEVELOPMENT PHASE OF KNN:

```
[ ] model.predict([[35,3,5,4]])

/usr/local/lib/python3.9/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but KNeighborsClassifier was fitted with feature names
  warnings.warn(
array([0])
```

Naive Bayes Classifier

```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB() nb.fit(xtrain,ytrain)
print("Training
Accuracy",nb.score(xtrain,ytrain))
print("Testing Accuracy",nb.score(xtest,ytest))
print("Overall
Accuracy:",nb.score(sst.transform(x),y))
y_pred=nb.predict(xtest) y_pred from
sklearn.metrics import confusion_matrix
nbb=confusion_matrix(ytest,y_pred)
print("Confusion matrix:\n",nbb) DECISION
TREE:
from sklearn.tree import
DecisionTreeClassifier
dtc=DecisionTreeClassifier()
dtc.fit(xtrain,ytrain)
print("Training
Accuracy",dtc.score(xtrain,ytrain))
print("Testing Accuracy",dtc.score(xtest,ytest))
print("Overall
Accuracy:",dtc.score(sst.transform(x),y))
y_pred=dtc.predict(xtest) y_pred from
sklearn.metrics import confusion_matrix
dtcs=confusion_matrix(ytest,y_pred)
print("Confusion matrix:\n",dtcs)  from sklearn
import tree plt.figure(figsize=(15,10))
tree.plot_tree(dtc,filled=True
```

SUPPORT VECTOR MACHINE

```
from sklearn import svm
SVM= svm.SVC()
SVM.fit(xtrain, ytrain) print("Training
Accuracy",SVM.score(xtrain,ytrain))
print("Testing Accuracy",SVM.score(xtest,ytest))
print("Overall
Accuracy:",SVM.score(sc_x.transform(x),y))
y_pred=SVM.predict(xtest) y_pred
from sklearn.metrics import confusion_matrix
SVMS=confusion_matrix(ytest,y_pred)
print("Confusion matrix:\n",SVMS)
```

ACCURACYSCORE:

```
training_data_accuracy=accuracy_score(x_train_prediction,y_train)
print("THE ACCURACY SCORE OF OUR MODEL
IS:",training_data_accuracy)
model=KNeighborsClassifier(n_neighbors=3)
model.fit(x_train,y_train)  x_train_prediction=model.predict(x_train)
training_data_accuracy=accuracy_score(x_train_prediction,y_train)
print("THE ACCURACY SCORE OF OUR MODEL
IS:",training_data_accuracy) x_test_prediction=model.predict(x_test)
test_data_accuracy=accuracy_score(x_test_prediction,y_test)
print("THE TESTING ACCURACY OF OUR MODEL
IS:",test_data_accuracy) print("Training
Accuracy",SVM.score(xtrain,ytrain)) print("Testing
Accuracy",SVM.score(xtest,ytest)) print("Overall
Accuracy:",SVM.score(sc_x.transform(x),y))
y_pred=SVM.predict(xtest) y_pred
```

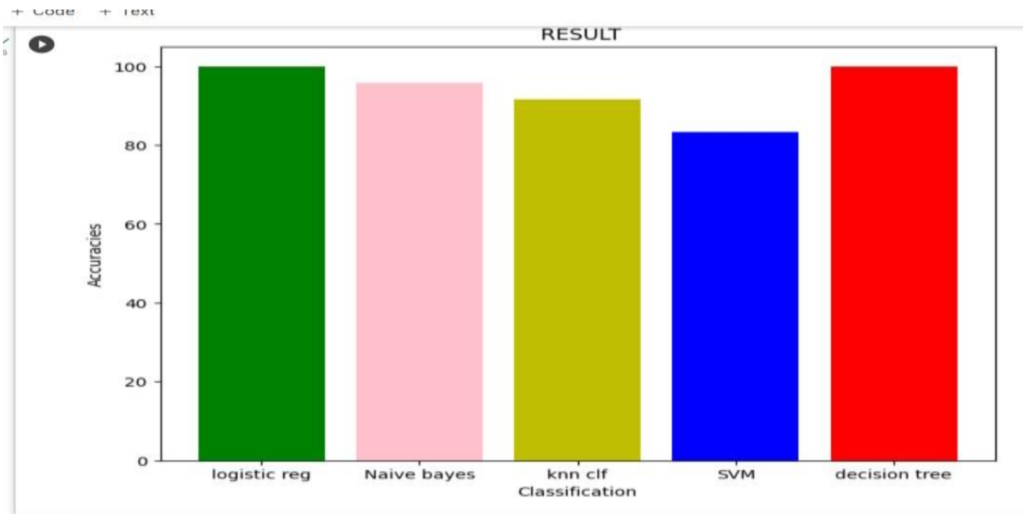
```
from sklearn.metrics import
confusion_matrix
SVMS=confusion_matrix(ytest,y_pred)
print("Confusion matrix:\n",SVMS)
print("Training
Accuracy",dtc.score(xtrain,ytrain))
print("Testing
Accuracy",dtc.score(xtest,ytest))
```

