NEW HORIZON
COLLEGE OF ENGINEERING
Autonomous College, Affiliated to VTU | Approved by AICTE New Delhi & UGC
Accredited by NAAC with 'A' Grade & Accredited by NBA

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

A MINI PROJECT REPORT ON

"PREDICTION AND CLASSIFICATION OF THYROID DETECTION"

Submitted in the partial fullfillment of the requirements in the 4th semester of

BACHELOR OF ENGINEERING
IN

INFORMATION SCIENCE AND ENGINEERING

FOR

COURSE :NAME MINI PROJECT
22ISE48 :COURSE CODE

By
G.Shasi Vardhan Reedy-1NH22IS048
K.Akash-1NH22IS063

Under the guidance of
Mrs.Shalini
B.E , M.Tech
Senior Assistant Professor

NEW HORIZON
COLLEGE OF ENGINEERING
Autonomous College, Affiliated to VTU | Approved by AICTE New Delhi & UGC
Accredited by NAAC with 'A' Grade & Accredited by NBA

CERTIFICATE

Certified that the project work entitled "PREDICTION AND CLASSIFICATION OF THYROID DETECTION" carried out by G.Shasi vardhan Reddy and K.Akash bearing usn 1NH22IS048 and 1NH22IS063 respectively, bonafide students of IV semester in partial fulfillment for the award of Bachelor of Engineering in Information Science & Engineering of New Horizon College of Engineering, an autonomous institute affiliated to the Visvesvaraya Technological University, Belagavi during the year 2023-24. It is certified that all corrections / suggestions indicated for Internal Assessment have been incorporated. The project report has been approved as it satisfies the academic requirements in respect of Mini Project work prescribed for the said Degree.

Name & Signature of Guide
Mrs.A. Shalini

Name & Signature of HOD
.Dr Vandana C.P

Examiners :

	Name	Signature
1.	…….	……..
2.	……	……

ACKNOWLEDGEMENT

Any project task a is of great enormity and it cannot be accomplished an by individual without support and guidance. We are grateful to a number of individuals whose professional guidance and encouragement has made this project completion a reality.

We have a great pleasure in expressing our deep sense of gratitude to the beloved Chairman **Mohan Manghnani .Dr** for having provided us with a great infrastructure and well-furnished labs.

We take this opportunity to express our profound gratitude to the Principal **.Dr Manjunatha** for his constant support and management.

We are grateful to **P.Vandana C .Dr** , Associate Professor and Head of Department of ISE, New Horizon College of Engineering, Bengaluru for her strong enforcement on perfection and quality during the course of our mini project work.

We would like to express our thanks to the guide **Mrs.Shalini**, Senior Assistant Professor, Department of ISE, New Horizon College of Engineering, Bengaluru who has always guided us in detailed technical aspects throughout our mini project.

We would like to mention special thanks all to the Teaching and Non-Teaching staff members of Information Science and Engineering Department, New Horizon College of Engineering, Bengaluru for their invaluable support and guidance.

G.Shasi vardhan Reddy

-1NH22IS048

K.Akash

-1NH22IS063

TABLE OF CONTENTS

Abstract

Thyroid gland diseases are among the most prevalent endocrine disorders in the world, affecting people of all age groups from children to the elderly. It is the general concept for a medical problem that prevents one's thyroid from producing enough hormones. Early diagnosis and care of such diseases are thus vital in preventing disease advancement and even death in patients. Machine Learning (ML) techniques play a major role in such cases for making correct decisions, proper disease diagnosis and saving cost and time of patients.
The purpose of this study is prediction of thyroid disease using classification Predictive Modelling. Data Collection, Data Preprocessing and Several classification models including K- Nearest Neighbours (k-NN), Naive Bayes, Decision Tree models are fit to the data to perform
predictions.

We believe that this work helps in providing a better classification model for prediction and diagnosis of thyroid gland diseases. The result of this proposed work can be used for the forthcoming identification to keep track on important factors affecting thyroid gland related diseases. The various algorithm results are based on speed, accuracy and performance of the model and cost for the treatment. Also, the classification of effective data are helps to find the treatment to the thyroid patients with better cost and facilitates the management.

CHAPTER 1

INTRODUCTION

Introduction

Thyroid illnesses and disorders are widespread hormonal problems that affect most of the world's population. There are two general types of thyroid disease, each with their own symptoms. one is hyperthyroid and the other is hypothyroid5.So, for example , Random Forests,KNN, Support Vector Machine (SVM), and Decision Trees are all machine learning methods widely used for predicting thyroid diseases. Machine Learning (ML) is an artificial intelligence field based on the idea that computers can learn from data, spot patterns, and make judgments with little or no human intervention.

Problem Definition

A lot of people find it difficult to maintain new habits or break old ones. Typical difficulties consist of:

1. According to statistics, thyroid disorders are on the rise in India. Approximately 1 in 10 Indian adults suffer from thyroid problem.

2. It has been estimated that around 42 million peoples suffer from thyroid disease.

3. Predicting thyroid disorder by doctor is a tedious process which might lead to negative prediction, only experienced doctor can examine the case properly.

Motivations

The following reasons drove the development of an habit monitoring website:

1. **Prevalence of Thyroid Disorders:** Thyroid disorders, including hypothyroidism and hyperthyroidism, are common globally, affecting millions of people. Early detection can significantly improve patient outcomes.

2. **Impact on Quality of Life:** Thyroid conditions can severely impact a person's quality of life, leading to symptoms such as fatigue, weight changes, depression, and cognitive impairments. Accurate prediction and detection can help mitigate these effects.

3. **Prevention of Complications:** Untreated thyroid disorders can lead to serious health complications, including cardiovascular diseases, infertility, and osteoporosis. Early detection allows for timely treatment to prevent these issues.

4. **Cost-Effective Healthcare:** Early and accurate detection of thyroid disorders can reduce healthcare costs by minimizing the need for extensive diagnostic procedures and preventing costly complications

5. **improvement in Treatment Outcomes:** Accurate prediction models can help tailor treatment plans to individual patients, leading to better management of the disease and improved treatment outcomes.

6. **Public Health Significance:** Thyroid disorders have significant public health implications. A reliable prediction and detection system can support public health initiatives by identifying high-risk populations and enabling targeted interventions.

