ADVANCED ACADEMIC CENTER

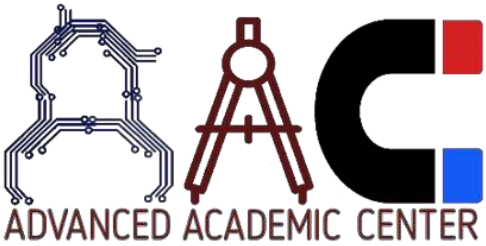
A CENTER FOR INTER-DISCIPLINARY RESEARCH
2020-21

PREDICTION OF BREAST CANCER

SUPERVISED BY

ABHINAV CHAKILAM

GOKARAJU RANGARAJU
INSTITUTE OF ENGINEERING AND TECHNOLOGY
AUTONOMOUS



Advanced Academic Center

(A Center for Inter-Disciplinary Research)

PREDICTION OF BREAST CANCER

is a bonafide work carried out by the following students in partial fulfillment of the requirements for Advanced Academic Centre intern, submitted to the chair, AAC during the academic year 2020-21.

NAME	ROLL NO.	BRANCH
ABHINAV PENDELA	20241A04S4	ECE
ADARI NIKHIL DATTA SAI	20241A05U2	CSE
MEDISHETTY SUMANA	20241A12F1	IT
MENGJI DYUTI	20241A0538	CSE
S. ASHWIN KUMAR	20241A0350	MECH

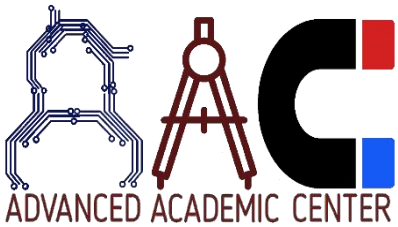
This work was not submitted or published earlier for any study

Dr/Ms./Mr.

Project Supervisor

Dr.B.R.K. Reddy
Program Coordinator

Dr.Ramamurthy Suri
Associate Dean, AAC



ACKNOWLEDGEMENTS

We express our deep sense of gratitude to our respected Director, Gokaraju Rangaraju Institute of Engineering and Technology, for the valuable guidance and for permitting us to carry out this project.

With immense pleasure, we extend our appreciation to our respected Principal, for permitting us to carry out this project.

We are thankful to the Associate Dean, Advanced Academic Centre, for providing us an appropriate environment required for the project completion.

We are grateful to our project supervisor who spared valuable time to influence us with their novel insights.

We are indebted to all the above-mentioned people without whom we would not have concluded the project.

CONTENTS PAGE NO

1. ABSTRACT

Cancer has been characterized as a heterogeneous disease consisting of many different sub-types. The early diagnosis and prognosis of a cancer type have become a necessity in cancer research, as it can facilitate the subsequent clinical management of patients. The importance of classifying cancer patients into high or low risk groups has led many research teams, from the biomedical and the bioinformatics field, to study the application of machine learning (ML) methods. Therefore, these techniques have been utilized as an aim to model the progression and treatment of cancerous conditions. In addition, the ability of ML tools to detect key features from complex datasets reveals their importance. A variety of these techniques, including Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Support Vector Machines (SVMs) and Decision Trees (DTs) have been widely applied in cancer research for the development of predictive models, resulting in effective and accurate decision making.

Even though it is evident that the use of ML methods can improve our understanding of cancer progression, an appropriate level of validation is needed in order for these methods to be considered in the everyday clinical practice. In this work, we present a review of recent ML approaches employed in the modeling of cancer progression. The predictive models discussed here are based on various supervised ML techniques as well as on different input features and data samples. Given the growing trend on the application of ML methods in cancer research, we present here the most recent publications that employ these techniques as an aim to model cancer risk or patient outcomes.

2. INTRODUCTION

2.1 What is Cancer?

Cancer starts when cells in the body change (mutate) and grow out of control. Your body is made up of tiny building blocks called cells. Normal cells grow when your body needs them, and die when your body doesn't need them any longer. Cancer is made up of abnormal cells that grow even though your body doesn't need them. In most types of cancer, the abnormal cells grow to form a lump or mass called a tumor.

2.2 What is Breast Cancer?

Breast cancer is cancer that starts in cells in the breast. The ducts and the lobules are the 2 parts of the breast where cancer is most likely to start. Breast cancer is one of the most common types of cancer in women in the U.S. Once breast cancer forms, cancer cells can spread to other parts of the body (metastasize), making it life-threatening. The good news is that breast cancer is often found early, when it's small and before it has spread.

2.3 Machine Learning and Breast Cancer Detection

In this project, the random forest algorithm is used to analyze the medical case diagnosis of breast cancer. The random forest algorithm can combine the characteristics of multiple eigenvalues, and the combined results of multiple decision trees can be used to improve the prediction accuracy. In this project, a random forest algorithm is used to discuss the case of breast cancer case diagnosis and obtain high prediction accuracy. Random forest is one of many classification techniques, and it is an algorithm for big data classification. Random forest classification is applied here to achieve a more accurate and reliable classification performance. The accuracy in this project is 96.50%.

3. MACHINE LEARNING

3.1 Introduction to Machine Learning

Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they carry out certain tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step.

