THYROID PREDICTION IN PREGNANT WOMEN USING KNN ALGORITHM

A Report submitted to the Rajiv Gandhi University of Knowledge and Technologies in partial fulfillment of the degree of

Bachelor of Technology in Computer Science and Engineering

Submitted by

G.Lavanya (S180024)

N.V.V.Ramani (S180188)

G.Jeevana (S180111)

3rd year BTech 2nd Semester

Under the supervision of

Dr. Y Kumar Sekhar

Asst. Professor-Department of CSE RGUKT Srikakulam



Department of Computer Science and Engineering

Rajiv Gandhi University of Knowledge and Technologies, Srikakulam

S.M. Puram (V), Etcherla (M), Srikakulam (Dt) – 532410



CERTIFICATE

This is to certify that the report entitled "THYROID PREDICTION IN PREGNANT WOMEN USING KNN ALGORITHM" was submitted by Lavanya.G, bearing ID.No. S180024, N.V.V.Ramani, bearing ID.No. S180188, Jeevana.G, bearing ID.No. S180111, in partial fulfilment of the requirements for the award of Bachelor of Technology in Computer Science is a Bonafide work carried out by them under my supervision and guidance.

The report has not been submitted previously in part or in full to this or any other University or Institution to award any degree or diploma.

Y Kumar Sekhar, Sesha Kumar Nalluri,

Project Guide, Head of the Department,

Department of CSE, Department of CSE,

RGUKT, SRIKAKULAM RGUKT, SRIKAKULAM

DECLARATION

We Lavanya Gunna, Ramani Nekkala and Jeevana Gummdi hereby declare that this report entitled "THYROID DETECTION IN PREGNANT WOMEN USING KNN ALGORITHM" submitted by us under the guidance and supervision of Y Kumar Sekhar is a Bonafede work.

We also declare that it has not been submitted previously in part or in full to this University or other University or Institution to award any degree or diploma.

(Lavanya Gunna)

I.D NO. S180024

(Ramani Nekkala)

I.D NO. S180188

(Jeevana Gummidi)

I.D NO. S180111

ACKNOWLEDGEMENTS

We would like to express my sincere gratitude to, my project Guide Y Kumar Sekhar, for valuable suggestions and keen interest throughout the progress of my course of research.

We are grateful to Sesha Kumar Nalluri, HOD CSE, for providing excellent computing facilities and a congenial atmosphere for progressing with my project.

At the outset, we would like to thank Rajiv Gandhi University of Knowledge and Technologies, Srikakulam for providing all the necessary resources for the successful completion of our course work. At last, but not least we thank our classmates and other students for their physical and moral support.

With Sincere Regards

Lavanya Gunna,

Ramani Nekkala,

Jeevana Gummdi.

ABSTRACT

The proposed system for thyroid detection using K-Nearest Neighbors (KNN) algorithm is designed to accurately detect early signs of thyroid dysfunction in expecting mothers. The system analyzes a large dataset of thyroid function test results and identifies patterns and relationships between different variables. Based on these patterns, the algorithm can predict the likelihood of an individual having thyroid dysfunction. This system also includes a user-friendly interface that allows healthcare professionals to input patient data and receive immediate results, thus saving time and improving accuracy. This revolutionary solution aims to address the shortcomings of the existing system by providing a more accurate and efficient way of detecting thyroid dysfunction in pregnant women. With the help of machine learning algorithms, healthcare professionals can provide better care and treatment to expecting mothers, ultimately leading to improved health outcomes for both the mother and baby.