CHAPTER 2

LITERATURE SURVEY

1. Existing System

Thyroid disorders are prevalent, and early detection and accurate diagnosis are crucial for effective management. There are several existing systems and methods for the prediction, detection, and diagnosis of thyroid conditions. These systems can be categorized into diagnostic techniques, predictive analytics, and monitoring systems. Here's a comprehensive overview of the main systems and methods currently in use:

- **Blood Tests**

Thyroid-Stimulating Hormone (TSH) Test: Measures the amount of TSH in the blood, which indicates how well the thyroid gland is functioning.
Free T4 (Thyroxine) Test: Measures the amount of free thyroxine in the blood, which helps in assessing thyroid function.
Free T3 (Triiodothyronine) Test: Measures the amount of free triiodothyronine in the blood.
Thyroid Antibody Tests: Includes TPOAb (Thyroid Peroxidase Antibodies) and TgAb (Thyroglobulin Antibodies) tests for autoimmune thyroid conditions.

- **Imaging Techniques**

Ultrasound: Used to detect abnormalities in the thyroid gland, such as nodules or cysts.
Radioactive Iodine Uptake Test: Measures how much radioactive iodine the thyroid gland absorbs, which can indicate thyroid function.
Thyroid Scintigraphy (Scan): Provides images of the thyroid gland to identify nodules and assess their activity.

- **Machine Learning Models**

Risk Prediction Models: Algorithms such as logistic regression, random forests, and support vector machines are used to predict the likelihood of thyroid disorders based on patient data.

Deep Learning Algorithms: Convolutional Neural Networks (CNNs) and other deep learning models are used to analyze medical images and improve the detection of thyroid abnormalities.

2. Proposed System

Developing a new system for thyroid prediction and detection involves creating a comprehensive solution that integrates advanced technologies and methodologies to improve accuracy, accessibility, and efficiency in identifying thyroid disorders. Below is a detailed outline of a proposed system for thyroid prediction and detection, including its components, features, technologies, and implementation strategies.

A. Data Collection

1.1 Clinical Data Input

- **Data Sources:** Electronic Health Records (EHR), patient self-reported symptoms, and clinical examination data.
- **Features:** Interface for inputting lab results (TSH, Free T4, Free T3, Thyroid Antibodies), medical history, and physical examination findings.

1.2 Imaging Data Integration

- **Data Sources:** Ultrasound, Thyroid Scintigraphy, and other imaging techniques.
- **Features:** Integration with PACS (Picture Archiving and Communication System) for retrieving and displaying thyroid imaging results.

1.3 Home Testing Data Collection

- **Data Sources:** At-home thyroid test kits.
- **Features:** Interface for uploading and analyzing home test results, with secure transmission of data to the system.

1.4 Symptom Tracking and Wearable Devices

- **Data Sources:** Wearable health monitors and mobile apps.
- **Features:** Integration with wearable devices that track symptoms like fatigue, weight changes, and heart rate.

B. Diagnostic and Predictive Analytics

2.1 AI-Based Analysis Engine

- **Technologies:** Machine Learning and Deep Learning algorithm
- **Features:**

- **Predictive Models:** AI models such as logistic regression, random forests, and neural networks for risk assessment and disorder prediction.
- **Image Analysis:** Convolutional Neural Networks (CNNs) for analyzing thyroid ultrasound images and detecting abnormalities.
- **Natural Language Processing (NLP):** To extract relevant information from unstructured clinical notes and patient-reported data.

2.2 Decision Support System

- **Technologies:** Rule-based expert systems and AI decision support.
- **Features:**
  - **Clinical Decision Support:** Provides diagnostic recommendations based on input data and AI analysis.
  - **Treatment Suggestions:** Offers treatment options based on established clinical guidelines and patient-specific data.

C. User Interface

3.1 Patient Portal

- **Features:**
  - **Self-Assessment Tools:** Interactive questionnaires for symptom assessment.
  - **Test Result Access:** View and interpret lab results and imaging reports.
  - **Appointment Scheduling:** Schedule consultations with endocrinologists.

3.2 Healthcare Provider Dashboard

- **Features:**
  - **Data Management:** Access patient data, test results, and imaging.
  - **Diagnostic Tools:** View AI-generated risk assessments and diagnostic suggestions.
  - **Reporting:** Generate and export patient reports and treatment plans.

D. Monitoring and Feedback

4.1 Real-Time Monitoring

- **Technologies:** IoT (Internet of Things) for wearable devices and symptom tracking.
- **Features:** Continuous monitoring of patient symptoms and health metrics, with alerts for abnormal patterns.

3. Objectives of the Proposed System

1. Enhance Early Detection of Thyroid Disorders

Objective: To leverage advanced artificial intelligence (AI) algorithms to identify thyroid disorders at an early stage.

Explanation: The system employs machine learning models to analyze patient data, including lab results and symptoms, to predict the risk of conditions like hypothyroidism, hyperthyroidism, and thyroid cancer.

2. Automate Thyroid Ultrasound Image Analysis

Objective: To improve the accuracy and efficiency of thyroid ultrasound image interpretation through automated analysis.

Explanation: Convolutional Neural Networks (CNNs) are used to automatically detect and classify thyroid abnormalities in ultrasound images, such as nodules or cysts.

3. Integrate Multi-Source Clinical Data into a Unified Platform

Objective: To consolidate various sources of thyroid health data into a single, comprehensive platform for better management and analysis.

Explanation: The system integrates clinical lab results, imaging data, and home test results into one interface for seamless access and evaluation by healthcare professionals.

4. Provide Real-Time Monitoring of Thyroid-Related Symptoms

Objective: To offer tools for continuous, real-time monitoring of thyroid-related symptoms through wearable devices and mobile applications.

Explanation: Wearable technology and mobile apps are used to track symptoms such as fatigue, weight changes, and temperature fluctuations, which are then analyzed for trends and deviations.

5. Develop a User-Friendly Patient Portal for Health Management

Objective: To create an intuitive and accessible portal for patients to manage their thyroid health, including viewing test results and scheduling appointments.

Explanation: The patient portal provides a user-friendly interface for patients to access their health data, complete self-assessments, and manage their care.

CHAPTER 3

SYSTEM REQUIREMENTS SPECIFICATION

1. Hardware Requirements

The following are needed to efficiently use the application.

