BREAST CANCER PREDICTION USING MACHINE LEARNING

By

DHINISHA CHRISTY J	**713521AM012**
FAARIZ HUSSAIN S	**713521AM014**
HARISH KUMAR P	**713521AM018**
MANIGANDAN G	**713521AM027**

MINI PROJECT REPORT-II

Submitted to the

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

SNS COLLEGE OF TECHNOLOGY
(AN AUTONOMOUS INSTITUTION)

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DECEMBER 2022

BONAFIDE CERTIFICATE

Certified that this Mini Project Report titled "**BREAST CANCER PREDICTION USING MACHINE LEARNING**" is the Bonafide record of work done by the students "**DHINISHA CHRISTY J (713521AM012), FAARIZ HUSSAIN S (713521AM014), HARISH KUMAR P (713521AM018), MANIGANDAN G (713521AM027)**" had carried out the 19ITP203-Mini project work under my supervision. Certified further that to the best of my knowledge the work reported here is does not from part of any other project report or research work on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROJECT GUIDE

Mr.K.S MOHAN

Assistant Professor,
Department of Information Technology,
SNS College of Technology,
Coimbatore – 641 035.

HEAD OF DEPARTMENT

Dr.M.UDHAYAMOORTHI,

Associate Professor & HOD(i/c)
Department of Information Technology,
SNS College of Technology,
Coimbatore – 641 035.

Submitted for the Mini Project -II examination held on.....

Internal Examiner **External Examiner**

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
1	EMPATHY	7
	1.1 RISK FACTORS	7
	1.2 MAJAOR PROBLEMS	10
2	DEFINE	11
	2.1 CAUSES	11
	2.2 INHERITED BREAST CANCER	11
3	IDEATE	12
	3.1 EXPLORATORY DATA ANALYSIS	12
	3.2 DATA VISUALIZATION	13
	3.3 CORRELATION ANALYSIS	13
	3.4 SOFTWARE AND LIBRARIES	14
4	PROTOTYPE	16
	4.1 PYTHON3 SOFTWARE	16
	4.2 PANDAS	16
	4.3 NUMPY	17
	4.4 SK LEARN	17
	4.5 MATPLOTLIB	18
	4.6 SEABORN	18
	4.7 FLASK	19

	4.8 ARCHITECTURAL DIAGRAM	**20**
	4.9 SUPPORT VECTOR MACHINE ALGORITHM	**20**
	4.10 PREDICTION ACCURACY	**21**
5	**TESTING**	**22**
6	**CONCLUSION**	**28**
	6.1 CONCLUSION	**28**
	6.2 FUTURE WORKS	**28**
	REFERENCE	**29**

LIST OF FIGURES

FIG NO.	TITLE	PAGE. NO
3.1	EXPLORATORY DATA ANALYSIS	12
3.2	DATA VISUALIZATION	13
3.3	CORRELATION ANALYSIS	14
3.4	SOFTWARE AND LIBRARIES	15
4.1	PYTHON3 SOFTWARE	16
4.2	PANDAS LIBRARY	17
4.3	NUMPY LIBRARY	17
4.4	SCIKIT-LEARN LIBRARY	18
4.5	MATPLOTLIB LIBRARY	18
4.6	SEABORN LIBRARY	19
4.7	STREAMLIT LIBRARY	19
4.8	ARCHITECTURAL DIAGRAM	20
4.10	PREDICTION ACCURACY	21

ABSTRACT

Women are seriously threatened by breast cancer with high morbidity and mortality. The lack of robust prognosis models results in difficulty for doctors to prepare a treatment plan that may prolong patient survival time. Hence, the requirement of time is to develop the technique which gives minimum error to increase accuracy.

Four algorithm Classification and Regression Trees (CART), Support Vector Machine (SVM), Gaussian Naïve Bayes (NB) and K-Nearest Neighbors (KNN) which predict the breast cancer by inducing the training data into the ML models. All experiments are executed within a simulation environment and conducted in JUPYTER NOTEBOOK platform. Out of these algorithms, SVM gives better accuracy comparing to the rest. Then the ML model is integrated with the front-end using Python Flask.