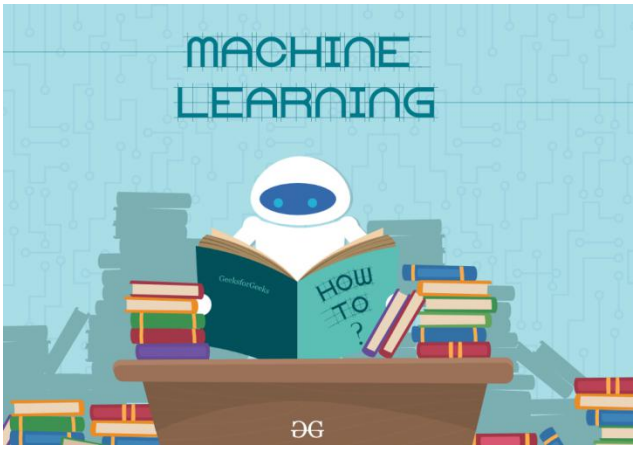


Fig : Machine Learning

3.2 Data in ML

Data can be any unprocessed fact, value, text, sound or picture that is not being interpreted and analyzed. Data is the most important part of all Data Analytic, Machine Learning, Artificial Intelligence. Without data, we can't train any model and all modern research and automation will go vain.

3.2.1 Types of Data in ML:

Training Data: The part of data we use to train our model. This is the data which your model actually sees(both input and output) and learn from.

Validation Data: The part of data which is used to do a frequent evaluation of model, fit on training dataset along with improving involved hyperparameters (initially set parameters before the model begins learning). This data plays it's part when the model is actually training.

Testing Data: Once our model is completely trained, testing data provides the unbiased evaluation. When we feed in the inputs of Testing data, our model will predict some values(without seeing actual output). After prediction, we evaluate our model by comparing it with actual output present in the testing data. This is how we evaluate and see how much our model has learned from the experiences feed in as training data, set at the time of training.

3.3 ML Algorithms

An “algorithm” in machine learning is a procedure that is run on data to create a machine learning “model.” Machine learning algorithms perform “pattern recognition.” Algorithms “learn” from data, or are “fit” on a dataset. There are many machine learning algorithms.

3.3.1 Classification in ML:

Classification is the task of “classifying things” into sub-categories. Classification is of two types:

1. Binary Classification : When we have to categorize given data into 2 distinct classes. Example – On the basis of given health conditions of a person, we have to determine whether the person has a certain disease or not.
2. Multi class Classification : The number of classes is more than 2. For Example – On the basis of data about different species of flowers, we have to determine which specie does our observation belong to.

3.3.2 Types of Classifiers (Algorithms):

There are various types of classifiers. Some of them are :

1. Linear Classifiers : Logistic Regression
2. Tree Based Classifiers : Decision Tree Classifier
3. Support Vector Machines
4. Artificial Neural Networks
5. Bayesian Regression
6. Gaussian Naive Bayes Classifiers
7. Stochastic Gradient Descent (SGD) Classifier
8. Ensemble Methods : Random Forests, AdaBoost, Bagging Classifier, Voting
9. Classifier, ExtraTrees Classifier

3.3.3 Practical Applications of Classifiers:

Here are a few practical applications of classifiers.

- Google's self driving car uses deep learning enabled classification techniques which enables it to detect and classify obstacles.
- Spam E-mail filtering is one of the most widespread and well recognized uses of Classification techniques.
- Detecting Health Problems, Facial Recognition, Speech Recognition, Object Detection, Sentiment Analysis all use Classification at their core.

3.3.4 More about Random Forest:

Every decision tree has high variance, but when we combine all of them together in parallel then the resultant variance is low as each decision tree gets perfectly trained on that particular sample data and hence the output doesn't depend on one decision tree but multiple decision trees. In the case of a classification problem, the final output is taken by using the majority voting classifier. In the case of a regression problem, the final output is the mean of all the outputs. This part is Aggregation.

Step 1 : Import the required libraries.

Step 2 : Import and print the dataset

Step 3 : Select all rows and column 1 from dataset to x and all rows and column 2 as y

