

Jaypee Institute of Information Technology, Noida
MINOR PROJECT - I
Project Report



GROUP MEMBERS DETAIL:-

GROUP NO. - “31”

SNO	ENROLLMENT NO.	BATCH	NAME
1.	19103077	B2	DIVYANSH DHINGRA
2.	19103081	B2	CHETANYA KUMAR JHA
3.	19103085	B2	HARSH CHAUHAN

GROUP MENTOR:- Ms. Deepti Singh

TITLE OF PROJECT:-

“Cure&Care”

(WE CARE ABOUT YOUR HEALTH)

ACKNOWLEDGEMENT:-

*It gives us immense pleasure to express our gratitude & appreciation towards our project supervisor **Ms. Deepti Singh** for her valuable guiding hand. Her useful suggestions for this whole work and cooperative behavior are sincerely acknowledged. We would like to express our sincere thanks to her for giving us this opportunity to undertake this project. We would also like to express our indebtedness to our parents as well as our family members whose blessings and support always helped us to face the challenges ahead.*

INTRODUCTION:-

Healthcare is a part of life we all consider to be something we are entitled to – it is our right to have access to healthcare whenever we need it. However, for many people across the world, this is simply not an option. Those living in poverty or dealing with the devastating effects of war and conflict do not have access to healthcare – even the most basic medical supplies are out of reach for many vulnerable people. Without access to healthcare, people are far more at risk. Diseases, malnutrition, pregnancy – all can become incredibly dangerous without proper help and support. An efficient health care system can contribute to a significant part of a country's economy, development, and industrialization. Health care is conventionally regarded as an important determinant in promoting the general physical and mental health and well-being of people around the world.

The purpose of “Cure&Care” is to provide better care for patients and help achieve health equity. We are living during difficult times because of the coronavirus pandemic. Those challenges can affect us both physically and emotionally. It’s hard to keep the same routine when you have to physically distance yourself from others. This pandemic has also affected us on a monetary basis. People with less income find it hard and unmanageable to spend on the health sector and get their basic tests done because of the fees charged. So our project provides solutions to all such issues. A person can log in and get to know about the importance of health care and use various functionalities.

The motivation behind our project is to apply the knowledge gained during the time being and provide the solution in form of a website that is integrated with machine learning models which gives users a convenient way for the confirmation of their hazardous diseases. It is a Research cum Development based project which required the detailed study of various research papers, documentation regarding machine learning, and also the knowledge of various medical terms.

In this project “Cure & Care” we have given a glimpse of various predictive tests which include the Heart Disease Prediction Test, Liver Cancer Prediction Test, Kidney Cancer Prediction Test, Diabetes Prediction Test. All the predictive analyses have been done using various classification models.

DATASET:-

For **DIABETES PREDICTION MODEL** - Pima Indians Diabetes Database has been selected from the KAGGLE website which includes:-

(<https://www.kaggle.com/uciml/pima-indians-diabetes-database>)

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

For **HEART DISEASE PREDICTION MODEL** - Heart Disease UCI Dataset has been selected from the KAGGLE website which includes:-

(<https://www.kaggle.com/ronitf/heart-disease-uci>)

```
heart_dis = pd.read_csv("heart.csv")
```

```
heart_dis.head()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

For **LIVER CANCER PREDICTION MODEL** - Indian Liver Patient Records Dataset has been selected from the KAGGLE website which includes:-

(<https://www.kaggle.com/uciml/indian-liver-patient-records>)

```
liver_dis = pd.read_csv("indian_liver_patient.csv")
```

```
liver_dis.head()
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin
0	65	Female	0.7	0.1	187	16	18	6.8	3.3
1	62	Male	10.9	5.5	699	64	100	7.5	3.2
2	62	Male	7.3	4.1	490	60	68	7.0	3.3
3	58	Male	1.0	0.4	182	14	20	6.8	3.4
4	72	Male	3.9	2.0	195	27	59	7.3	2.4

For **KIDNEY CANCER PREDICTION MODEL** - Indian Liver Patient Records Dataset has been selected from the KAGGLE website which includes:-

(<https://www.kaggle.com/mansoordaku/ckdisease>)

```
kidney_dis = pd.read_csv("kidney_disease.csv")
```