The proposed work can be used to predict the outcome based on the users inputs that is the patient is affected by cancer or not. This research is carried out to predict the accuracy. The future research can be carried out to predict the other different parameters and breast cancer research can be categorises on basis of other parameters.

CHAPTER 1
EMPATHY

Breast cancer was the most frequent malignancy in women, occupying 30% of all new cancer cases diagnosed in women from the published data of the American Cancer Society. The incidence of breast cancer in women was estimated to be increasing slowly by 0.5% per year, which contributed to 287,850 new cases of invasive breast cancer diagnoses and 43,250 deaths cases from breast cancer recorded among United States women in 2021. Like the United States, China has a huge number of breast cancer patients; the highest incidence of malignant tumor in women was still breast cancer, with about 304,000 new cases every year, which means the treatment and quality of life of breast cancer survivors are worth our extensive attention.

Cancer patients tended to have more emotional problems in terms of pain, such as fear and sadness in comparison with healthy individuals. Empathy is not only the temporary state of emotional cognition but also a relatively constant personality trait, which plays extremely important positive effects on individuals and in society. Pain empathy is a common component of emotion, which focused on understanding and experiencing others' painful feelings and responding to them emotionally and behaviorally. When doctors see patients suffering from pain, they experience compassion, so pain empathy has often been used to measure the professional ethics of medical staff.

Breast cancer patients would feel inferior and depressed due to the incomplete body after radical mastectomy, meanwhile, chemotherapy could also reduce their concern for surrounding things, resulting in emotional regulation and empathy impairment, which would even persist for several years.

1.1 RISK FACTORS

Studies have shown that your risk for breast cancer is due to a combination of factors. The main factors that influence your risk include being a woman and getting older. Most breast cancers are found in women who are 50 years old or older.

Some women will get breast cancer even without any other risk factors that they know of. Having a risk factor does not mean you will get the disease, and not all risk factors have the same effect. Most women have some risk factors, but most women

do not get breast cancer. If you have breast cancer risk factors, talk with your doctor about ways you can lower your risk and about screening for breast cancer.

Getting older: The risk for breast cancer increases with age. Most breast cancers are diagnosed after age 50.

Reproductive history: Starting menstrual periods before age 12 and starting menopause after age 55 expose women to hormones longer, raising their risk of getting breast cancer.

Personal history of breast cancer or certain non-cancerous breast diseases. Women who have had breast cancer are more likely to get breast cancer a second time. Some non-cancerous breast diseases such as atypical hyperplasia or lobular carcinoma in situ are associated with a higher risk of getting breast cancer.

A woman's risk for breast cancer is higher if she has a mother, sister, or daughter (first-degree relative) or multiple family members on either her mother's or father's side of the family who have had breast or ovarian cancer. Having a first-degree male relative with breast cancer also raises a woman's risk.

Previous treatment using radiation therapy. Women who had radiation therapy to the chest or breasts (for instance, treatment of Hodgkin's lymphoma) before age 30 have a higher risk of getting breast cancer later in life.

Exposure to the drug diethylstilbestrol (DES). DES was given to some pregnant women in the United States between 1940 and 1971 to prevent miscarriage. Women who took DES, or whose mothers took DES while pregnant with them, have a higher risk of getting breast cancer.

Cancer is the second leading cause of death in the United States, but many kinds of cancer can be prevented or caught early. Leading risk factors for preventable cancers are smoking, getting too much ultraviolet (UV) radiation from the sun or tanning beds, being overweight or having obesity, and drinking too much alcohol.

CDC is a leader in efforts to reduce preventable cancers, improve the health of cancer survivors, and help give every person an equal opportunity to achieve the best health possible. Several divisions within CDC's National Centre for Chronic Disease Prevention and Health Promotion work to reduce risk factors for preventable cancers, promote screening to catch cancer early, and collect data on all notifiable cancer cases in the United States.