Step 4 : Fit Random forest regressor to the dataset

Step 5 : Predicting a new result

Step 6 : Visualising the result

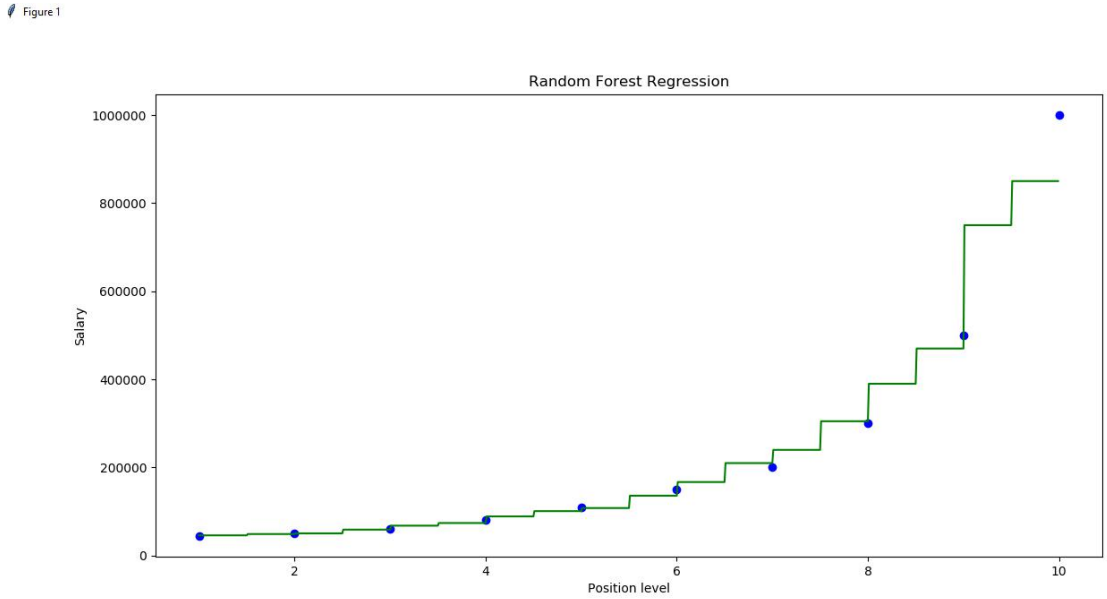
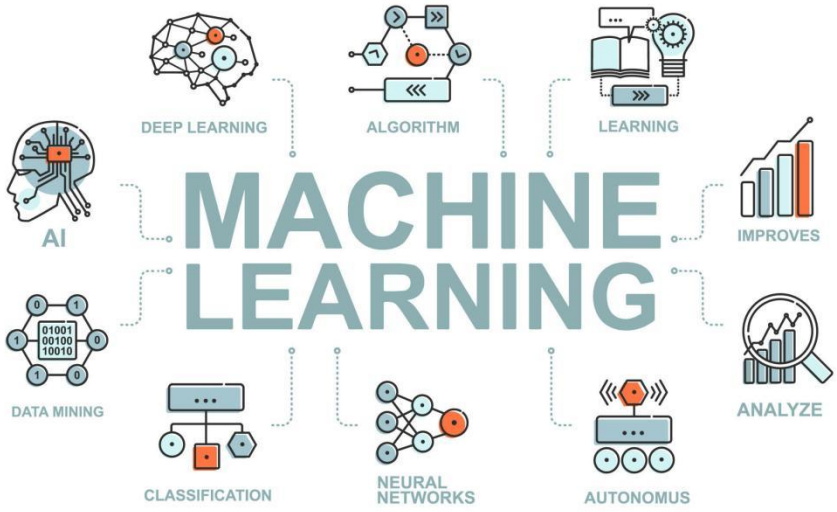


Fig : Random Forest Regression

3.5 History of ML

Tom M. Mitchell provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: **"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."** This definition of the tasks in which machine learning is concerned offers a fundamentally operational definition rather than defining the field in cognitive terms.

Modern day machine learning has two objectives, one is to classify data based on models which have been developed, the other purpose is to make predictions for future outcomes based on these models. A hypothetical algorithm specific to classifying data may use computer vision of moles coupled with supervised learning in order to train it to classify the cancerous moles. Where as, a machine learning algorithm for stock trading may inform the trader of future potential predictions.



3.6 Future of ML:

Machine Learning has been one of the hottest topics of discussion among the C-suite. With its incredible potential to compute and analyze huge amounts of data, advanced ML techniques are being used in businesses to perform complex tasks quicker and more efficiently.

The machine learning market is expected to grow from USD 1.03 Billion in 2016 to USD 8.81 Billion by 2022, at a Compound Annual Growth Rate (CAGR) of 44.1% during the forecast period.

Machine learning-driven solutions are being leveraged by organizations to improve customer experience, ROI, and to gain a competitive edge in business. Big players in the field like Google, IBM, Microsoft, Apple, and Salesforce are already leveraging ML benefits using Machine Learning.

3.7 Applications of ML:

a. Machine Learning in Education

Advances in AI are enabling teachers to realize a far better understanding of how their students are progressing with learning. AI will make big and positive changes in education helping students to enjoy the training process and have a far better understanding with their teachers. Students will not feel apprehensive towards their teachers and be frightened of being judged.

Fig : ML in Education

b. Machine learning in Search Engine

Search engines rely on machine learning to improve their services is no secret today. Implementing these Google has introduced some amazing services. Such as voice recognition, image search and many more. Google services like its image search and translation tools use sophisticated machine learning which permit computers to ascertain , listen and speak in much an equivalent way as human do.

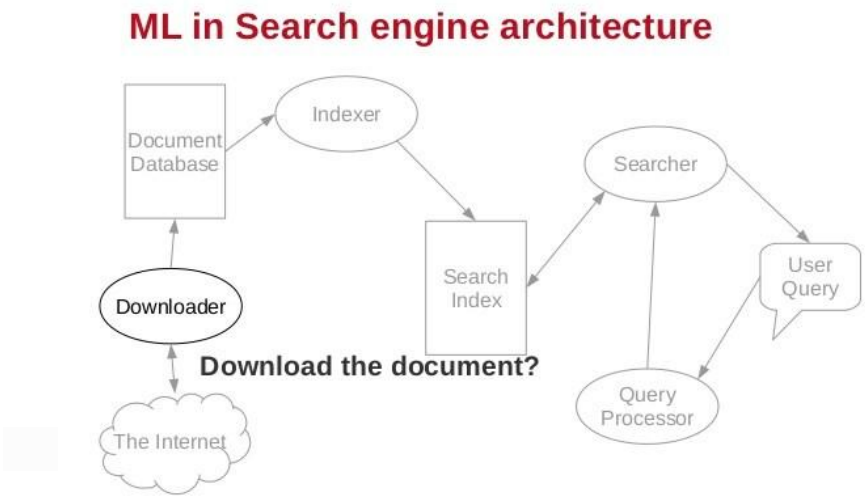


Fig : ML in Search Engine

c. Machine Learning in Digital Marketing

This is where machine learning can help significantly. Machine Learning is being implemented in digital marketing departments round the globe It allows a more relevant personalization. Thus, companies can interact and engage with the customer.