Keywords: K-Nearest Neighbor, prediction, Machine learning (Supervised)

CONTENTS

Certificate2
Declaration
Abstract5
1. INTRODUCTION
1.1 Introduction8
1.2 Motivation8
1.3 Problem Statement8
1.4 Objectives8-9
1.5 Goal9
1.6 Scope10
1.7 Applications
1.8 Limitations
2. LITERATURE SURVEY
2.1 Create Dataset11
2.2 Study12
2.3 Summary12
3. EXISTING SYSTEM
3.1 Existing System
3.2Disadvantages13

4.1 Proposed System	14
4.2 Advantages	14
4.3 System Requirements	14
5. METHODS AND ALGORITHMS	
5.1 Algorithm	15
5.2Factors	15

9. REFERENCES.......27

4. PROPOSED SYSTEM

CHAPTER-1

INTRODUCTION

1.1 Introduction

Thyroid disorders during pregnancy can have serious consequences for both the mother and the baby. It is important to detect these disorders early on to prevent complications. Today, we will be discussing the role of the thyroid gland in pregnancy and how the KNN algorithm can be used for early prediction .Did you know that the thyroid gland is responsible for producing hormones that regulate metabolism? During pregnancy, the demand for these hormones increases significantly. However, sometimes the thyroid gland may not function properly, leading to thyroid disorders. In this presentation, we will explore the different types of thyroid disorders that can occur during pregnancy and their potential risks.

1.2 Motivation

Detecting thyroid disease in pregnant women is of utmost importance due to the potential impact it can have on both the mother and the developing fetus. Thyroid disorders during pregnancy, such as hypothyroidism or hyperthyroidism, can lead to various complications and adverse outcomes if left undiagnosed and untreated. Therefore, there is a need for an accurate and efficient method of thyroid disease detection in pregnant women.

1.3 Problem Statement

Thyroid disorders are relatively common during pregnancy, affecting up to 5% of all pregnancies. The most common thyroid disorders during pregnancy are hypothyroidism and hyperthyroidism. Hypothyroidism occurs when the thyroid gland does not produce enough thyroid hormone, while hyperthyroidism occurs when the thyroid gland produces too much thyroid hormone. Symptoms of hypothyroidism during pregnancy include fatigue, weight gain, constipation, and depression. Symptoms of hyperthyroidism during pregnancy include rapid heartbeat, nervousness, insomnia, and weight loss. If left untreated, thyroid disorders during pregnancy can lead to complications such as premature birth, low birth weight, and preeclampsia.

1.4 Objectives

The objectives of Thyroid prediction can be summarized as follows:

- 1. **Identify thyroid dysfunction:** The primary objective is to detect any abnormality or dysfunction in Identify thyroid dysfunction: The primary objective is to detect any abnormality or dysfunction in the thyroid gland during pregnancy. Thyroid dysfunction in pregnancy can have adverse effects on both the mother and the developing fetus.
- 2. **Ensure optimal thyroid function:** Maintaining proper thyroid function is crucial during pregnancy. Adequate thyroid hormone levels are necessary for normal fetal growth and development. By predicting thyroid dysfunction, appropriate interventions can be implemented to optimize thyroid function and minimize any potential risks to the mother and the baby.
- 3. **Prevent complications:** Thyroid dysfunction during pregnancy can lead to various complications, such as miscarriage, preterm birth, preeclampsia, and developmental issues in the baby. Early prediction of thyroid problems allows for timely management and treatment, reducing the likelihood of these complications.
- 4. **Guide treatment decisions:** Predicting thyroid dysfunction helps healthcare providers make informed decisions regarding treatment options. Depending on the specific condition and severity, treatment may involve medication, dietary modifications, or other interventions to normalize thyroid hormone levels and mitigate associated risks.
- 5. **Monitor thyroid function throughout pregnancy:** Pregnancy can have a significant impact on thyroid function, and the hormonal changes that occur during gestation can affect thyroid hormone levels. By predicting thyroid dysfunction, healthcare providers can monitor thyroid function regularly, ensuring that appropriate adjustments are made to the treatment plan if needed.
- 6. **Support maternal and fetal well-being:** The ultimate objective of predicting and managing thyroid dysfunction during pregnancy is to promote the overall well-being of both the mother and the baby. By addressing thyroid issues, healthcare providers aim to optimize maternal health, reduce the risk of complications, and support healthy fetal growth and development.

1.5 Goal

The goal of thyroid prediction in pregnant women based on symptoms is to recognize and identify the presence of thyroid dysfunction by assessing the characteristic signs and symptoms exhibited by the individual. While symptoms alone cannot provide a definitive diagnosis, they can serve as an initial screening tool to guide further evaluation and diagnostic testing.