Smoking and second-hand smoke cause about 80% to 90% of lung cancer deaths in the United States. Smoking also causes cancer of the voice box (larynx), mouth and throat, esophagus, urinary bladder, kidney, pancreas, cervix, colon, rectum, liver, and stomach, as well as a type of blood cancer called acute myeloid leukemia. About 34 million US adults smoke cigarettes, and every day, about 1,600 young people under age 18 try their first cigarette.

People who don't smoke but are exposed to second-hand smoke at home or at work have a 20% to 30% higher risk of lung cancer. Second-hand smoke causes more than 7,300 lung cancer deaths in this population each year. In the United States, 58 million people who don't smoke are exposed to second-hand smoke every year.

CDC's Office on Smoking and Health is at the forefront of the nation's efforts to reduce deaths and prevent chronic diseases that result from commercial* tobacco use, including cancer. OSH prioritizes health equity by creating opportunities for all people to be as healthy as possible.

CDC and its partners promote efforts to:

- Prevent young people from starting to use tobacco.
- Promote quitting among adults and young people.
- Reduce people's exposure to second-hand smoke.
- Advance health equity by identifying and eliminating tobacco-related disparities.

CDC's Tips From Former Smokers® (Tips®) campaign, the first federally funded tobacco education campaign, focuses on motivating US adults who smoke to try to quit. Tips features real people who are living with serious health conditions caused by smoking and second-hand smoke exposure. The newest Tips series adds compelling stories from family members who take care of loved ones affected by a smoking-related disease or disability.

Researchers have identified hormonal, lifestyle and environmental factors that may increase your risk of breast cancer. But it's not clear why some people who have no risk factors develop cancer, yet other people with risk factors never do. It's likely that breast cancer is caused by a complex interaction of your genetic makeup and your environment.

1.2MAJOR PROBLEMS

By empathizing the breast cancer in healthcare, there are some noticeable problems in the desired field

- The doctors and laboratory persons need to check for result after diagnosis which takes long time
- By collecting the symptom results, it is needed to be predicted before itself that the person affected or not

CHAPTER 2
DEFINE

2.1 CAUSES

Doctors know that breast cancer occurs when some breast cells begin to grow abnormally. These cells divide more rapidly than healthy cells do and continue to accumulate, forming a lump or mass. Cells may spread (metastasize) through your breast to your lymph nodes or to other parts of your body.

Breast cancer most often begins with cells in the milk-producing ducts (invasive ductal carcinoma). Breast cancer may also begin in the glandular tissue called lobules (invasive lobular carcinoma) or in other cells or tissue within the breast.

2.2 INHERITED BREAST CANCER

Doctors estimate that about 5 to 10 percent of breast cancers are linked to gene mutations passed through generations of a family. A number of inherited mutated genes that can increase the likelihood of breast cancer have been identified. The most well-known are breast cancer gene 1 (BRCA1) and breast cancer gene 2 (BRCA2), both of which significantly increase the risk of both breast and ovarian cancer. If you have a strong family history of breast cancer or other cancers, your doctor may recommend a blood test to help identify specific mutations in BRCA or other genes that are being passed through your family. Consider asking your doctor for a referral to a genetic counsellor, who can review your family health history. A genetic counsellor can also discuss the benefits, risks and limitations of genetic testing to assist you with shared decision-making.

CHAPTER 3
IDEATE

These problems can be resolved by developing a machine learning model to predict whether the patient has symptoms of breast cancer.

Finally, by selecting an appropriate machine learning algorithm, the prediction is made with better accuracy providing algorithm.

3.1 EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations. It is a good practice to understand the data first and try to gather as many insights from it. EDA is all about making sense of data in hand, before getting them dirty with it.