As consumer expectations grow for more personalized, relevant, and assistive experiences, machine learning is becoming a useful tool to assist meet those demands. Sophisticated segmentation focus on the appropriate customer at the right time. Also, with the right message.

Fig : ML in Digital Marketing

d. Machine Learning in Health Care

Machine learning, simply put, may be a sort of AI when computers are programmed to find out information without human intervention.

The foremost common healthcare use cases for machine learning are automating medical billing, clinical decision support and therefore the development of clinical care guidelines. More importantly, scientists and researchers are using machine learning (ML) to churn out variety of smart solutions which will ultimately help in diagnosing and treating an illness.

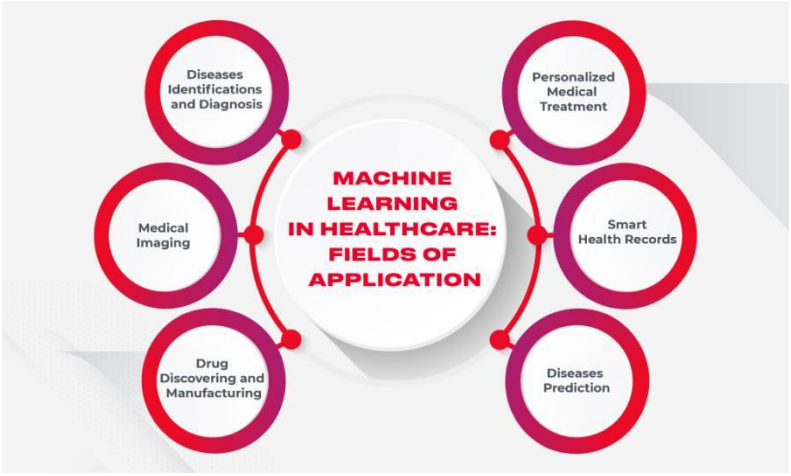


Fig : ML in Health Care

4.PROJECT WORKFLOW

4.1 Introduction

Machine learning has given the computer systems the abilities to automatically learn without being explicitly programmed. But how does a machine learning system work? So, it can be described using the life cycle of machine learning. Machine learning life cycle is a cyclic process to build an efficient machine learning project. The main purpose of the life cycle is to find a solution to the problem or project.

Machine learning life cycle involves seven major steps, which are given below:

1. Gathering Data
2. Data preparation
3. Data Wrangling
4. Analyse Data
5. Train the model
6. Test the model
7. Deployment

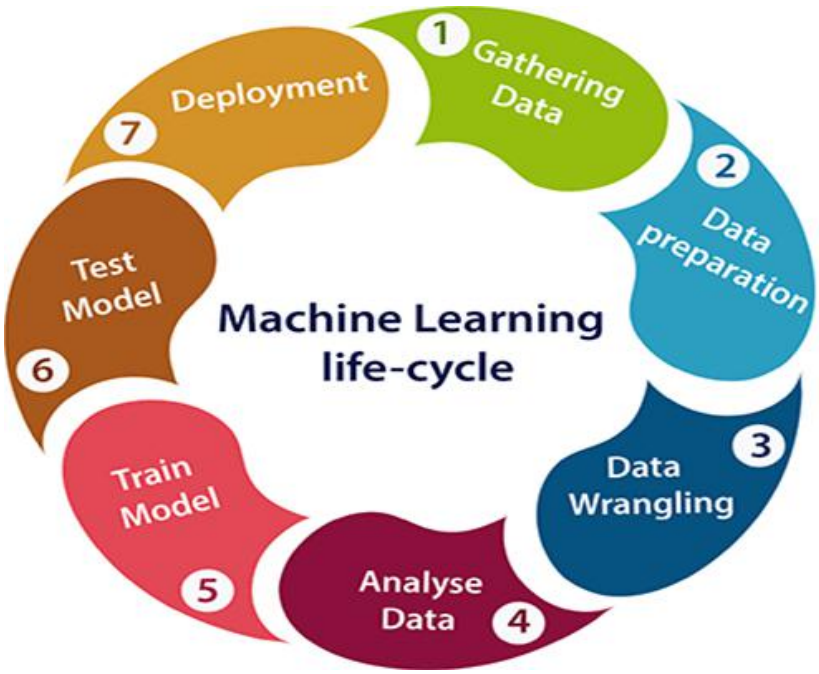


Fig : Steps in Machine Learning

4.2 Gathering Data

Data Gathering is the first step of our project. The goal of this step is to identify and obtain all data-related problems. In this step, we need to identify the different data sources, as data can be collected from various sources such as files, database, internet, or mobile devices. It is one of the most important steps. The quantity and quality of the collected data will determine the efficiency of the output. The more will be the data , the more accurate will be the prediction.