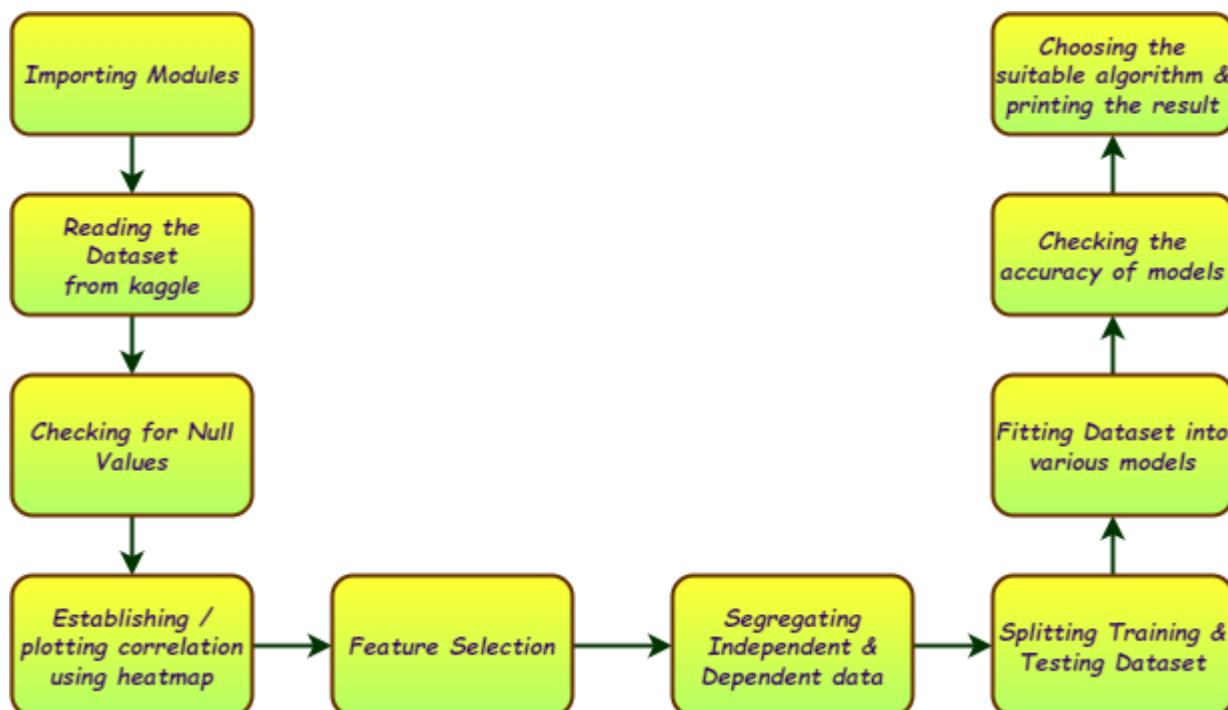
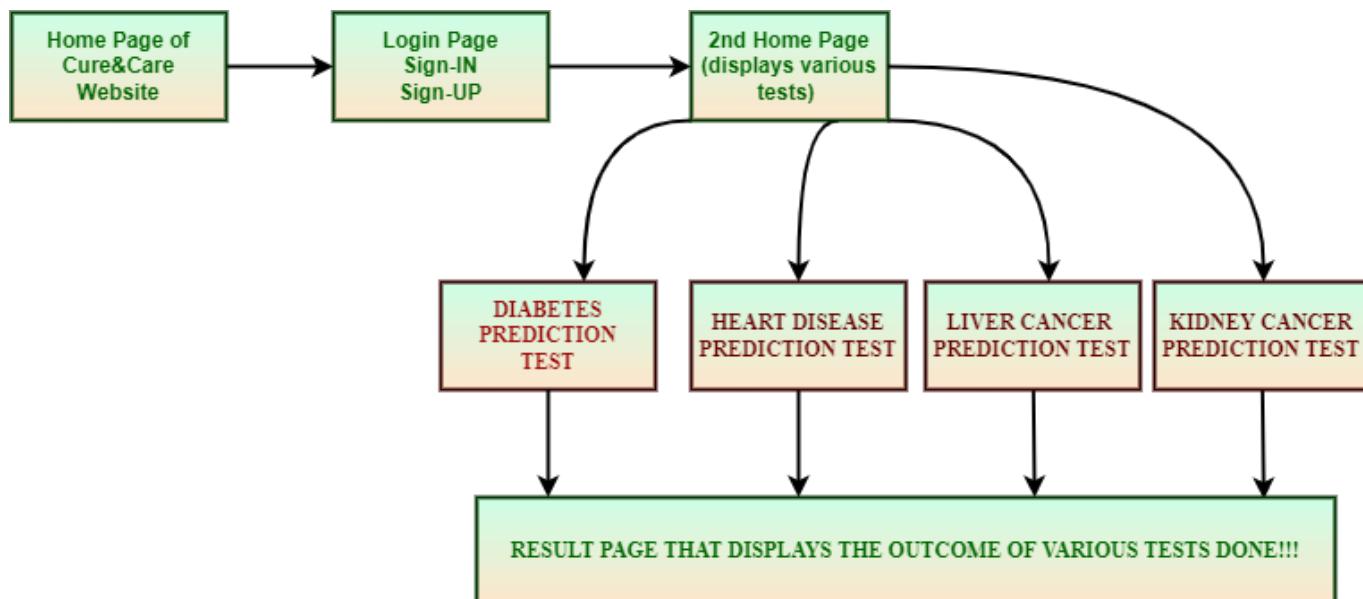
```
kidney_dis.head()
```

	id	age	bp	sg	al	su	rbc	pc	pcc	ba	...	pcv	wc	rc	htn	dm	cad	appet	pe	ane	classification
0	0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	...	44	7800	5.2	yes	yes	no	good	no	no	ckd
1	1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	...	38	6000	NaN	no	no	no	good	no	no	ckd
2	2	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	...	31	7500	NaN	no	yes	no	poor	no	yes	ckd
3	3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	...	32	6700	3.9	yes	no	no	poor	yes	yes	ckd
4	4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	...	35	7300	4.6	no	no	no	good	no	no	ckd

METHODOLOGY:-

1. Data Collection
2. Data Cleaning
3. Data Correlation & Pre-processing
4. Data Visualization
5. Train - Test Data Split
6. Model Building
7. Prediction on Testing Dataset
8. Visualization of Predicted model

FLOWCHART:-



DATA CLEANING:-

Rows with missing values have been appended using mean values whereas independent features were interpolated. Feature scaling was done using StandardScaler() & MinMaxScaler().

```
...    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 4  
    Dataset         0  
    dtype: int64
```

```
[23]    liver_dis['Gender'] = liver_dis['Gender'].map({'Male': 1, 'Female': 2})  
✓ 0.6s
```