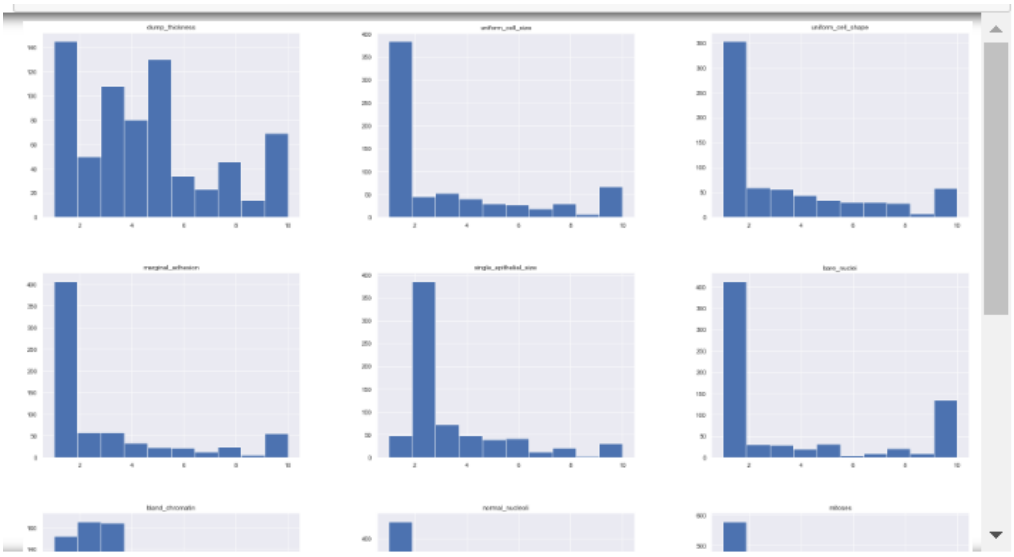
```
df.describe()
```

	id	clump_thickness	uniform_cell_size	uniform_cell_shape	marginal_adhesion	si
count	6.990000e+02	699.000000	699.000000	699.000000	699.000000	
mean	1.071704e+06	4.417740	3.134478	3.207439	2.806867	
std	6.170957e+05	2.815741	3.051459	2.971913	2.855379	
min	6.163400e+04	1.000000	1.000000	1.000000	1.000000	
25%	8.706885e+05	2.000000	1.000000	1.000000	1.000000	
50%	1.171710e+06	4.000000	1.000000	1.000000	1.000000	
75%	1.238298e+06	6.000000	5.000000	5.000000	4.000000	
max	1.345435e+07	10.000000	10.000000	10.000000	10.000000	

3.1 EXPLORATORY DATA ANALYSIS

3.2 DATA VISUALIZATION

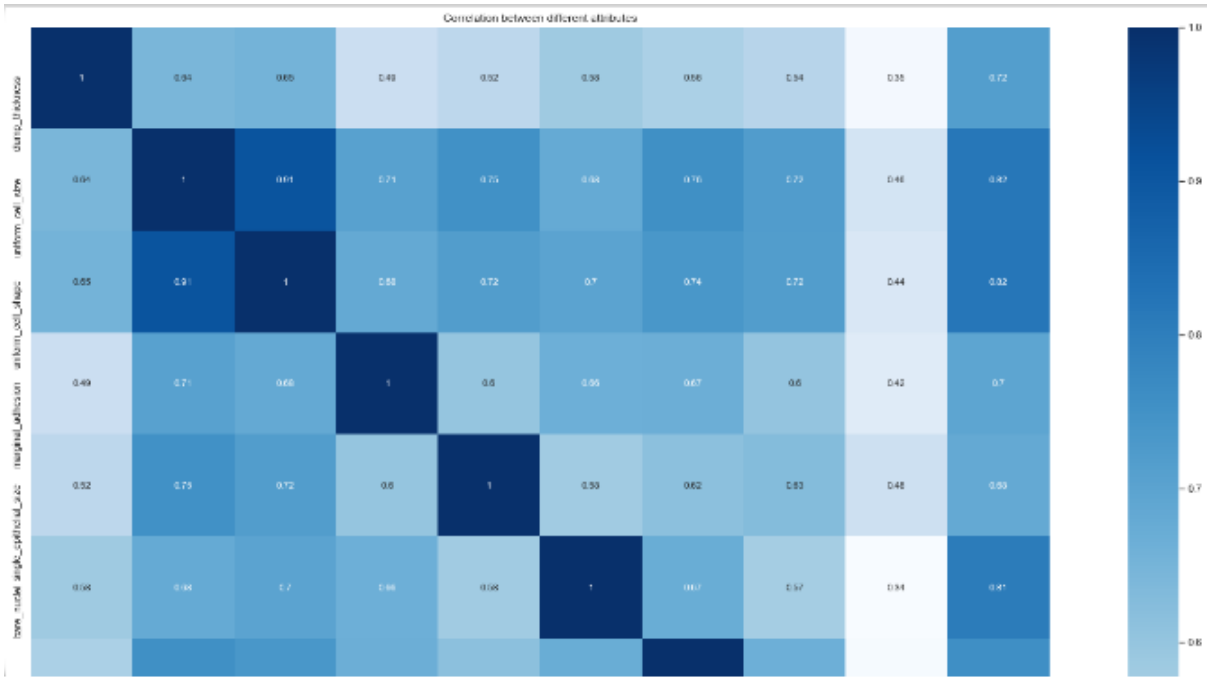
Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.