1.6 Scope

The scope of thyroid prediction in pregnant women based on symptoms involves identifying and correlating reported symptoms to screen for potential thyroid dysfunction, triaging individuals for diagnostic testing, and monitoring thyroid function and symptoms throughout pregnancy.

1.7 Applications

- Early detection and intervention
- Screening tool
- Resource allocation
- Treatment planning
- Patient education and support

1.8 Limitations

- Overlapping symptoms: The symptoms associated with thyroid dysfunction can be nonspecific and overlap with other common pregnancy-related symptoms. This makes it challenging to attribute symptoms solely to thyroid dysfunction, leading to potential misinterpretation or misdiagnosis.
- 2. **Subjectivity and variability**: The reporting and interpretation of symptoms can be subjective and vary among individuals. Symptoms may be described differently or perceived with varying degrees of severity, which can introduce inconsistency and uncertainty in symptom-based prediction.

- 3. **Limited predictive accuracy:** Symptom-based prediction alone may not provide sufficient accuracy for diagnosing thyroid dysfunction. Diagnostic tests, including measuring thyroid hormone levels and thyroid autoantibodies, are necessary for confirmation and precise evaluation of thyroid function.
- 4. **Complexity of thyroid disorders:** Thyroid dysfunction is a complex condition with various underlying causes and presentations. Symptoms may differ based on the type and severity of dysfunction, making it challenging to rely solely on symptoms for accurate prediction.
- 5. **Individual variations:** Pregnant women may exhibit different patterns of symptoms or may not exhibit any symptoms despite having thyroid dysfunction. Some individuals may also experience fluctuations in symptoms over time, further complicating the prediction process.

CHAPTER-2

LITERATURE SURVEY

2.1 Collect Information

We went through several IEEE papers for knowing about the project and we collect the dataset from various health surveys and we went through Google searching and YouTube platforms for knowing more information about the project.

2.2 Study

Key Features are

- 1. Collect the dataset.
- 2. Analyze the thyroid dataset columns by preprocessing, stemming, tokenization.
- 3. Apply the KNN algorithm technique on whole dataset then divided the dataset into train and testing.
- 4. We can enter the symptoms of the person then it predicts whether the person have thyroid or not.

2.3 Summary

Analysis of Thyroid is a prediction model developed for pregnant women. This prediction model facilitates to predict the thyroid with early signs. This model evaluates symptoms of person based on a given dataset to the model.

CHAPTER 3

EXISTING SYSTEM

3.1 Existing System

The Existing System is based on the all blood test reports for thyroid test.

That report includes:

- 1. T3 measured
- 2. TSH measured
- 3. T3 measured
- 4. TSH measured
- 5. TT4 measured
- 6. T4U measured
- 7. FTI measured
- 8. TBG measured

3.2 Disadvantages

- Doesn't provide any facilities for rural areas.
- Accuracy is low
- It is necessary to take blood test

CHAPTER-4

PROPOSED SYSTEM

4.1 Proposed System

In the proposed system we take the early symptoms of the patient without using blood test reports

Instead of taking the blood reports we take the visible early symptoms of the person

4.2 Advantages

- Less Cost
- Can take early precaution through this prediction
- Can be useful for rural areas

4.3 System Requirements

- Python IDLE/Google Collaboratory.
- ML modules like numpy, pandas, matplotlib, seaborn.
- Windows, MAC OS.

Hardware requirements:

1. RAM: 4GB or more.

CHAPTER-5

METHODS AND ALGORITHMS

5.1 K-Nearest Neighbor

Implementing thyroid prediction using the k-Nearest Neighbors (k-NN) algorithm involves several steps. Here's a high-level overview of the process:

- 1. **Data Collection:** Gather a dataset that includes features (also known as predictors) and target values (the thyroid condition). The features could include patient characteristics like age, fatigue, family history etc... While the target values indicate whether the patient has a thyroid condition or not.
- 2. **Data Preprocessing**: Clean and preprocess the dataset to ensure data quality. This step involves handling missing values, removing irrelevant features, and performing any necessary data transformations (e.g., normalization, standardization) to make the data suitable for the k-NN algorithm.
- 3. **Splitting the Dataset**: Divide the dataset into two subsets: a training set and a test set. The training set will be used to train the k-NN model, while the test set will be used to evaluate its performance.
- 4. **Feature Selection:** If needed, perform feature selection techniques to identify the most relevant features for the thyroid prediction task. This step helps to improve the model's performance and reduce computational complexity.
- 5. Choosing the Value of k: Determine the value of k, which represents the number of neighbors to consider when making predictions. You can choose this value based on experimentation or using techniques like cross-validation.
- 6. **Training the k-NN Model:** Use the training set to train the k-NN model. The k-NN algorithm does not involve a traditional training process; instead, it stores the training data in memory to be used for prediction.

- 7. **Predicting Thyroid Condition:** Apply the trained k-NN model on the test set to predict the thyroid condition for each test instance. The prediction is based on the majority class of the k nearest neighbors.
- **8. Evaluating Model Performance:** Assess the performance of the k-NN model using evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics help determine how well the model predicts the thyroid condition.
- **9. Hyperparameter Tuning:** Fine-tune the hyperparameters of the k-NN algorithm, such as the distance metric used, to optimize the model's performance. This can be done using techniques like grid search or random search.
- **10. Model Deployment:** Once satisfied with the performance, deploy the trained k-NN model to make predictions on new, unseen data. Ensure that the model is integrated into the appropriate application or system for real-world usage.

5.2 Factors

Factors include Age, Fatigue, Family History, and Sensitivity to cold, Hair Loss, Weight Changes, and Heart Rate

CHAPTER-6

EXPERIMENT RESULTS

The results of thyroid prediction using the k-Nearest Neighbors (k-NN) algorithm can vary depending on the dataset, feature selection, hyper parameter tuning, and the evaluation metrics used. Based on the symptoms of the patient the dataset divided into 2-groups, one is that the persons have thyroid, another is that the persons not having thyroid. Here we can predict the thyroid disease by using early symptoms. It can be useful for rural areas.

CHAPTER-7

SOURCE CODE

STEP-1 IMPORTING LIBRARIES

```
In [1]: import pandas as pd
   import numpy as np
   import matplotlib.pyplot as plt
   import seaborn as sns
   import warnings
   warnings.filterwarnings("ignore")
```

STEP-2 LOADING THE DATASET

```
In [2]: df=pd.read_csv('Thyroid.csv')
    df
```

STEP-3 DROPPING THE DUPLICATES

```
In [4]: df=df.drop_duplicates()
In [5]: df
```

STEP-4 CHECKING THE NULL VALUES AND DATA TYPE OF RECORDS

```
In [7]: df.isnull().sum()
In [6]: df.dtypes
```

STEP-5 PLOTTING THE DATA

```
In [8]: sns.countplot('Thyroid Disease',data=df)
In [9]: df.hist(bins=10,figsize=(10,10))
   plt.show()
```

STEP-6 SPLITTING THE DATA INTO TRAINING AND TESTING

```
In [10]: x=df.iloc[:,:-1].values
    y=df.iloc[:,-1].values

In [11]: from sklearn.model_selection import train_test_split
    x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
    x_test

In [12]: from sklearn.preprocessing import StandardScaler
    sc=StandardScaler()
    x_train=sc.fit_transform(x_train)
    x_test=sc.fit_transform(x_test)

In [13]: from sklearn.neighbors import KNeighborsClassifier
    classifier=KNeighborsClassifier(n_neighbors=11,metric='euclidean',p=2)
    classifier.fit(x_train,y_train)

In [14]: predictions = classifier.predict(x_test)
```

STEP-7 CHECKING THE ACCURACY SCORE

In [15]: from sklearn.metrics import accuracy_score
 accuracy_score(y_test, predictions)