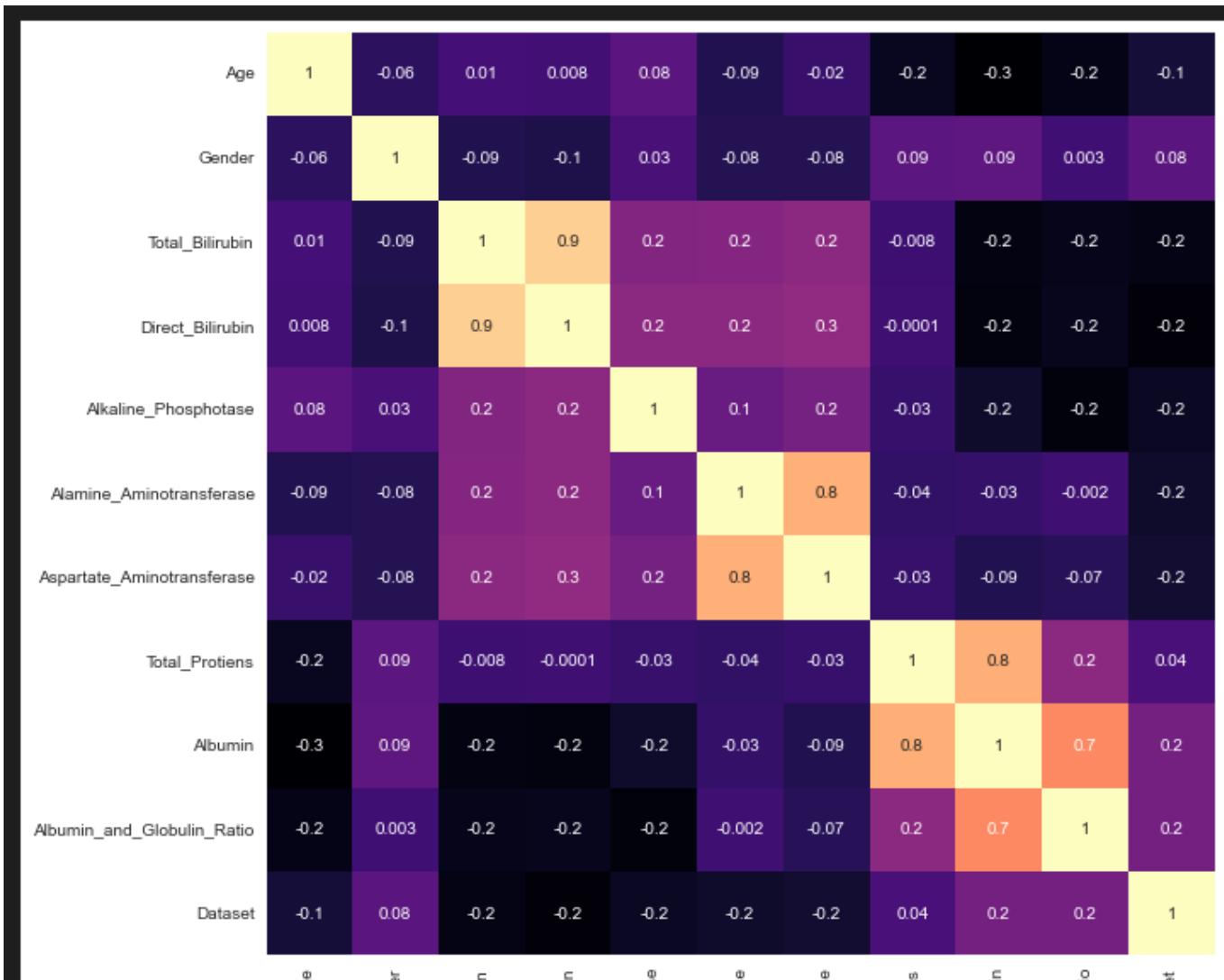
```
[24]    l=[]  
    for i in range(len(liver_dis['Albumin_and_Globulin_Ratio'].values)):  
        if pd.isnull(liver_dis['Albumin_and_Globulin_Ratio'].values[i]) != True:  
            l.append(liver_dis['Albumin_and_Globulin_Ratio'].values[i])  
    np.mean(l)  
✓ 0.7s  
... 0.9470639032815199
```

```
from sklearn.preprocessing import StandardScaler  
st_sc = StandardScaler()  
indep_train = st_sc.fit_transform(indep_train)  
indep_test = st_sc.transform(indep_test)
```

```
indep_train  
  
array([[-0.52639686, -1.15139792, -3.75268255, ..., -4.13525578,  
       -0.49073479, -1.03594038],  
      [ 1.58804586, -0.27664283,  0.68034485, ..., -0.48916881,  
       2.41502991,  1.48710085],  
      [-0.82846011,  0.56687102, -1.2658623 , ..., -0.42452187,  
       0.54916055, -0.94893896],  
      ...,  
      [ 1.8901091 , -0.62029661,  0.89659009, ...,  1.76054443,  
       1.981245 ,  0.44308379],  
      [-1.13052335,  0.62935353, -3.75268255, ...,  1.34680407,  
       -0.78487662, -0.33992901],  
      [-1.13052335,  0.12949347,  1.43720319, ..., -1.22614383,  
       -0.61552223, -1.03594038]])
```

DATA CORRELATION:-

Data correlation has been established and seen using the seaborn & matplotlib module. Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.



TRAIN TEST SPLIT:-

For training & testing data in our project, we have provided 70% to train the model & 30% to check or test the model's precision.