Processor	-	Intel Core i3 and above
Speed	-	GHz 2.5
RAM	-	8 GB (min)
Hard Disk	-	GB 50

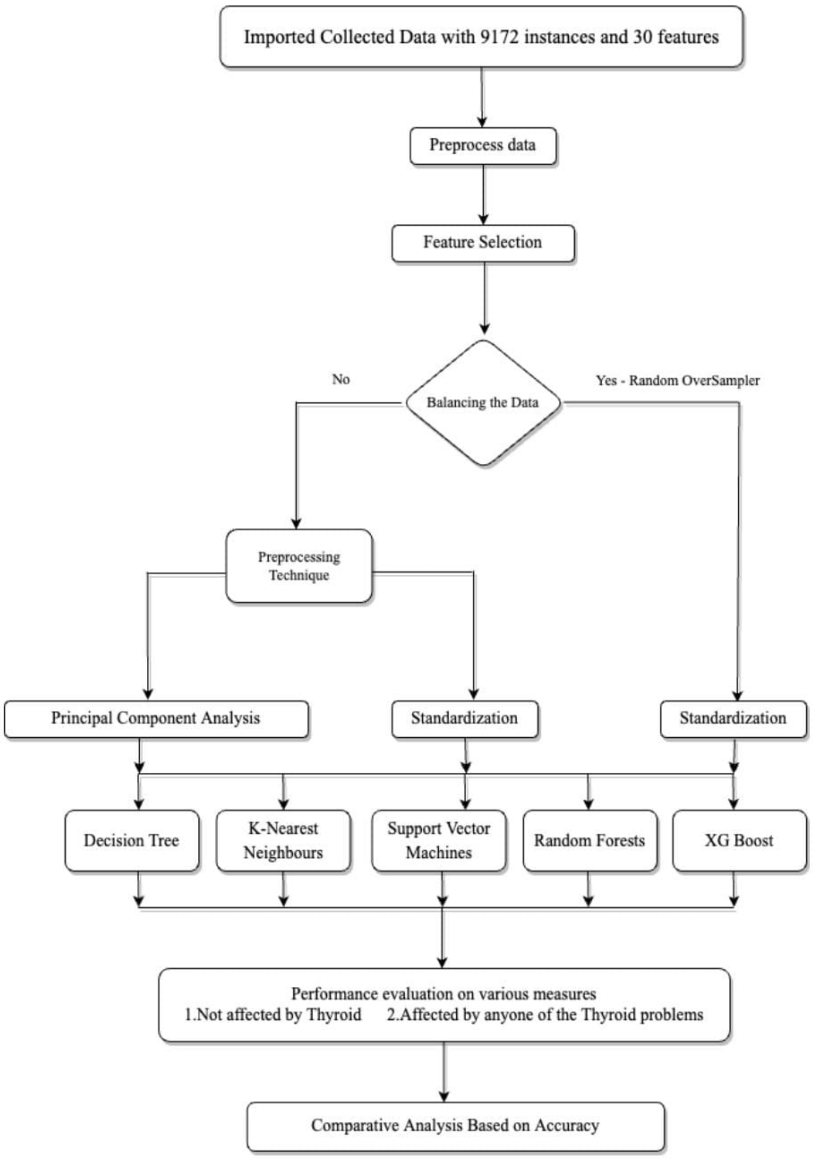
2. Software Requirements

Software requirements define software resource fundamentals that need to be installed on a workstation to provide optimum working of a software. The following are required for optimal development and usage of the application.

Operating System	-	and 7 Windows above
Programming Language	-	Python
Compiler	-	Google Colab

Chapter 4:

1. Flowchart of Proposed System



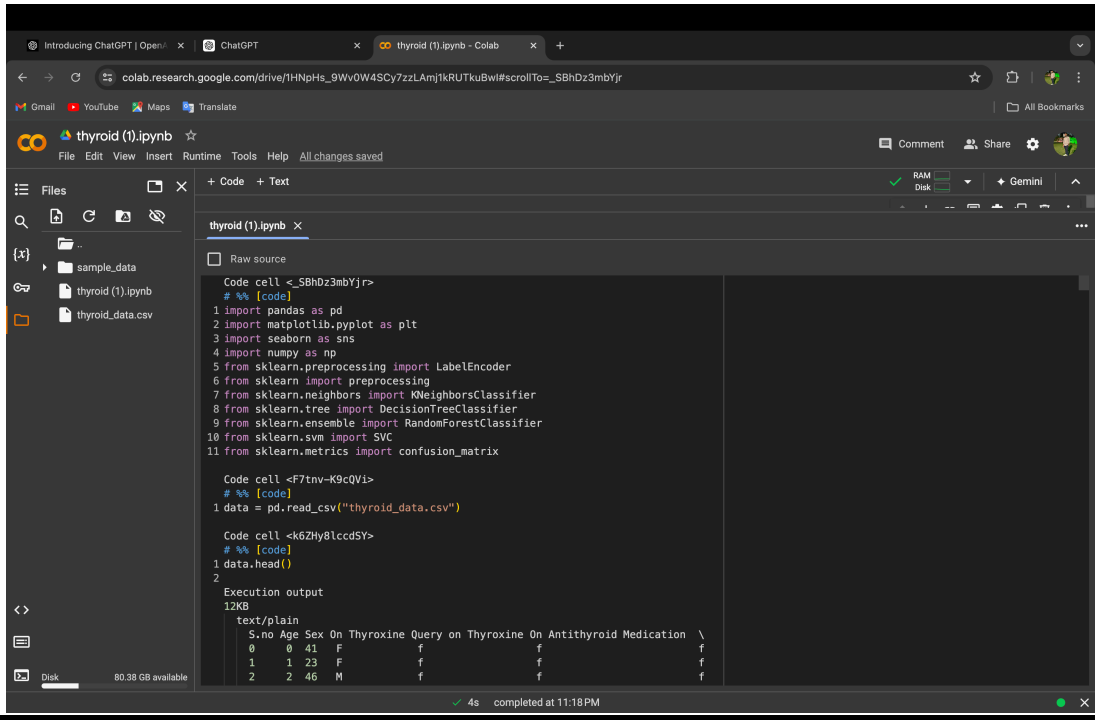
Methodology of our proposed thyroid Diagnosis