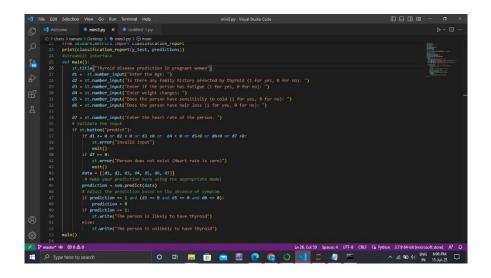
STEP-8 CLASSIFICATION REPORT

In [16]: from sklearn.metrics import classification_report
 print(classification_report(y_test, predictions))

STEP-9 TESTING THE MODEL

```
In [18]: d1 = float(input("Enter the Age: "))
         d2 = float(input("Is there any family history affected by thyroid (1 for yes, 0 for no): "))
         d3 = float(input("Enter if the person has fatigue (1 for yes, 0 for no): "))
         d4 = float(input("Enter weight changes: "))
         d5 = float(input("Does the person have sensitivity to cold (1 for yes, 0 for no): "))
         d6 = float(input("Does the person have hair loss (1 for yes, 0 for no): "))
         d7 = float(input("Enter the heart rate of the person: "))
         # Validate the input
         if d1 <= 0 or d4 < 0 or d7 <= 0:
             print("Invalid input")
             exit()
         if d7 == 0:
             print("Person does not exist (Heart rate is zero)")
             exit()
         data = [[d1, d2, d3, d4, d5, d6, d7]]
         # Make your prediction here using the appropriate model
         prediction = classifier.predict(data)
         # Adjust the prediction based on the absence of symptoms
         if prediction == 1 and (d3 == 0 and d5 == 0 and d6 == 0):
             prediction = 0
         if prediction == 1:
             print("The person is likely to have thyroid")
             print("The person is unlikely to have thyroid")
```

STEP-10 CREATING THE INTERFACE



Outputs:

Step-2

Out[2]:

	Age	Family History	Fatigue	Weight Changes	Sensitivity to Cold	Hair Loss	Heart Rate	Thyroid Disease
0	28	0	0	0	0	0	80	0
1	32	1	1	1	1	1	90	1
2	35	0	0	0	0	0	75	0
3	30	1	1	1	0	0	85	1
4	33	1	1	1	1	1	90	1
1166	33	1	1	1	1	1	80	1
1167	30	0	1	1	1	1	85	1
1168	35	0	1	1	1	1	82	1
1169	28	1	0	0	0	0	78	0
						-		

Step-3

Out[5]:

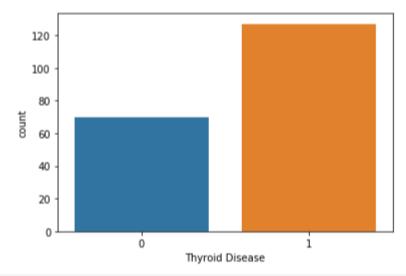
	Age	Family History	Fatigue	Weight Changes	Sensitivity to Cold	Hair Loss	Heart Rate	Thyroid Disease
0	28	0	0	0	0	0	80	0
1	32	1	1	1	1	1	90	1
2	35	0	0	0	0	0	75	0
3	30	1	1	1	0	0	85	1
4	33	1	1	1	1	1	90	1
702	31	0	1	0	0	0	88	1
742	28	1	0	0	0	0	78	1
842	31	1	0	0	0	0	88	1
940	31	0	0	1	1	1	88	1
1100	34	0	1	0	1	0	87	1

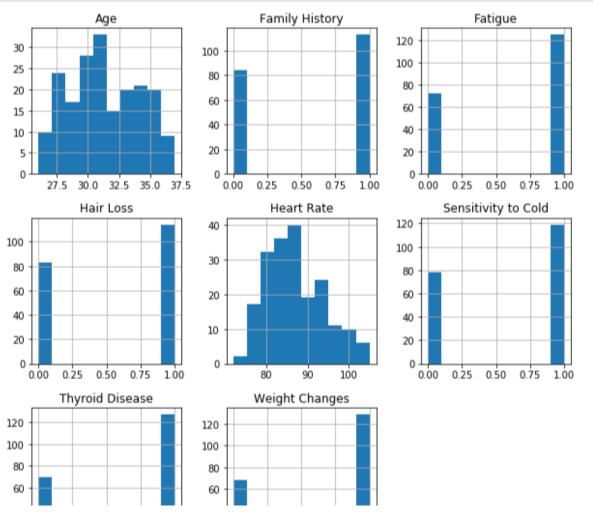
STEP-4

Out[6]: Age int64 Family History int64 Fatigue int64 Weight Changes int64 Sensitivity to Cold int64 Hair Loss int64 Heart Rate int64 Thyroid Disease int64 dtype: object