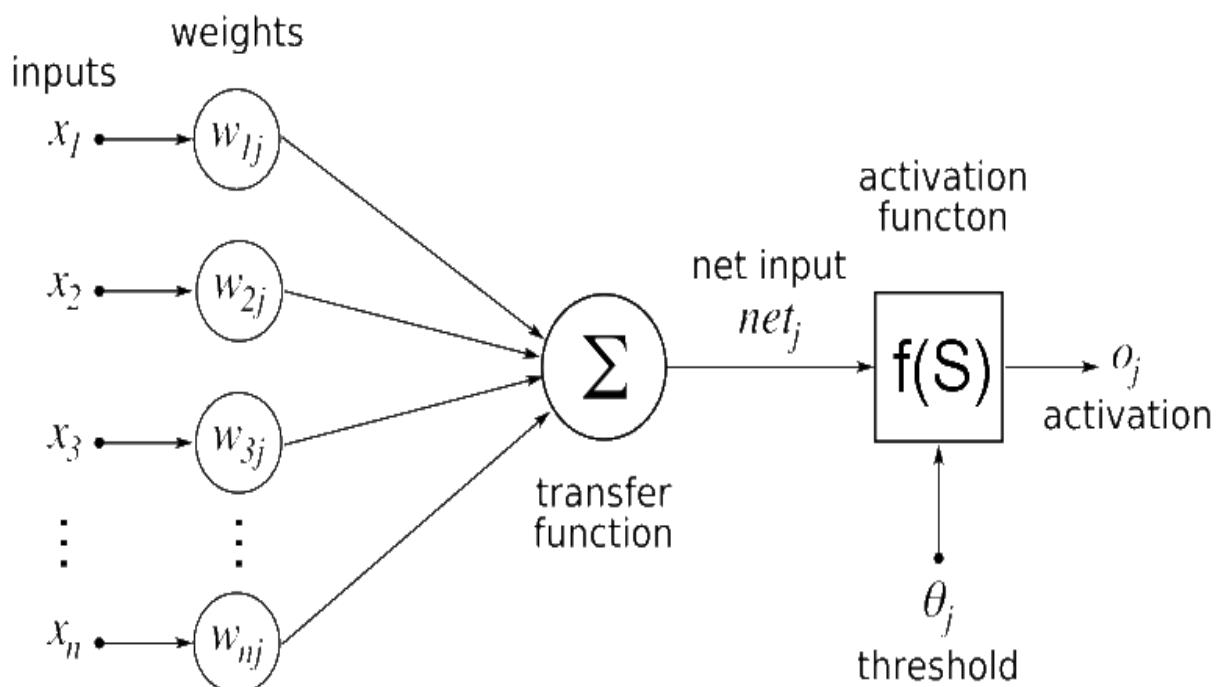
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
print('Shape training set: X:{} , y:{}' .format(X_train.shape, y_train.shape))
print('Shape test set: X:{} , y:{}' .format(X_test.shape, y_test.shape))
```

```
Shape training set: X:(212, 10), y:(212,)
Shape test set: X:(91, 10), y:(91,)
```

MODEL BUILDING:-

For our different predictive analysis, we have trained our dataset in different classification models which includes:-

1. **Logistic Regression**:- Logistic regression is a statistical machine learning algorithm that classifies the data by considering outcome variables on extreme ends and tries makes a logarithmic line that distinguishes between them.
2. **Random Forest Classifier**:- A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.
3. **Decision Tree Classifier**:- Decision tree builds classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.
4. **Support Vector Machine**:- The supervised learning algorithm Support Vector Regression is used to predict discrete values. SVMs and Support Vector Regression are both based on the same premise. The basic idea behind SVR is to find the best fit line. In SVR, the best fit line is the hyperplane that has the maximum number of points.
5. **K Nearest Neighbors**:- K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. The nearest neighbor-based classifiers use some or all the patterns available in the training set to classify a test pattern.
6. **Artificial Neural Network**:- Artificial Neural networks (ANN) or neural networks are computational algorithms. It is intended to simulate the behavior of biological systems composed of “neurons”. The process records one at a time and learns by comparing their classification of the record with the known actual classification of the record.



RESULTS & COMPARING ACCURACIES:-

Heart Disease Prediction Model:-

MODELS USED	ACCURACY SCORE (%age)
Logistic Regression	83.51648
Random Forest Classifier	76.22580
K-Nearest Neighbors	65.93406
Decision Tree Classifier	78.02197
Support Vector Machine	82.51676

Liver Cancer Prediction Model:-

Models Used	Accuracy (%age)
Random Forest Classifier	77.39726
Logistic Regression	71.91780
Decision Tree Classifier	66.43835
K-Nearest Neighbors	70.54794
Support Vector Machine	74.65753

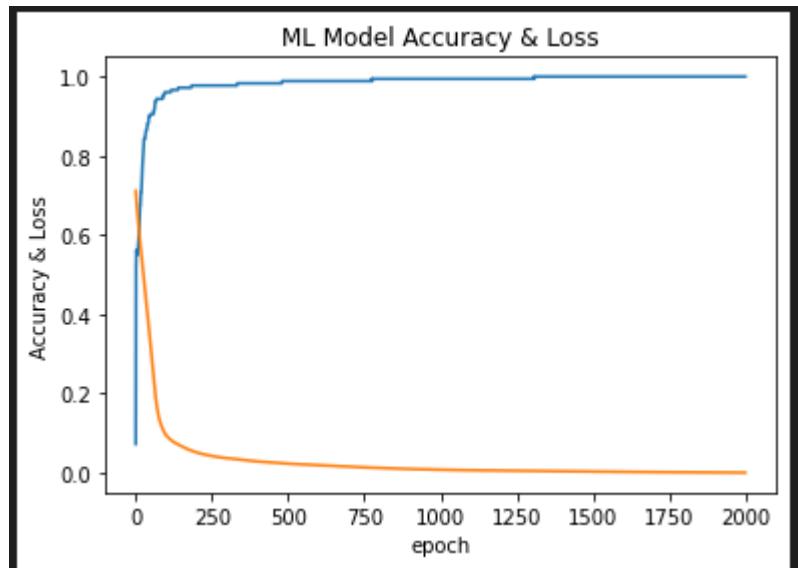
Diabetes Prediction Model:-

Models Used	Accuracy (%age)
Support Vector Machine	75.97402
Logistic Regression	75.32467
K-Nearest Neighbors	70.12987
Decision Tree Classifier	74.02597
Random Forest Classifier	73.37662

Kidney Cancer Prediction Model:-

For the Kidney Cancer Prediction Model, we have used Artificial Neural Network. These days the ANNs are broadly used for the motive of prognosis of diseases. Due to its wide studying abilities and fault tolerance, it's miles maximum famous in clinical prognosis. Artificial Neural Networks are a promising means for diagnosis.

Accuracy Score:-



Classification Report