This step includes the below tasks:

- Identify various data sources
- Collect data
- Integrate the data obtained from different sources

By performing the above task, we get a coherent set of data, also called as a dataset.

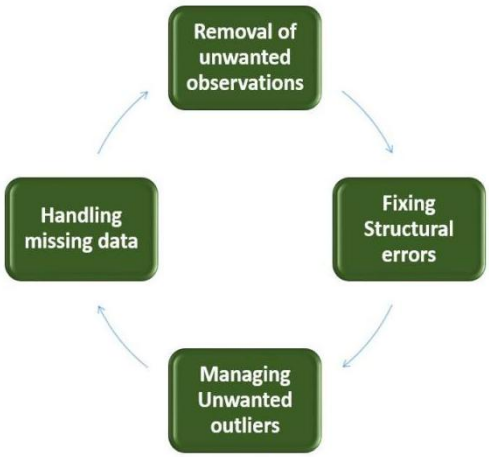
4.3 Data Wrangling

Data wrangling is the process of cleaning and converting raw data into a useable format. It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues. It is not necessary that data we have collected is always of our use as some of the data may not be useful. In real-world applications, collected data may have various issues, including:

- Missing Values
- Duplicate data
- Invalid data
- Noise

So, we use various filtering techniques to clean the data.

It is mandatory to detect and remove the above issues because it can negatively affect the quality of the outcome.

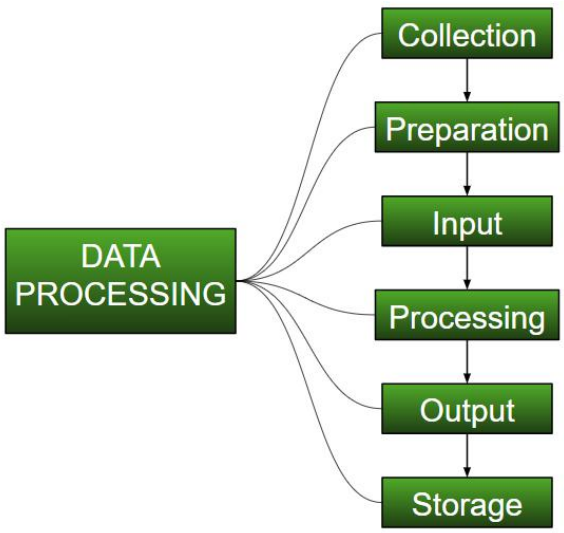


4.4 Data Preparation

After collecting the data, we need to prepare it for further steps. Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training. In this step, first, we put all data together, and then randomize the ordering of data. This step can be further divided into two processes:

Data exploration: It is used to understand the nature of data that we have to work with. We need to understand the characteristics, format, and quality of data. A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.

Data pre-processing: Now the next step is preprocessing of data for its analysis.



4.5 Data Analysis

Now the cleaned and prepared data is passed on to the analysis step. This step involves:

- Selection of analytical techniques
- Building models
- Review the result

The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome. It starts with the determination of the type of the problems, where we select the machine learning techniques such as Classification, Regression, Cluster analysis, Association, etc. then build the model using prepared data, and evaluate the model. Hence, in this step, we take the data and use machine learning algorithms to build the model.

4.6 Train Model

Now the next step is to train the model, in this step we train our model to improve its performance for better outcome of the problem.

We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and, features.

4.7 Test Model

Once our machine learning model has been trained on a given dataset, then we test the model. In this step, we check for the accuracy of our model by providing a test dataset to it. Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

4.8 Deployment

The last step of our project workflow is deployment, where we deploy the model in the real-world system. If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will check whether it is improving its performance using available data or not. The deployment phase is similar to making the final report for a project.

5.CODE

THIS PROGRAM DETECTS BREAST CANCER BASED ON DATA

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#LOAD THE DATA
from google.colab import files
uploaded=files.upload()
df=pd.read_csv('data.csv')
df.head(10)

#COUNTING THE NUMBER OF ROWS AND COLUMNS IN THE DATA SET
df.shape

#COUNTING ALL THE MISSING VALUES
df.isna().sum()
df.dropna(axis=1)
df.shape
df['diagnosis'].value_counts()
sns.countplot(df['diagnosis'],label='count')
df.dtypes

from sklearn.preprocessing import LabelEncoder
labelencoder_Y = LabelEncoder()
df.iloc[:,1]= labelencoder_Y.fit_transform(df.iloc[:,1].values)
print(labelencoder_Y.fit_transform(df.iloc[:,1].values))
 sns.pairplot(df.iloc[:,1:9])
```

```
df.iloc[:,1:12].corr()
plt.figsize=(10,10)

sns.heatmap(df.iloc[:,1:12].corr(),annot=True,fmt='.0%')
X=df.iloc[:,2:31].values
Y=df.iloc[:,1].values
type(X)

from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.25,random_state=0)

from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
def models(X_train,Y_train):

  #USING LOGISTIC REGRESSION
  from sklearn.linear_model import LogisticRegression
  log = LogisticRegression(random_state = 0)
  log.fit(X_train, Y_train)

  #USING KNEIGHBORSCLASSIFIER
  from sklearn.neighbors import KNeighborsClassifier
  knn = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
  knn.fit(X_train, Y_train)