RESULTS

```
import matplotlib.pyplot as plt fig = plt.figure() ax = fig.add_axes([0,0,1,1])
algo = ['logistic reg','Naive bayes','knn clf','SVM','decision tree']
accuracy=[accuracy_model*100,accuracynb*100,accuracy_knn*100,accuracy_
SVM*100,ac
curacy_dtc*100]
ax.bar(algo[0],accuracy[0],color = 'b')
ax.bar(algo[1],accuracy[1],color = 'y')
ax.bar(algo[2],accuracy[2],color = 'pink')
ax.bar(algo[3],accuracy[3],color = 'green')
ax.bar(algo[4],accuracy[4],color = 'r')
plt.xlabel('Classifiers------------>')
plt.ylabel('Accuracies------------->')
plt.title('ACCURACIES      RESULTED')
plt.show()
```

CHAPTER-5
RESULTS AND DISCUSSION

S.NO	MODEL	ACCURACY
1.	Lasso regression	1.0
2.	Navie bayes	0.95666666666
3.	KNN	0.91566666666
4.	SVM	0.833333333334
5.	Decision Tree	1.0



CHAPTER 6

CONCLUSION AND FUTURE SCOPE:

In conclusion, lung cancer prediction is an important area of research and healthcare that aims to detect and diagnose lung cancer at an early stage, assess an individual's risk of developing the disease, and provide personalized treatment plans to improve survival rates and reduce mortality rates. Lung cancer prediction can also play a role in prevention by identifying individuals at high risk and implementing measures to reduce their risk of developing the disease. The objectives of lung cancer prediction include early detection, risk assessment, personalized treatment, improving survival rates, and prevention. Developing and implementing accurate and reliable lung cancer prediction models and algorithms can have a significant impact on reducing the burden of lung cancer and improving the overall health outcomes for individuals at risk of or living with the disease.

The future scope of lung cancer prediction is promising and includes several advancements that are likely to improve the accuracy, efficiency, and effectiveness of lung cancer prediction. The use of AI and ML algorithms can help to improve the accuracy of lung cancer prediction models by analyzing large datasets and identifying patterns that may not be apparent to humans.Development of non-invasive biomarkers is Research is underway to identify non-invasive biomarkers that can be used to detect lung cancer in its early stages, such as circulating tumor cells, DNA, RNA, and proteins.Incorporation of imaging technology is Advances in imaging technology, such as computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI), can help to improve the accuracy of lung cancer diagnosis and prediction by providing detailed images of the lungs and surrounding tissues.

CHAPTER 7

REFERENCES:

Rebecca Siegel, Deepa Naishadham and Ahmedin Jemal, "Cancer statistics 2013", *CA: a* cancer journal for clinicians, vol. 63, no. 1, pp. 11-30, 2013

"Reduced lung-cancer mortality with low-dose computed tomographic screening", *New England Journal of Medicine*, vol. 365, no. 5, pp. 395-409, 2011.

J St Thomas, D Lamb, T Ashcroft, B Corrin, CW Edwards, AR Gibbs, et al., "How reliable is the diagnosis of lung cancer using small biopsy specimens? report of a ukcccr lung cancer working party", *Thorax*, vol. 48, no. 11, pp. 1135-1139, 1993

Dinggang Shen, Guorong Wu and Heung-Il Suk, "Deep learning in medical image analysis", *Annual review of biomedical engineering*, vol. 19, pp. 221-248, 2017.

Zhang Li, Zheyu Hu, Jiaolong Xu, Tao Tan, Hui Chen, Zhi Duan, Ping Liu, Jun Tang, Guoping Cai, Quchang Ouyang et al., "Computer-aided diagnosis of lung carcinoma using deep learning-a pilot study", *arXiv preprint arXiv:1803.05471*, 2018.