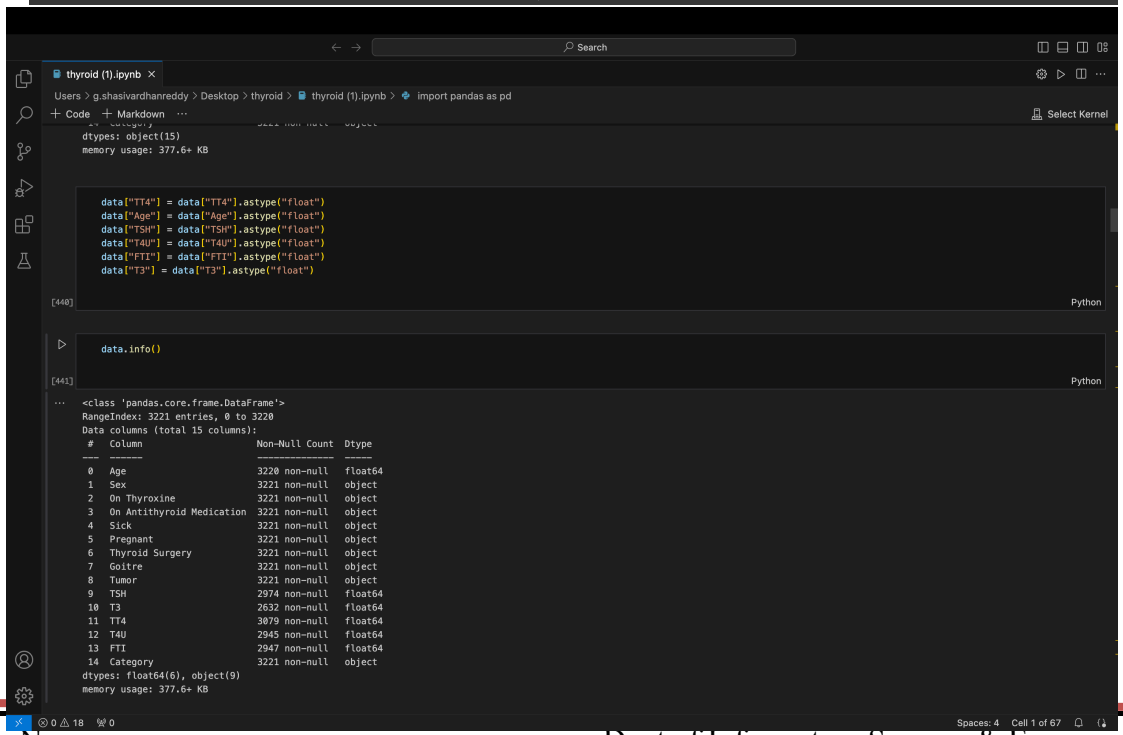
Fig Flowchart :4.2 of proposed system

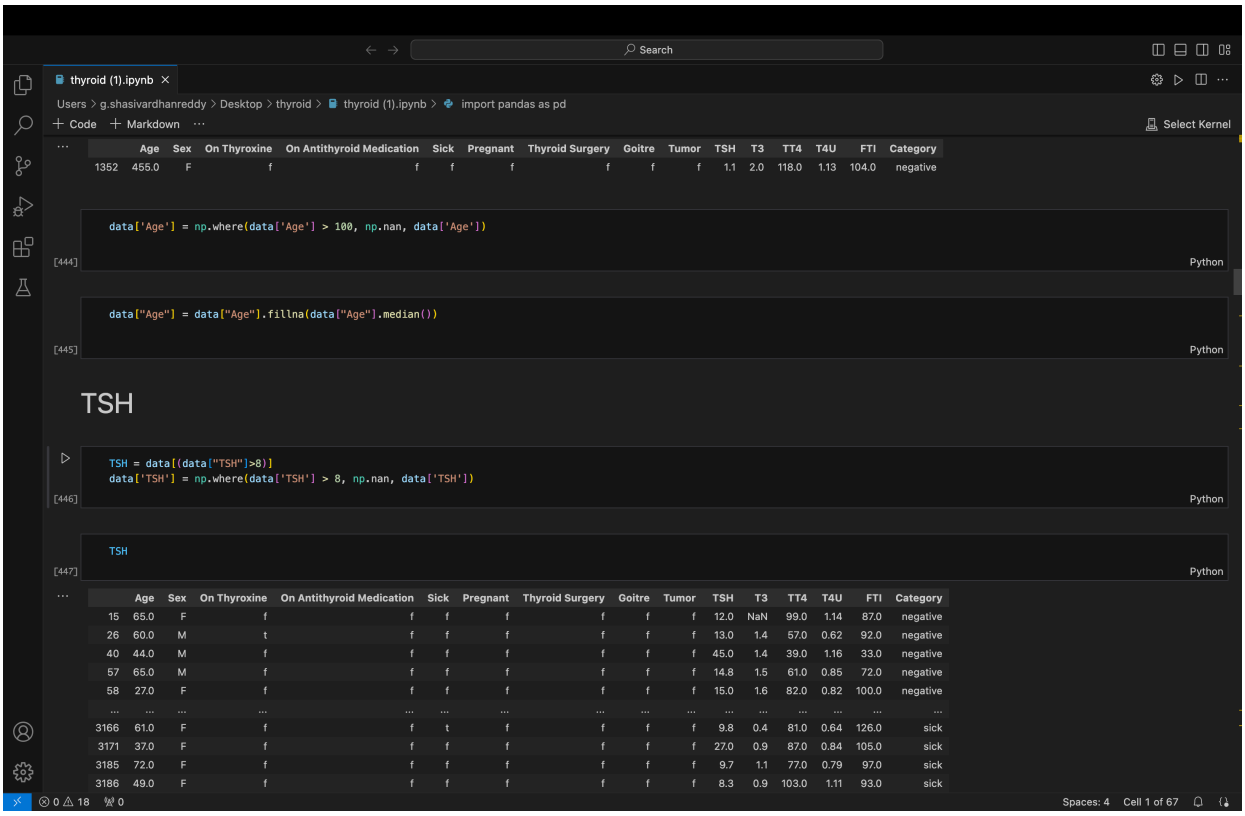
CHAPTER 5

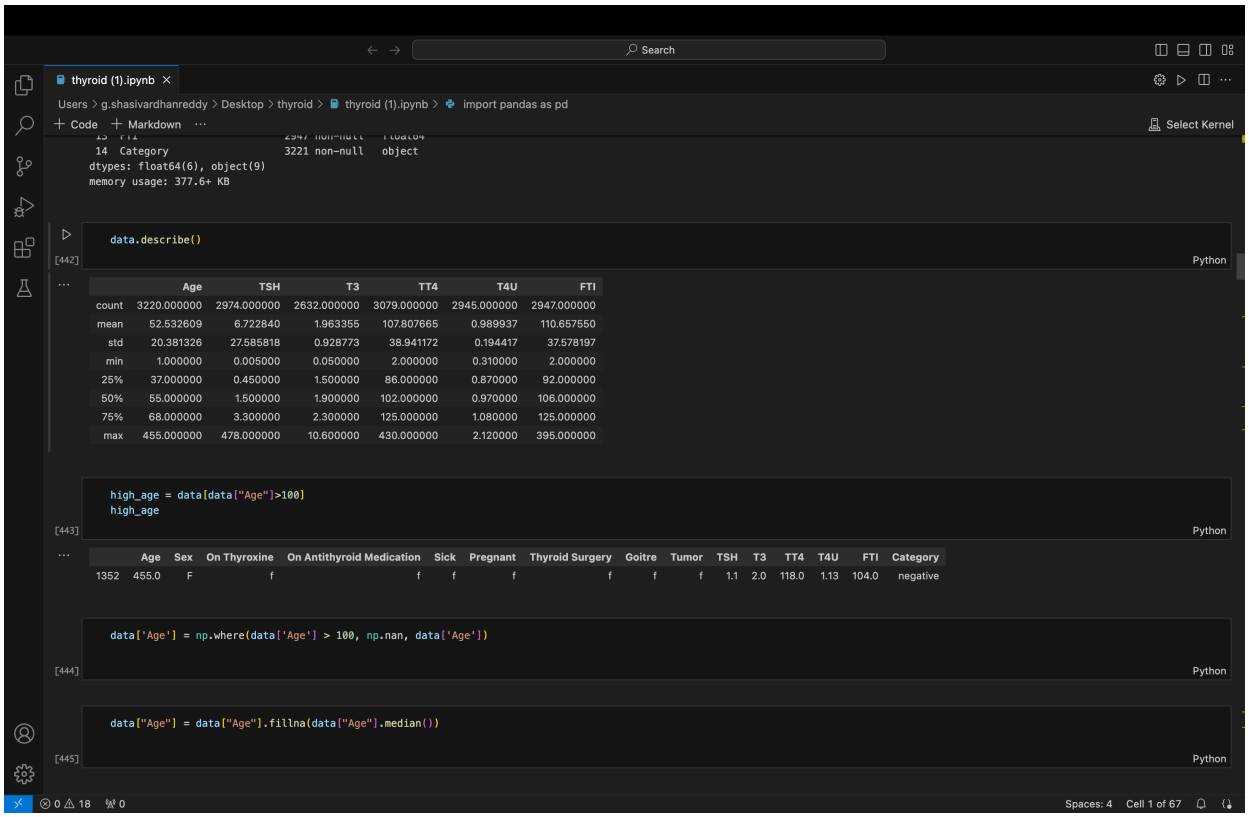
IMPLEMENTATION

1. Code Implementation





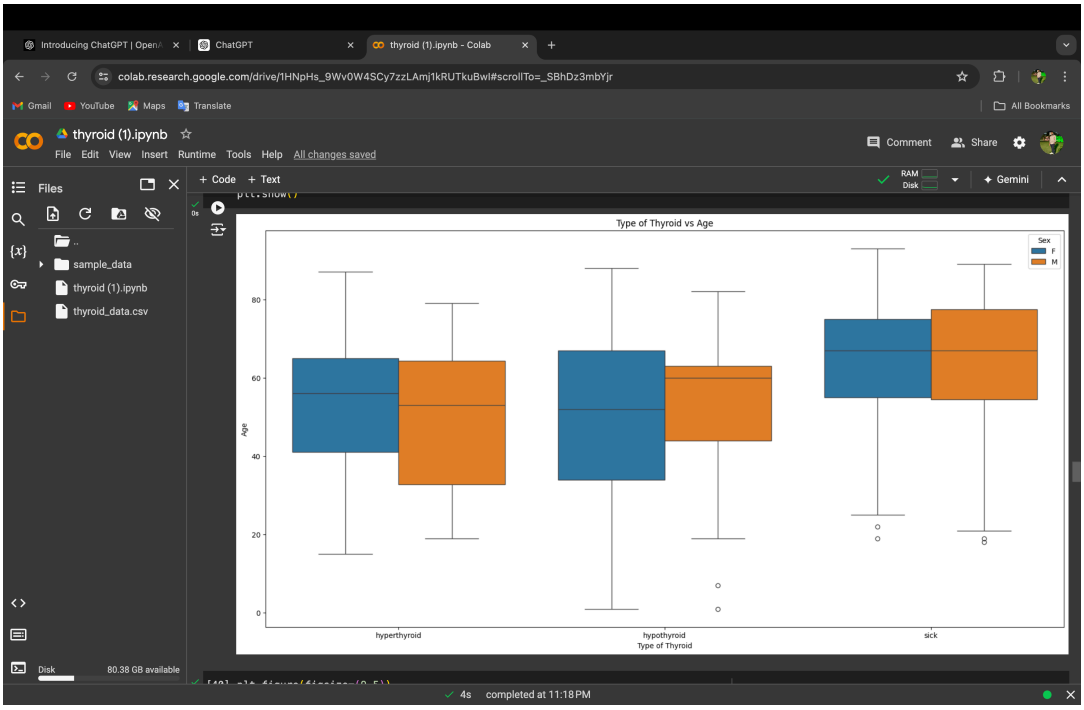

thyroid (1).ipynb
Users > g.shasivardhanreddy > Desktop > thyroid > thyroid (1).ipynb > import pandas as pd
+ Code + Markdown
Select Kernel
Age Sex On Thyroxine On Antithyroid Medication Sick Pregnant Thyroid Surgery Goitre Tumor TSH T3 TT4 T4U FTI Category
1352 455.0 F f f f f f f 1.1 2.0 118.0 1.13 104.0 negative
data['Age'] = np.where(data['Age'] > 100, np.nan, data['Age'])
[444] Python
data["Age"] = data["Age"].fillna(data["Age"].median())
[445] Python
TSH
TSH = data[(data["TSH"]>8)]
data['TSH'] = np.where(data['TSH'] > 8, np.nan, data['TSH'])
[446] Python
TSH
[447] Python
Age Sex On Thyroxine On Antithyroid Medication Sick Pregnant Thyroid Surgery Goitre Tumor TSH T3 TT4 T4U FTI Category
15 65.0 F f f f f f f 12.0 NaN 99.0 1.14 87.0 negative
26 60.0 M t f f f f f f 13.0 1.4 57.0 0.62 92.0 negative
40 44.0 M f f f f f f f 45.0 1.4 39.0 1.16 33.0 negative
57 65.0 M f f f f f f f 14.8 1.5 61.0 0.85 72.0 negative
58 27.0 F f f f f f f 15.0 1.6 82.0 0.82 100.0 negative
3166 61.0 F f f t f f f f 9.8 0.4 81.0 0.64 126.0 sick
3171 37.0 F f f f f f f f 27.0 0.9 87.0 0.84 105.0 sick
3185 72.0 F f f f f f f f 9.7 1.1 77.0 0.79 97.0 sick
3186 49.0 F f f f f f f f 8.3 0.9 103.0 1.11 93.0 sick
Spaces: 4 Cell 1 of 67