3.2 DATA VISUALIZATION

3.3 CORRELATION ANALYSIS

Correlation analysis in research is a statistical method used to measure the strength of the linear relationship between two variables and compute their association. Simply put - correlation analysis calculates the level of change in one variable due to the change in the other. A high correlation points to a strong relationship between the two variables, while a low correlation means that the variables are weakly related. When it comes to market research, researchers use correlation analysis to analyze quantitative data collected through research methods like surveys and live polls. They try to identify the relationship, patterns, significant connections, and trends between two variables or datasets. There is a positive correlation between two variables when an increase in one variable leads to the increase in the other. On the other hand, a negative correlation means that when one variable increases, the other decreases and vice-versa.



3.3 CORRELATION ANALYSIS

3.4 SOFTWARE & LIBRARIES

Normally, a library is a collection of books or is a room or place where many books are stored to be used later. Similarly, in the programming world, a library is a collection of precompiled codes that can be used later on in a program for some specific well-defined operations. Other than pre-compiled codes, a library may contain documentation, configuration data, message templates, classes, and values, etc.

A Python library is a collection of related modules. It contains bundles of code that can be used repeatedly in different programs. It makes Python Programming simpler and convenient for the programmer. As we don't need to write the same code again and again for different programs. Python libraries play a very vital role in fields of Machine Learning, Data Science, Data Visualization, etc.

SOFTWARE / LIBRARIES	USAGE
Python3 software	Programming language
Pandas	To work with data in data frames and data exploration
NumPy	For scaling and reshaping data
Sklearn	Pre-processing, model selection, metrics
Matplotlib, Seaborn	For data visualization
Pickle	To save the trained model
HTML, CSS, JavaScript	To make the front-end
Flask	Web Framework to integrate the ML model to frontend

3.4 SOFTWARE AND LIBRARIES

CHAPTER 4
PROTOTYPE

4.1 PYTHON3 SOFTWARE

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Python is a multi-paradigm programming language. Object-oriented programming and structured programming are fully supported, and many of their features support functional programming and aspect-oriented programming (including metaprogramming and metaobjects). Many other paradigms are supported via extensions, including design by contract and logic programming.

4.1 PYTHON SOFTWARE

4.2 PANDAS

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Its name is a play on the phrase "Python data analysis" itself. Wes McKinney started building what would become pandas at AQR Capital while he was a researcher there from 2007 to 2010.

4.2 PANDAS LIBRARY

4.3 NUMPY

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. NumPy is open-source software and has many contributors. NumPy is a NumFOCUS fiscally sponsored project.

4.3 NUMPY LIBRARY

4.4 SK LEARN

Scikit-learn (formerly **scikits.learn** and also known as **sklearn**) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support-vector machines, random forests, gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. Scikit-learn is a NumFOCUS fiscally sponsored project. scikit-learn is an open-source Python library that implements a range of machine learning, pre-processing, cross-validation, and visualization

algorithms using a unified interface. It is simple and efficient tools for data mining and data analysis. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means, etc. It is accessible to everybody and reusable in various contexts. Built on the top of NumPy, SciPy, and matplotlib. It is open source, commercially usable – BSD license.