  #USING SVC LINEAR
  from sklearn.svm import SVC
  svc_lin = SVC(kernel = 'linear', random_state = 0)
  svc_lin.fit(X_train, Y_train)
```

```
#USING SVC RBF
from sklearn.svm import SVC
svc_rbf = SVC(kernel = 'rbf', random_state = 0)
svc_rbf.fit(X_train, Y_train)

#USING GAUSSIANNB
from sklearn.naive_bayes import GaussianNB
gauss = GaussianNB()
gauss.fit(X_train, Y_train)

#USING DECISIONTREECLASSIFIER
from sklearn.tree import DecisionTreeClassifier
tree = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
tree.fit(X_train, Y_train)

#USING RANDOMFORESTCLASSIFIER METHOD OF ENSEMBLE CLASS TO USE RANDOM FOREST CLASSIFICATION ALGORITHM
from sklearn.ensemble import RandomForestClassifier
forest = RandomForestClassifier(n_estimators = 10, criterion = 'entropy', random_state = 0)
forest.fit(X_train, Y_train)

#PRINT MODEL ACCURACY ON THE TRAINING DATA.
print('[0]Logistic Regression Training Accuracy:', log.score(X_train, Y_train))
print('[1]K Nearest Neighbor Training Accuracy:', knn.score(X_train, Y_train))
print('[2]Support Vector Machine (Linear Classifier) Training Accuracy:', svc_lin.score(X_train, Y_train))
print('[3]Support Vector Machine (RBF Classifier) Training Accuracy:', svc_rbf.score(X_train, Y_train))
print('[4]Gaussian Naive Bayes Training Accuracy:', gauss.score(X_train, Y_train))
print('[5]Decision Tree Classifier Training Accuracy:', tree.score(X_train,
```

```
Y_train))
 print('[6]Random Forest Classifier Training Accuracy:', forest.score(X_train,
Y_train))
return log, knn, svc_lin, svc_rbf, gauss, tree, forest
model = models(X_train,Y_train)

from sklearn.metrics import confusion_matrix
for i in range(len(model)):
cm = confusion_matrix(Y_test, model[i].predict(X_test))
  TN = cm[0][0]
  TP = cm[1][1]
  FN = cm[1][0]
  FP = cm[0][1]
  print(cm)
  print('Model[{}] Testing Accuracy = "{}!"'.format(i,  (TP + TN) / (TP + TN + FN
  + FP)))
  print()

#WAYS TO GET THE CLASSIFICATION ACCURACY & OTHER METRICS
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score

for i in range(len(model)):
print('Model ',i)
#Check precision, recall, f1-score
print( classification_report(Y_test, model[i].predict(X_test)) )
#Another way to get the models accuracy on the test data
print( accuracy_score(Y_test, model[i].predict(X_test)))
print()
```

```
#PRINT PREDICTION OF RANDOM FOREST CLASSIFIER MODEL
pred = model[6].predict(X_test)
print(pred)
print()