```
● clf_report = classification_report(y_test, pred_data)
print('Classification report')
print("_____")
print(clf_report)
print("_____")
```

✓ 0.9s

Classification report

	precision	recall	f1-score	support
0	0.81	0.83	0.82	41
1	0.86	0.84	0.85	50
accuracy			0.84	91
macro avg	0.83	0.83	0.83	91
weighted avg	0.84	0.84	0.84	91

```
clf_rep = classification_report(y_test, pred_d)
print('Classification report')
print("_____")
print(clf_rep)
print("_____")
```

✓ 0.1s

Classification report

	precision	recall	f1-score	support
0	0.63	0.59	0.61	41
1	0.68	0.72	0.70	50
accuracy			0.66	91
macro avg	0.66	0.65	0.65	91
weighted avg	0.66	0.66	0.66	91

WEB UI:-

A web UI has also been created for our project which we will successfully integrate with our models in the future. This UI provides various functionalities like a login page or creating your account, as soon as you login into the website it opens the second home page for you which provides various functionalities like Diabetes Detection, Liver Cancer detection, Heart Disease Detection, Kidney Cancer Detection. Web UI has been created using HTML5 & CSS3. In the future, we hope to carry forward this project and build more functionalities, create API & various other things. Some snapshots of our Web UI are as follows -

First Home Page



Dr. Stephanie Wosniack

OUR TEAM

Dr. Stephanie Wosniack is dedicated to providing her patients with the best possible care. We at Cure and Care are focused on helping you. After receiving successful care for various aches and pains over the years, Dr. Wosniack found her calling to help others get well.

A handwritten signature in black ink that reads "Stephanie".



Heart Specialist

Derrick Gray

[Find out more!!](#)



Kidney Specialist

Cary Bond

[Find out more!!](#)

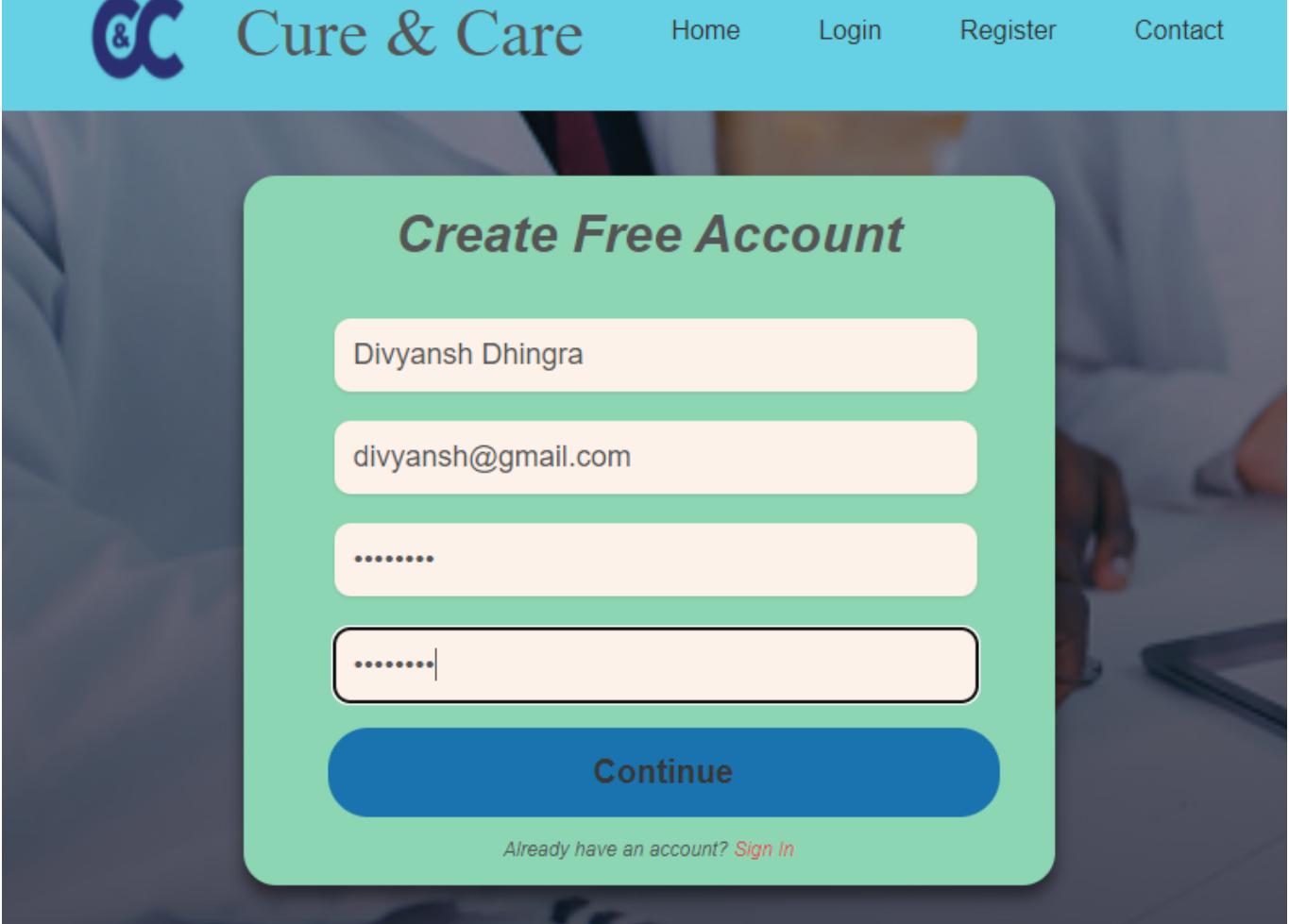


Liver Specialist

Mary Osborne

[Find out more!!](#)

Create Account Page



The page features a teal header bar with the Cure & Care logo and navigation links for Home, Login, Register, and Contact. A large central form is titled "Create Free Account". It contains four input fields: Name (Divyansh Dhingra), Email (divyansh@gmail.com), and two password fields (both showing masked input). A blue "Continue" button is at the bottom, and a "Sign In" link is at the bottom right.

Create Free Account

Divyansh Dhingra

divyansh@gmail.com

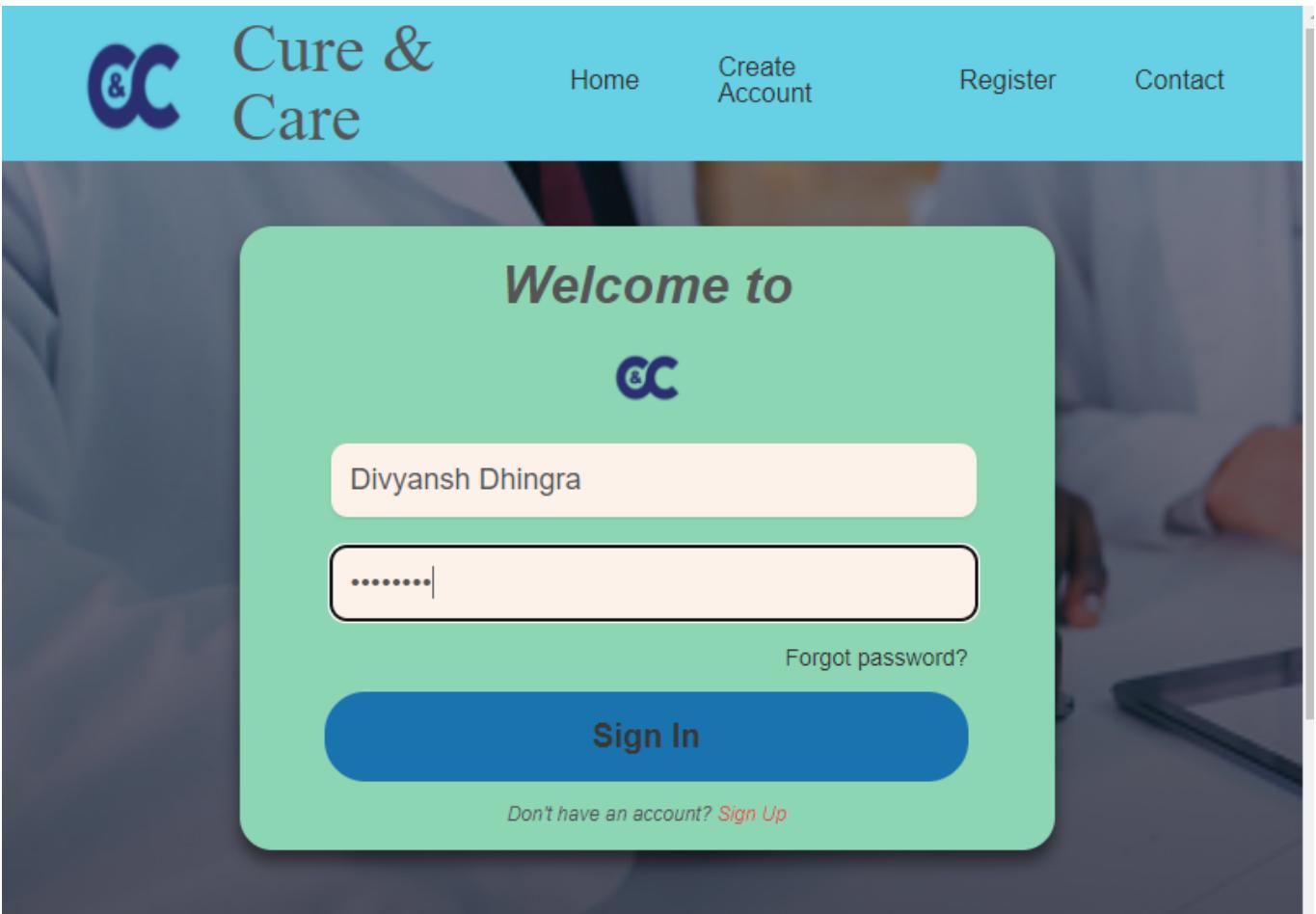
.....

.....|

Continue

Already have an account? [Sign In](#)

Login page\



The page features a teal header bar with the Cure & Care logo and navigation links for Home, Create Account, Register, and Contact. A large central form is titled "Welcome to" with the Cure & Care logo below it. It contains two input fields: Email (Divyansh Dhingra) and Password (masked input). Below the fields are links for "Forgot password?" and "Sign In". At the bottom left is a "Sign Up" link.