4.4 SCIKIT-LEARN LIBRARY

4.5 MATPLOTLIB

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged.[3] SciPy makes use of Matplotlib.

4.5 MATPLOTLIB LIBRARY

4.6 SEABORN

Seaborn is one of an amazing library for visualization of the graphical statistical plotting in Python. Seaborn provides many color palettes and defaults beautiful styles to make the creation of many statistical plots in Python more attractive. Seaborn library aims to make a more attractive visualization of the central part of understanding and exploring data. It is built on the core of the

matplotlib library and also provides dataset-oriented APIs. Seaborn is also closely integrated with the Panda's data structures, and with this, we can easily jump between the various different visual representations for a given variable to better understand the provided dataset.

4.6 SEABORN LIBRARY

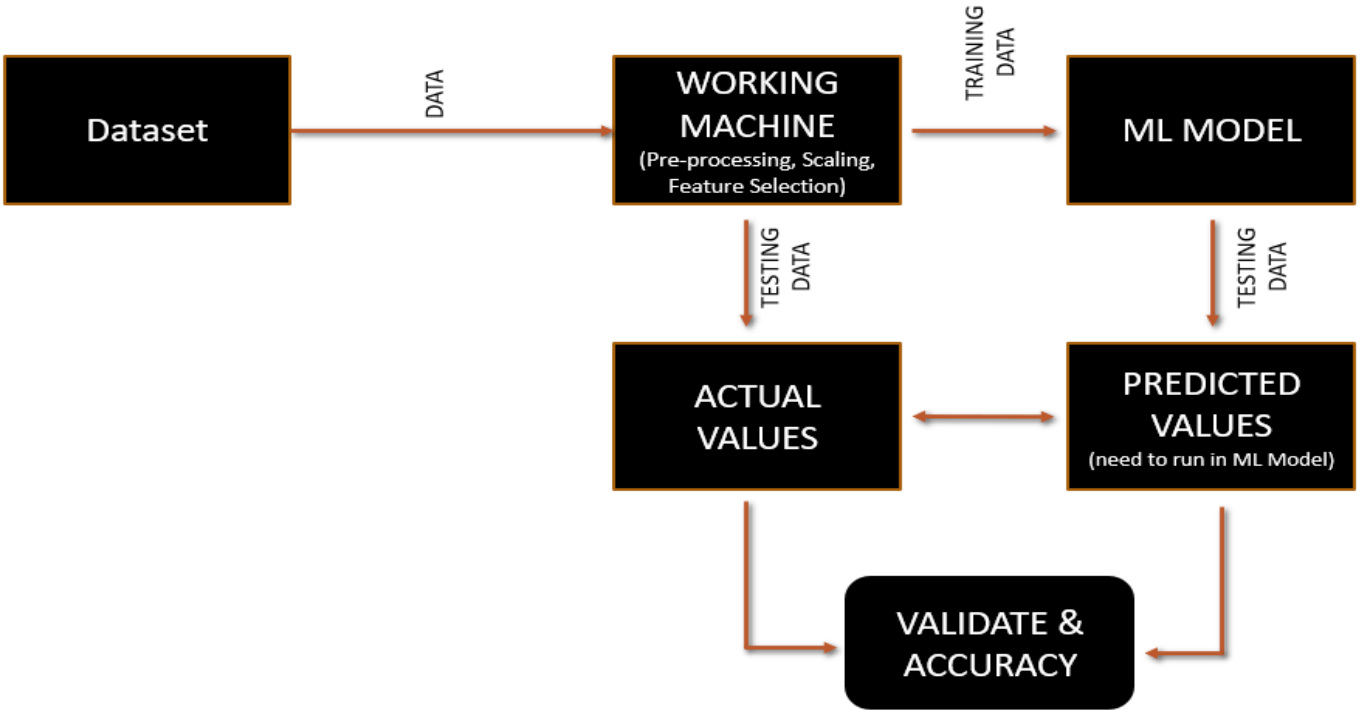
4.7 FLASK

Flask is a web framework, it's a Python module that lets you develop web applications easily. It's has a small and easy-to-extend core: it's a microframework that doesn't include an ORM (Object Relational Manager) or such features. It does have many cool features like url routing, template engine. It is a WSGI web app framework.