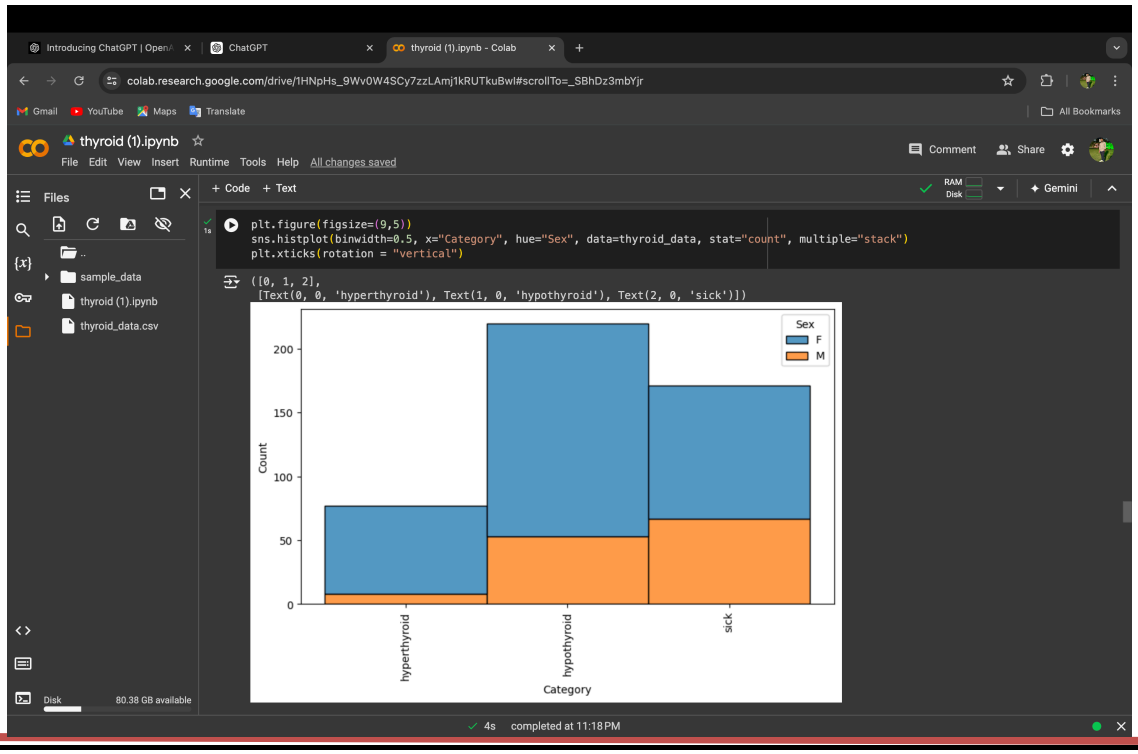

thyroid (1).ipynb
Users > g.shasivardhanreddy > Desktop > thyroid > thyroid (1).ipynb > import pandas as pd
+ Code + Markdown
Select Kernel
14 Category 3221 non-null object
dtypes: float64(6), object(9)
memory usage: 377.6+ KB
data.describe()
[442] Python
Age TSH T3 TT4 T4U FTI
count 3220.000000 2974.000000 2632.000000 3079.000000 2945.000000 2947.000000
mean 52.532609 6.722840 1.963355 107.807665 0.989937 110.657550
std 20.381326 27.585818 0.928773 38.941172 0.194417 37.578197
min 1.000000 0.005000 0.050000 2.000000 0.310000 2.000000
25% 37.000000 0.450000 1.500000 86.000000 0.870000 92.000000
50% 55.000000 1.500000 1.900000 102.000000 0.970000 106.000000
75% 68.000000 3.300000 2.300000 125.000000 1.080000 125.000000
max 455.000000 478.000000 10.600000 430.000000 2.120000 395.000000
high_age = data[data["Age"]>100]
high_age
[443] Python
Age Sex On Thyroxine On Antithyroid Medication Sick Pregnant Thyroid Surgery Goitre Tumor TSH T3 TT4 T4U FTI Category
1352 455.0 F f f f f f f 1.1 2.0 118.0 1.13 104.0 negative
data['Age'] = np.where(data['Age'] > 100, np.nan, data['Age'])
[444] Python
data["Age"] = data["Age"].fillna(data["Age"].median())
[445] Python
Spaces: 4 Cell 1 of 67