#PRINT THE ACTUAL VALUES
print(Y_test)
```

OUTPUTS:

TRAINING OUTPUT

[0] Logistic Regression Training Accuracy: 0.9906103286384976

[1] K Nearest Neighbor Training Accuracy: 0.9765258215962441

[2] Support Vector Machine (Linear Classifier) Training Accuracy: 0.9882629107981221

[3] Support Vector Machine (RBF Classifier) Training Accuracy: 0.9835680751173709

[4] Gaussian Naive Bayes Training Accuracy: 0.9507042253521126

[5] Decision Tree Classifier Training Accuracy: 1.0

[6] Random Forest Classifier Training Accuracy: 0.9953051643192489

TESTING OUTPUT

Model[0] Testing Accuracy = "0.9440559440559441!"

Model[1] Testing Accuracy = "0.958041958041958!"

Model[2] Testing Accuracy = "0.965034965034965!"

Model[3] Testing Accuracy = "0.965034965034965!"

Model[4] Testing Accuracy = "0.9230769230769231!"

Model[5] Testing Accuracy = "0.951048951048951!"

Model[6] Testing Accuracy = "0.965034965034965!"

6.ML MODEL ANALYSIS :

Model 0

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.96	0.96	0.96	90
1	0.92	0.92	0.92	53
ACCURACY			0.94	143
MACRO AVG.	0.94	0.94	0.94	143
WEIGHTED AVG.	0.94	0.94	0.94	143

ACCURACY: 0.9440559440559441

Model 1

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.95	0.99	0.97	90
1	0.98	0.91	0.94	53
ACCURACY			0.96	143
MACRO AVG.	0.96	0.95	0.95	143
WEIGHTED AVG.	0.96	0.96	0.96	143

ACCURACY: 0.958041958041958

Model 2

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.98	0.97	0.97	90
1	0.94	0.96	0.95	53
ACCURACY			0.97	143
MACRO AVG.	0.96	0.96	0.96	143
WEIGHTED AVG.	0.97	0.97	0.97	143

ACCURACY:0.965034965034965

Model 3

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.97	0.98	0.97	90
1	0.96	0.94	0.95	53
ACCURACY			0.97	143
MACRO AVG.	0.96	0.96	0.96	143
WEIGHTED AVG.	0.96	0.97	0.96	143

ACCURACY: 0.965034965034965

Model 4

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.93	0.94	0.94	90
1	0.90	0.89	0.90	53
ACCURACY			0.92	143
MACRO AVG.	0.92	0.92	0.92	143
WEIGHTED AVG.	0.92	0.92	0.92	143

ACCURACY: 0.9230769230769231

Model 5

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.99	0.93	0.96	90
1	0.90	0.98	0.94	53
ACCURACY			0.95	143
MACRO AVG.	0.94	0.96	0.95	143
WEIGHTED AVG.	0.95	0.95	0.95	143

ACCURACY: 0.951048951048951

Model 6

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.98	0.97	0.97	90
1	0.94	0.96	0.95	53
ACCURACY			0.97	143
MACRO AVG.	0.96	0.96	0.96	143
WEIGHTED AVG.	0.97	0.97	0.97	143

ACCURACY: 0.965034965034965

7.GRAPHICAL ANALYSIS

Graph 1:

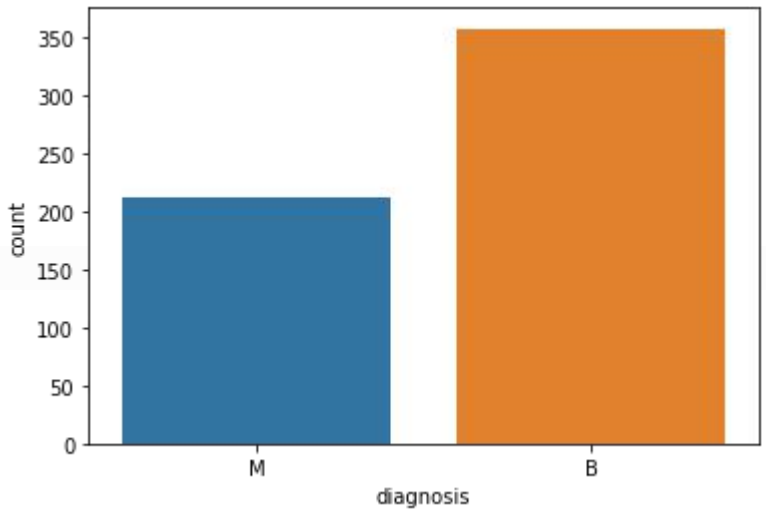


Fig: Prediction of the type of Cancer (M- Malignant B- Benign)

Graph 2:

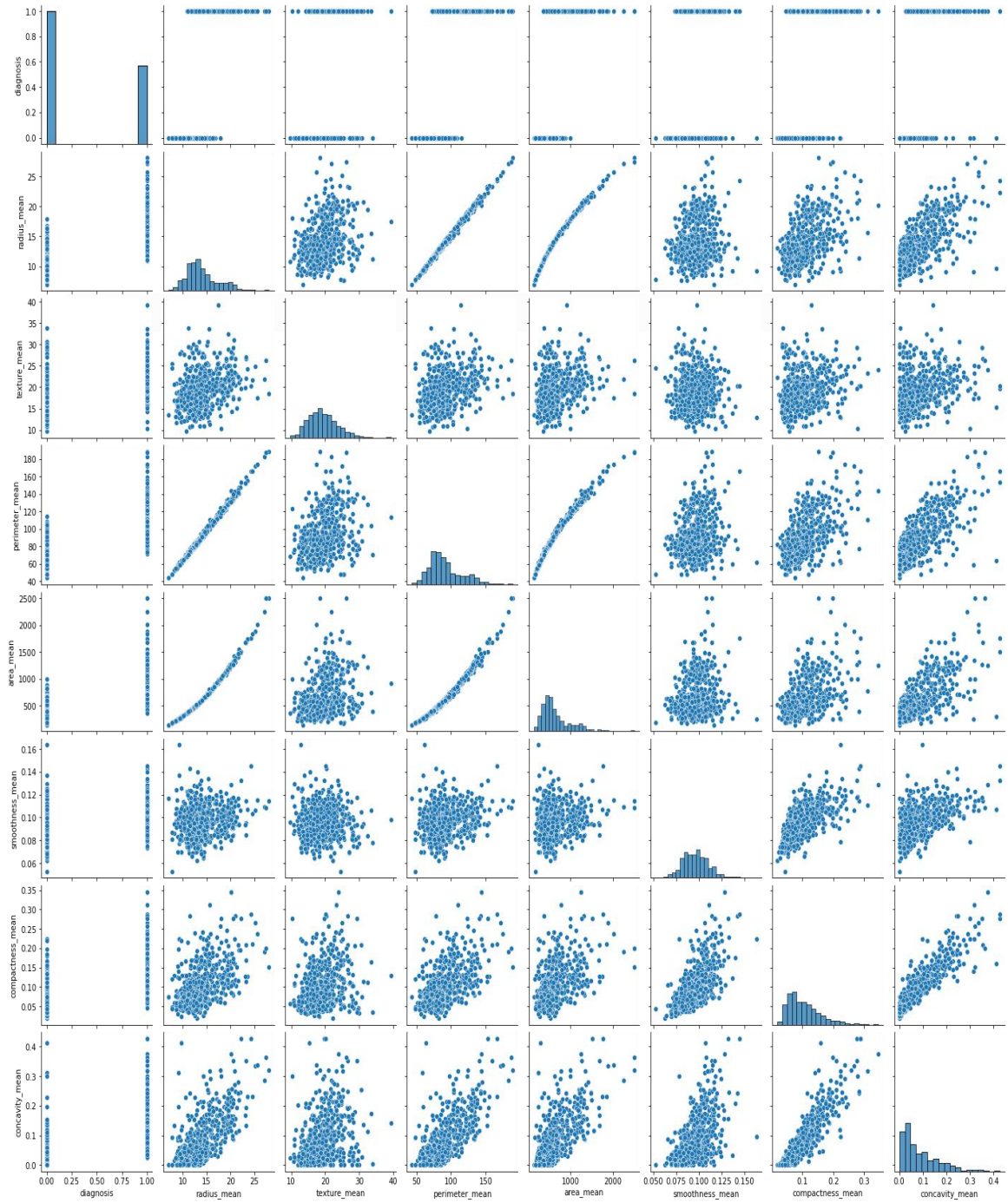


Fig: Plot of Various Parameters (Independent Variables)

Graph 3:

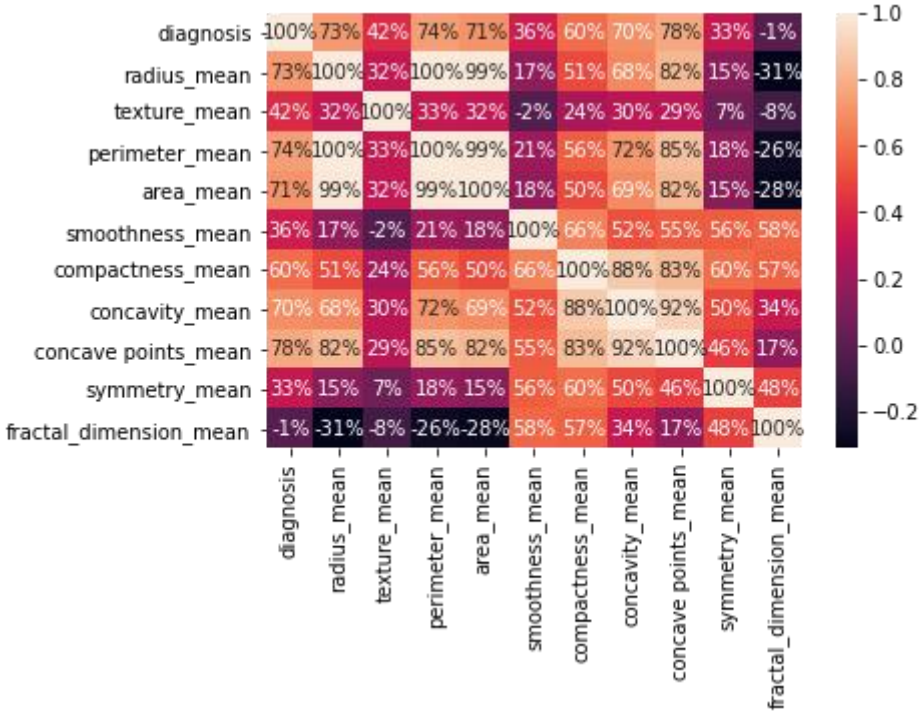