Flask is a web application framework written in Python. It was developed by Armin Ronacher, who led a team of international Python enthusiasts called Poocco. Flask is based on the Werkzeg WSGI toolkit and the Jinja2 template engine.Both are Pocco projects.

4.7 FLASK FRAMEWORK

4.8 ARCHITECTURAL DIAGRAM



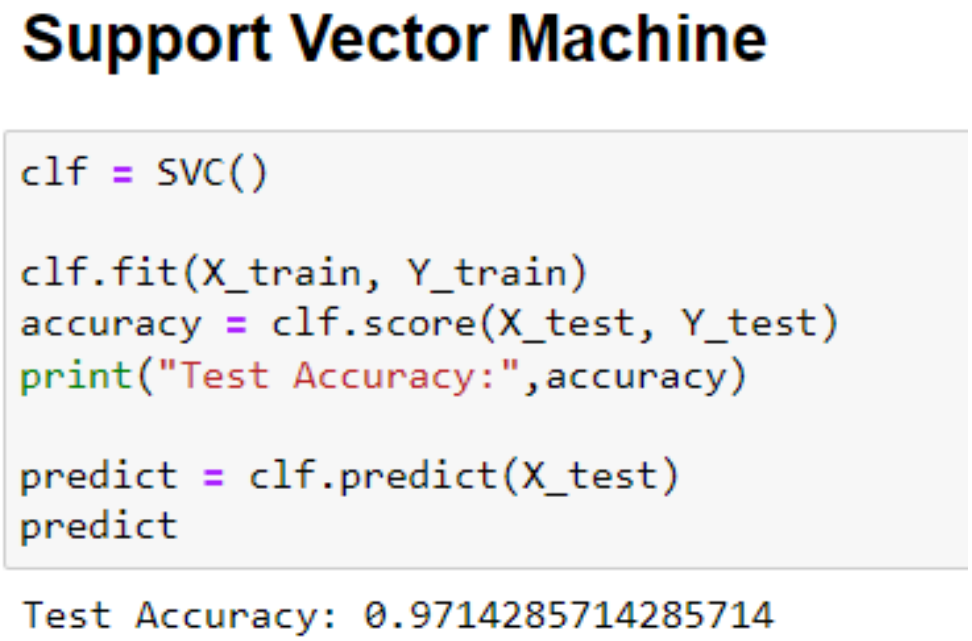
4.8 ARCHITECTURAL DIAGRAM

4.9 SUPPORT VECTOR MACHINE ALGORITHM

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification, implicitly mapping their inputs into high-dimensional feature spaces.

4.10 PREDICTION ACCURACY



4.10 PREDICTION ACCURACY

CHAPTER 5
TESTING

	id	clump_thickness	uniform_cell_size	uniform_cell_shape	marginal_adhesion
count	6.990000e+02	699.000000	699.000000	699.000000	699.000000
mean	1.071704e+06	4.417740	3.134478	3.207439	2.806867
std	6.170957e+05	2.815741	3.051459	2.971913	2.855379
min	6.163400e+04	1.000000	1.000000	1.000000	1.000000
25%	8.706885e+05	2.000000	1.000000	1.000000	1.000000
50%	1.171710e+06	4.000000	1.000000	1.000000	1.000000
75%	1.238298e+06	6.000000	5.000000	5.000000	4.000000
max	1.345435e+07	10.000000	10.000000	10.000000	10.000000

```
#Shape of the Dataset
df.shape
```

Data pre-processing

```
df.drop(['id'],axis=1,inplace = True)
```

```
# Columns in the dataset
df.columns
```

```
Index(['clump_thickness', 'uniform_cell_size', 'uniform_cell_shape',
       'marginal_adhesion', 'single_epithelial_size', 'bare_nuclei',
       'bland_chromatin', 'normal_nucleoli', 'mitoses', 'class'],
      dtype='object')
```