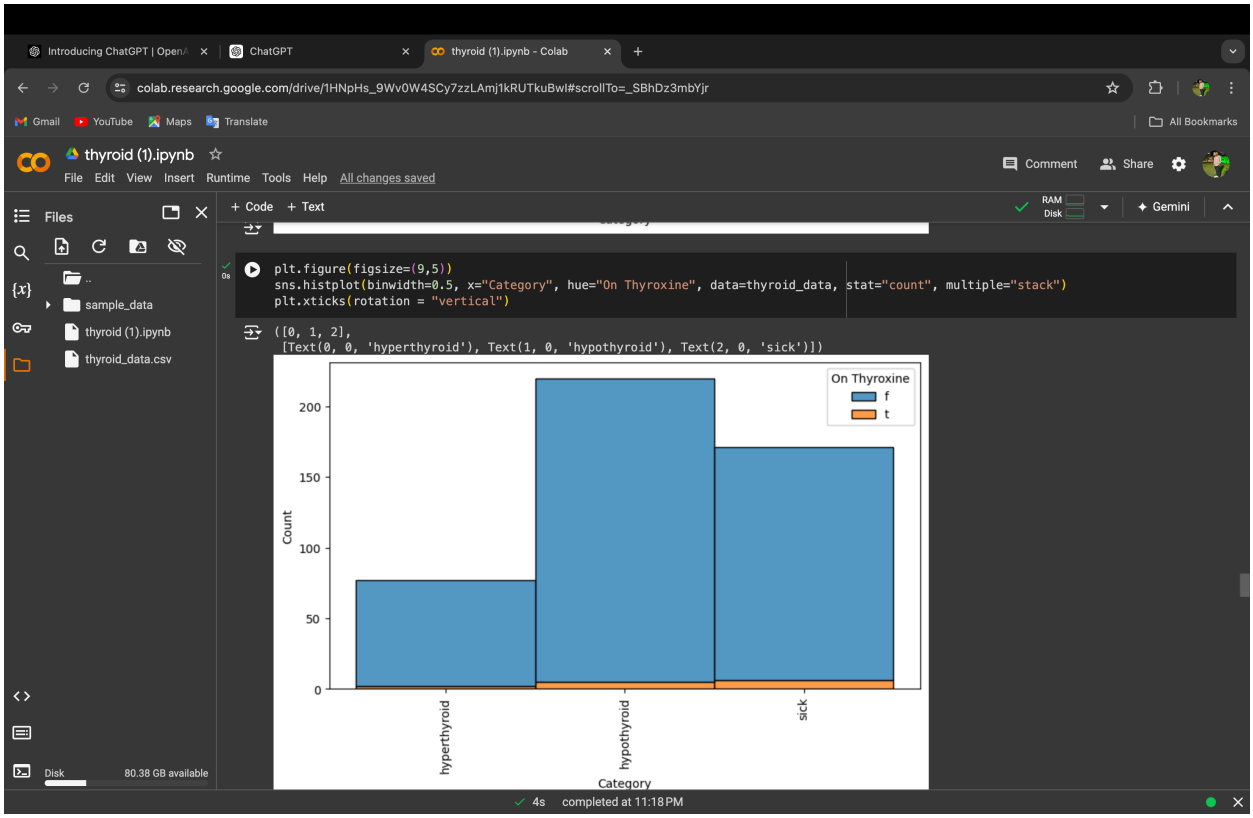
CHAPTER 6

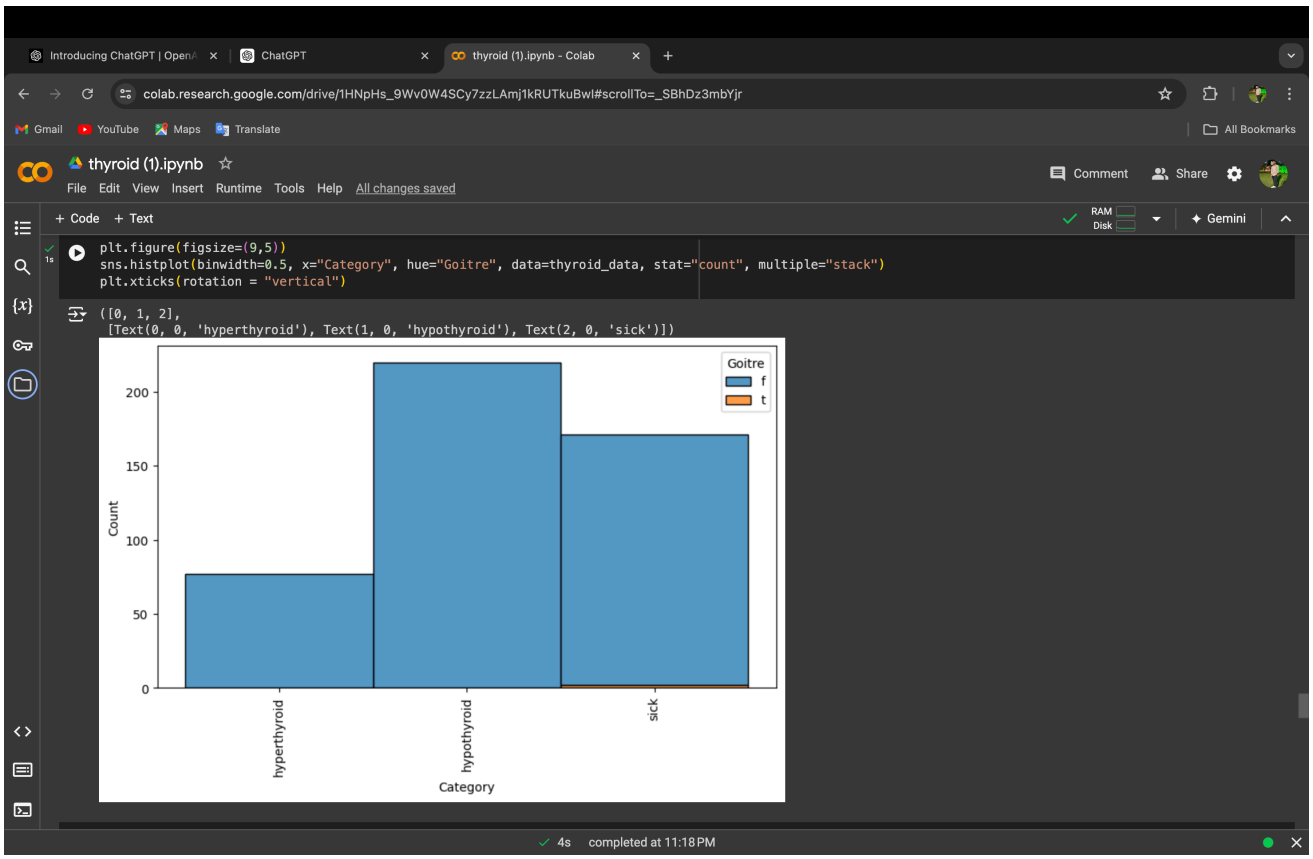
EXPERIMENTAL RESULTS

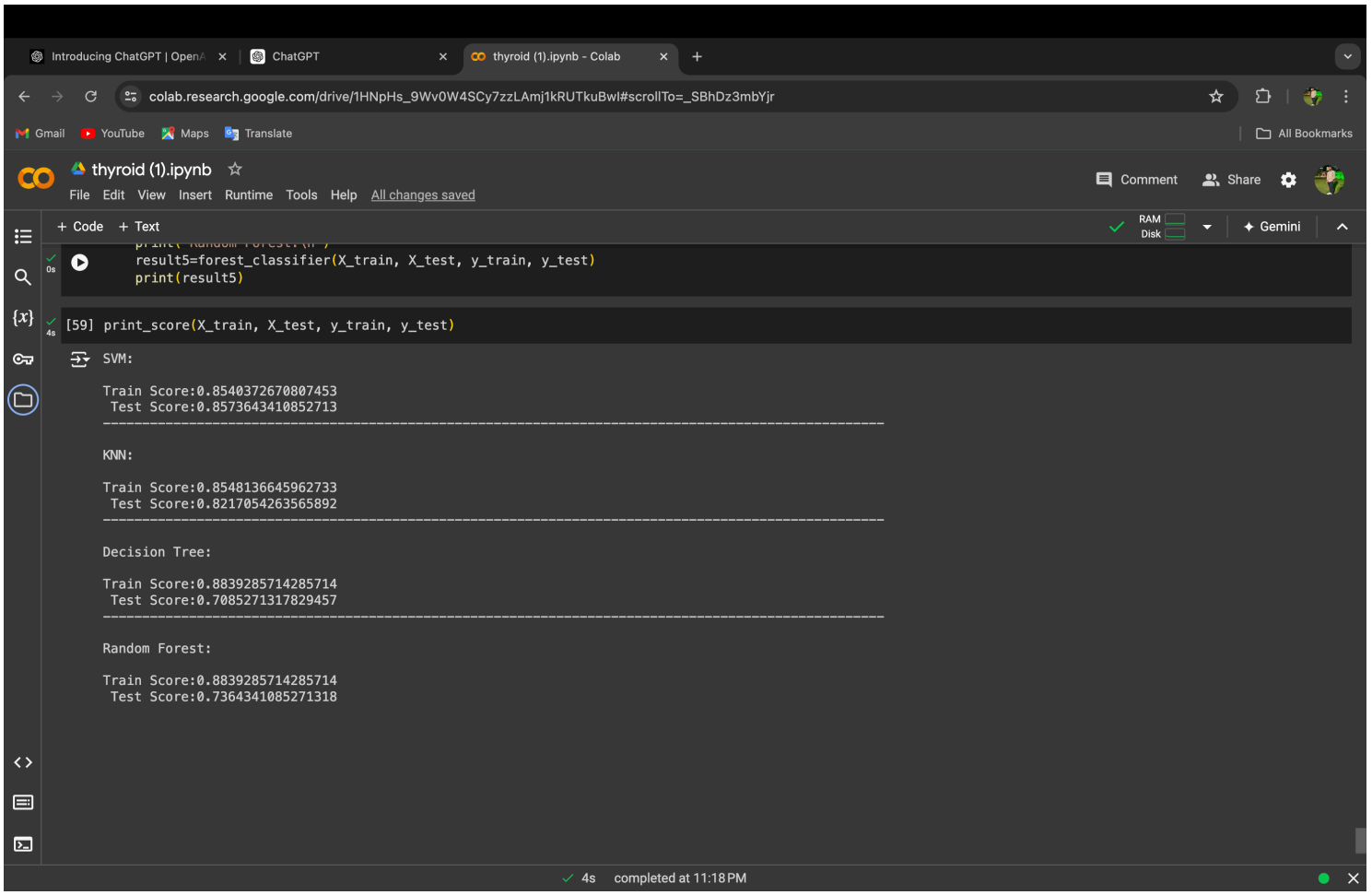
1. Outcome of Proposed System






thyroid (1).ipynb - Colab
colab.research.google.com/drive/1HNpHs_9Wv0W4SCy7zzLAmj1kRUTkuBwl#scrollTo=_SBhDz3mbYjr
thyroid (1).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
Gemini
Files
sample_data
thyroid (1).ipynb
thyroid_data.csv
Disk 80.38 GB available
plt.figure(figsize=(9,5))
sns.histplot(binwidth=0.5, x="Category", hue="On Thyroxine", data=thyroid_data, stat="count", multiple="stack")
plt.xticks(rotation = "vertical")
([0, 1, 2],
[Text(0, 0, 'hyperthyroid'), Text(1, 0, 'hypothyroid'), Text(2, 0, 'sick')])
On Thyroxine
f
t
Count
Category
hyperthyroid
hypothyroid
sick
4s completed at 11:18 PM


thyroid (1).ipynb - Colab
colab.research.google.com/drive/1HNpHs_9Wv0W4SCy7zzLAmj1kRUTkuBwl#scrollTo=_SBhDz3mbYjr
thyroid (1).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
Gemini
plt.figure(figsize=(9,5))
sns.histplot(binwidth=0.5, x="Category", hue="Goitre", data=thyroid_data, stat="count", multiple="stack")
plt.xticks(rotation = "vertical")
([0, 1, 2],
[Text(0, 0, 'hyperthyroid'), Text(1, 0, 'hypothyroid'), Text(2, 0, 'sick')])
Goitre
f
t
Count
Category
hyperthyroid
hypothyroid
sick
4s completed at 11:18 PM


Introducing ChatGPT | OpenA
ChatGPT
thyroid (1).ipynb - Colab
colab.research.google.com/drive/1HNpHs_9Wv0W4SCy7zzLAmj1kRUTkuBwl#scrollTo=_SBhDz3mbYjr
Gmail
YouTube
Maps
Translate
All Bookmarks
thyroid (1).ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Gemini
result5=forest_classifier(X_train, X_test, y_train, y_test)
print(result5)
[59] print_score(X_train, X_test, y_train, y_test)