Welcome to

Cure & Care

Divyansh Dhingra

.....|

[Forgot password?](#)

Sign In

[Don't have an account? Sign Up](#)



1234567890



cureandcare@gmail.com



Cure & Care

[Heart Test](#)[Liver Test](#)[Kidney Test](#)[Diabetes Test](#)[Our Doctors](#)

Dr. Stephanie Wosniack

OUR TEAM

Dr. Stephanie Wosniack is dedicated to providing her patients with the best possible care. We at Cure and Care are focused on helping you. After receiving successful care for various aches and pains over the years, Dr. Wosniack found her calling to help others get well.



Heart Disease Prediction Web page

HEART DISEASE TEST

"Healthy habits are your heart's desire."

Age e.g 20

Gender

Chest Pain Type 0,1,2 or 3

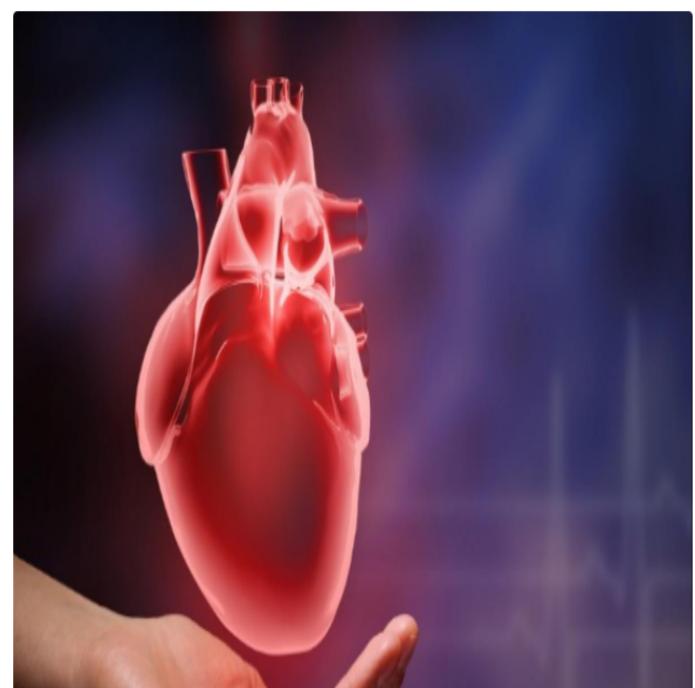
Resting Blood Pressure

Cholesterol

Fasting Blood Sugar 0 or 1

Resting ECG Result 0,1 or 2

Maximum Heart Rate



KIDNEY DISEASE TEST

"Protect your kidneys, Lead a healthier life."

Specific Gravity	<input type="text"/>
Albumin	<input type="text"/>
Serum Creatinine	<input type="text"/>
Haemoglobin	<input type="text"/>
Packed Cell Volume	<input type="text"/>
WBC Count	<input type="text"/>
RBC Count	<input type="text"/>
Hypertension	<input type="text"/> 0 or 1



Liver Cancer Prediction Web page

LIVER DISEASE TEST

"I only weigh 98 pounds. The rest of this is liver."

Age	<input type="text"/> e.g 20
Gender	<input type="text"/> Male/Female
Total Bilirubin	<input type="text"/>
Alkaline Phosphotase	<input type="text"/>
Alanine Aminotransferase	<input type="text"/>
Aspartate Aminotransferase	<input type="text"/>
Total Proteins	<input type="text"/>
Albumin	<input type="text"/>



Diabetes Prediction Web page

DIABETES DISEASE TEST

"I am stronger than diabetes. I am going to be around, till there is a cure."

Age	<input type="text"/>
BMI	<input type="text"/>
Insulin	<input type="text"/>
Glucose	<input type="text"/>
Blood Pressure	<input type="text"/>
Pregnancies	<input type="text"/>
Skin Thickness	<input type="text"/>
Diabetes Pedigree Function	<input type="text"/>



!! Test Yourself !!



Heart Disease Test

[TAKE A TEST](#)



Liver Disease Test

[TAKE A TEST](#)



Kidney Disease Test

[TAKE A TEST](#)



Diabetes Disease Test

[TAKE A TEST](#)

MODULES IMPORTED:-

- *Numpy*
- *Sklearn*
 1. *train_test_split*
 2. *LabelEncoder*
 3. *StandardScaler*
 4. *MinMaxScaler*
 5. *accuracy_score*
 6. *classification_report*
 7. *confusion_matrix*
- *Pandas*
- *Matplotlib*
- *Seaborn*
- *Keras*
 1. *Sequential*
 2. *load_model*
 3. *Dense*
- *Flask*
 1. *render_template*
 2. *request*

CONCLUSION:-