Handling missing values

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 699 entries, 0 to 698
Data columns (total 10 columns):
 #   Column                  Non-Null Count  Dtype 
---  ------                  --------------  ----- 
 0   clump_thickness         699 non-null    int64 
 1   uniform_cell_size       699 non-null    int64 
 2   uniform_cell_shape      699 non-null    int64 
 3   marginal_adhesion       699 non-null    int64 
 4   single_epithelial_size  699 non-null    int64 
 5   bare_nuclei             699 non-null    object
 6   bland_chromatin         699 non-null    int64 
 7   normal_nucleoli         699 non-null    int64 
 8   mitoses                 699 non-null    int64 
 9   class                   699 non-null    int64 
dtypes: int64(9), object(1)
memory usage: 54.7+ KB
```

```
#Diagnosis class Malignant = 4 and Benign = 2
#The number of Benign and Maglinant cases from the dataset
df['class'].value_counts()
```

```
2    458
4    241
Name: class, dtype: int64
```

```
df.fillna(method='ffill', inplace=True)
```

```
df.isna().sum()
```

```
clump_thickness              0
uniform_cell_size            0
uniform_cell_shape           0
marginal_adhesion            0
single_epithelial_size       0
bare_nuclei                  0
bland_chromatin              0
normal_nucleoli              0
mitoses                      0
class                        0
dtype: int64
```

Train and Test Model

```
#Split the data into predictor variables and target variable, following by break

Y = df['class'].values
X = df.drop('class', axis=1).values

X_train, X_test, Y_train, Y_test = train_test_split (X, Y, test_size = 0.30, ran
```

```
# Define models to train
models= []
models.append(('CART', DecisionTreeClassifier()))
models.append(('SVM', SVC()))
models.append(('NB', GaussianNB()))
models.append(('KNN', KNeighborsClassifier()))

# evaluate each model in turn
results = []
names = []

for name, model in models:
    kfold = KFold(n_splits=10)
    cv_results = cross_val_score(model, X_train, Y_train, cv=kfold,
    results.append(cv_results)
    names.append(name)
    msg = "For %s Model:Mean accuracy is %f (Std accuracy is %f)" %
    print(msg)
```

```
For CART Model:Mean accuracy is 0.957058 (Std accuracy is 0.023173)
For SVM Model:Mean accuracy is 0.971386 (Std accuracy is 0.013512)
For NB Model:Mean accuracy is 0.963223 (Std accuracy is 0.025463)
For KNN Model:Mean accuracy is 0.971386 (Std accuracy is 0.016306)
```

Support Vector Machine

```
clf = SVC()

clf.fit(X_train, Y_train)
accuracy = clf.score(X_test, Y_test)
print("Test Accuracy:",accuracy)

predict = clf.predict(X_test)
predict
```

```
Test Accuracy: 0.9714285714285714
```

```
example_measures = [[4,2,1,1,1,2,3,2,1]]
prediction = clf.predict(example_measures)
print(prediction)
```

```
[2]
```

Breast Cancer Detection

Enter Cell Details

Clump Thickness	1
Uniform Cell size	2
Uniform Cell shape	1
Marginal Adhesion	4
Single Epithelial Cell Size	5
Bare Nuclei	1
Bland Chromatin	3
Normal Nucleoli	1
Mitoses	2

Predict Cancer

Patient has Breast cancer

CHAPTER 6
CONCLUSION

6.1 CONCLUSION

The final accuracy value is 97.14% by the Support Vector Machine algorithm. The Machine learning Project done successfully with 97.14% accuracy which is great for deployment. Now, we are ready to deploy our ML model for the healthcare purposes.

6.2 FUTURE WORKS

This type of AI technologies can change the world into massive automation. To make their work more automotive and easier, the model can be deployed in the app or website. By integrating this ML model to the backend of the app, when user provides the required data, the ML model process the output easily.

REFERENCE

- https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470
- https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html
- https://www.cancer.gov/types/breast
- https://en.wikipedia.org/wiki/Machine_learning
- https://www.foreseemed.com/blog/machine-learning-in-healthcare