SVM:
Train Score:0.8540372670807453
Test Score:0.8573643410852713
KNN:
Train Score:0.8548136645962733
Test Score:0.8217054263565892
Decision Tree:
Train Score:0.8839285714285714
Test Score:0.7085271317829457
Random Forest:
Train Score:0.8839285714285714
Test Score:0.7364341085271318
4s completed at 11:18 PM

CHAPTER
7

CONCLUSION AND FUTURE ENHANCEMENT

Final Thoughts

In this project We looked into different techniques for the thyroid classification methods like SVM,KNN, random forest and SVM tree in which resulted: -test 20% and train 80% train score :85%and test score:85%.From this we can say that desicion SVM tree is the good classification algorithm with accuracy 85%and SVM is the best classification algorithm for thyroid prediction

Thyroid Detection using Machine Learning is a project idea that aims a smart and precise way to predict thyroid disease. We have made use of logistic regression algorithm to train our dataset and to predict thyroid disease with more accuracy. Here the machine is trained to detect whether the person normal, hyper-hypo thyroid ism based on the user's input. So when user enters data in web app the data will be processed in backend (model) and the result will be displayed onthe screen. Our objective was to give society an efficient and precise way of machine learning which can be used in applications aiming to perform disease
detection.

Further development can be do by using image processing of ultrasonic scanning of thyroid images top redict thyroid nodules and cancer, which cannot be recognized in blood test report.By combining both the results, thyroid diseaseprediction can cover all thyroid related diseases.

REFERENCES

1. Research Papers
2. Web Links
3. Books