The main objective of this project was to put forth our learnings & knowledge gained during the course time in form of our imagination. This Cure & Care website aims to give a handy experience to the users. Furthermore, many functionalities are to be implemented with proper backend knowledge. Nowadays, since the Healthcare sector is thriving and people are getting more conscious of their health with the advent time so we contemplate the idea of carrying forward this website & providing an easy end-to-end solution to the people.

FUTURE SCOPE:-

We envisioned creating a web interface for the users that would provide them with an easy-to-go experience. We intend to add various functionalities as we carry forward our project. Some of the functionalities that we are planning to implement -

- Fully integrating our ML/AI models to the web interface
- Providing the function for self-care
- Integrating OpenCV
- Building a user-friendly interface for the novice users
- Integrating Google maps for locating the nearby doctors & aid rooms
- And many more functionalities.....

DIVISION OF WORK:-

Enrollment No	Name	Contribution
19103077	Divyansh Dhingra	Majorly worked on developing machine learning models & applying the knowledge gained in predictive analysis. Tried integrating the models to the website using flask framework
19103081	Chetanya Kumar Jha	Contributed to developing models & analyzing the dataset & models. Also helped in report creation.
19103085	Harsh Chauhan	Contributed to developing the front-end website. Created various web pages & integrated them. Helped in designing part.

RESEARCH PAPER ANALYSIS:-

MACHINE LEARNING WITH HEALTHCARE PERSPECTIVE (By Jyotir Moy Chatterjee)

The value of machine learning in health care is its ability to process huge datasets beyond the scope of human capability and then reliably convert analysis of that data into clinical insights that aid physicians in planning and providing care, ultimately leading to better outcomes, lower costs of care, and increased patient satisfaction.

CLASSIFICATION USING SUPERVISED LEARNING & INTEGRATING WITH HEALTHCARE (By Shovan Chowdhury, Marco P. Schoen) -

The ultimate objective is to extract meaningful information from published abstracts. In pursuing this objective, the ML techniques are utilized to classify different publications into various fields: Healthcare, Science, etc. The comparative study of various classification models has been very well delineated & analyzed.

SKLEARN DOCUMENTATION

The detailed study of different machine learning concepts & models is very well explained. It gives an insight into the thorough study of regression models, classification models, clustering models & so on.

ARTIFICIAL NEURAL NETWORKS - A STUDY (By Er. Parveen Kumar, Er. Pooja Sharma)

An introduction to an artificial neural network gives a complete overview of its implementation. The research paper concludes the working of Neural Networks, characteristics of Neural Networks, Limitations & Applications.