Fig: Dependency of Parameters on each other

8. FUTURE DEVELOPMENTS

The rapid advancement of machine learning and especially deep learning continues to fuel the medical imaging community's interest in applying these techniques to improve the accuracy of cancer screening. These findings show that automatic deep learning methods can be readily trained to attain high accuracy on heterogeneous mammography platforms, and hold tremendous promise for improving and developing clinical tools to reduce false positive and false negative screening mammography results.

In future, clinical tests will be performed, either at a clinic or at home. Data is inputted into a pathological ML system. A few minutes later, an email is received with a detailed report that has an accurate prediction about the development of cancer. We can expect ML to replace local pathologist in the coming decades, and it's pretty exciting!

A potential future development of the presented work is to apply ML models to other data with different features, concerning the survival prognosis of the patients. We also plan to make some substantial improvements of our Python-based workflow and in particular to make it a web-based application with additional services. The future research can be carried out to predict the other different parameters and breast cancer research can be categorizes on basis of other parameters. The analysis of the results signifies that the integration of multidimensional data along with different classification, feature selection and dimensionality reduction techniques can provide auspicious tools for inference in this domain.

Further research in this field should be carried out for the better performance of the classification techniques so that it can predict on more variables. Future developments include parametrizing classification techniques to achieve high accuracy, looking into many datasets and how further Machine Learning algorithms can be used to characterize Breast Cancer, reducing the error rates with maximum accuracy.

9. REFERENCES

[1] B. Akbugday, "Classification of Breast Cancer Data Using Machine Learning Algorithms," 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 2019.

[2] https://www.kaggle.com/

[3] https://www.geeksforgeeks.org/

[4] https://www.nature.com/articles/s41598-019-48995-4#Sec15

[5] https://www.javatpoint.com/machine-learning

[6] https://www.wikipedia.org/

[7] https://www.researchgate.net/publication/341508593_BREAST_CANCER_PREDICTION_USING_MACHINE_LEARNING

[8] https://www.sciencedirect.com/science/article/pii/S2001037014000464

[9] https://data-flair.training/blogs/future-of-machine-learning/

[10] https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94