A NOVEL APPROACH FOR HEART DISEASE PREDICTION (By Yin Kia Chiam, Wan Azman, Armin Yazdani)

The study is aimed at predicting heart disease based on the scores of significant features using Weighted Associative Rule Mining. This study managed to provide a significant contribution in computing the strength scores with significant predictors in heart disease prediction.

DIAGNOSIS OF CHRONIC KIDNEY DISEASE USING EFFECTIVE CLASSIFICATION ALGORITHMS & RECURSIVE FEATURE ELIMINATION TECHNIQUES (By Ebrahime Mohammed Senan, Nizar Alsharif, Mukti E. Jadhav)

The novelty of this study lies in developing the diagnosis system to detect chronic kidney diseases. Recursive Feature Elimination (RFE) was applied. Four classification algorithms applied in this study were support vector machine (SVM), k -nearest neighbors (KNN), decision tree, and random forest.

REFERENCES:-

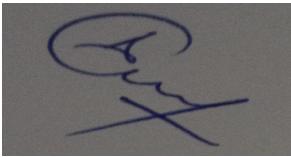
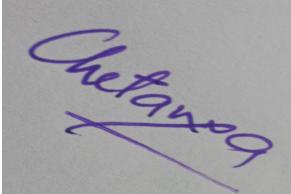
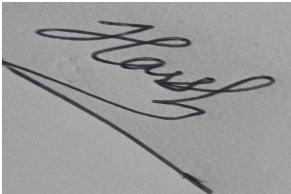
- 1) https://scikit-learn.org/stable/modules/linear_model.html
- 2) [1. Supervised learning — scikit-learn 0.23.2 documentation](#)
- 3) <https://www.kaggle.com/datasets?search=PIMA+diabetes>
- 4) <https://www.hindawi.com/journals/jhe/2021/1004767/>
- 5) <https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012072/meta>
- 6) <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01527-5>
- 7) <https://www.kaggle.com/ronitf/heart-disease-uci>
- 8) <https://www.kaggle.com/uciml/indian-liver-patient-records>
- 9) <https://www.kaggle.com/mansoordaku/ckdisease>
- 10) https://www.researchgate.net/publication/319943588_Research_study_on_applications_of_artificial_neural_networks_and_e-learning_personalization

Students' Self Declaration for Open Source libraries and other source code usage in Minor Project

I / We **Divyansh Dhingra, Harsh Chauhan, Chetanya Kumar Jha** hereby declare the following usage of the open-source code and prebuilt libraries in our minor project in **Vth Semester** with the consent of our supervisor. We also measure the similarity percentage of pre-written source code and our source code and the same is mentioned below. This measurement is true to the best of our knowledge and abilities.

List of pre-build libraries -

- *Numpy*
- *Sklearn*
- *Pandas*
- *Matplotlib*
- *Seaborn*
- *Keras*
- *Flask*

Student ID	Student Name	Student signature
19103077	Divyansh Dhingra	
19103081	Chetanya Kumar Jha	
19103085	Harsh Chauhan	

Declaration by Supervisor (To be filled by Supervisor only)

I, **Deepti Singh** (Name of Supervisor) declare that I above submitted project with Titled **Cure&Care** was conducted under my supervision. The project is original and neither the project was copied from External sources nor it was submitted earlier in JIIT. I authenticate this project.

Signature (Supervisor)