22 | Page

STEP-5
<matplotlib.axes._subplots.AxesSubplot at 0x1bdfdcdf508>





STEP-6

```
Out[11]: array([[ 27,
                                 0,
                                      1, 105],
                    1, 1,
                             0,
                             0,
              27,
                    1, 1,
                                 0,
                                      0, 79],
              [ 28,
                                        78],
                    1,
                         1,
                             0,
                                 0,
                                      0,
              [ 30,
                                     1, 80],
                    1, 1,
                             1,
                                 1,
                                      1, 81],
              [ 36,
                     0,
                       1,
                             1,
                                 1,
              [ 33,
                       1,
                             1,
                                     1, 93],
                    1,
                                 1,
              29,
                    1,
                        1,
                             0,
                                 0,
                                      1, 96],
                         0,
                                 0,
              [ 30,
                                      1, 85],
                    1,
                             1,
              28,
                    1, 1,
                            1,
                                 1,
                                     1, 86],
              [ 35,
                    1,
                       1,
                             1,
                                      1, 85],
                                 1,
              29,
                    0, 0,
                             0,
                                 0, 0, 81],
                             0,
                                 0,
              27,
                    0, 0,
                                     0, 82],
                                     0, 79],
              [ 31,
                    1, 0,
                            0,
                                 0,
                            1,
              [ 33,
                    1, 1,
                                 1, 1, 80],
              [ 31,
                    1,
                       0,
                             0,
                                 0,
                                     0, 88],
              [ 30,
                    1, 1,
                            1,
                                 1, 1, 85],
              [ 34,
                                      1, 95],
                    1,
                        1,
                            1,
                                 1,
              [ 30,
                    0, 0,
                            0,
                                 0,
                                     0, 76],
              [ 30,
                    1, 1,
                            0,
                                 0,
                                     1, 105],
                                 0,
              [ 28,
                    0, 1,
                                     0, 78],
                             1,
              [ 31,
                    0, 1,
                            1,
                                 1, 1, 88],
              [ 29,
                                     0, 81],
                    1,
                        1,
                             1,
                                 0,
              [ 35,
                    1, 1,
                            1,
                                 1, 1, 89],
              [ 31,
                             0,
                                     1, 96],
                    0, 1,
                                 0,
              [ 32,
                    0, 1,
                            1,
                                 0, 1, 78],
              27,
                    1, 1,
                            0,
                                 0, 0, 82],
                            1,
              [ 32,
                    1, 1,
                                 1, 1, 84],
              [ 34,
                    1, 1,
                            1, 1, 1, 90],
              34,
                            0,
                                     1, 99],
                    0, 1,
                                 1,
              [ 31,
                    0, 0,
                             0, 1,
                                     1, 96],
              [ 32,
                    1,
                       1,
                             1,
                                1,
                                      1, 94],
```

Out[13]: KNeighborsClassifier(metric='euclidean', n_neighbors=11)

STEP-7

Out[15]: 0.825

STEP-8

	precision	recall	f1-score	support
0	0.80	0.75	0.77	16
1	0.84	0.88	0.86	24
accuracy			0.82	40
macro avg	0.82	0.81	0.82	40
weighted avg	0.82	0.82	0.82	40

STEP-9

```
Enter the Age: 35
Is there any family history affected by thyroid (1 for yes, 0 for no): 0
Enter if the person has fatigue (1 for yes, 0 for no): 0
Enter weight changes: 0
Does the person have sensitivity to cold (1 for yes, 0 for no): 0
Does the person have hair loss (1 for yes, 0 for no): 0
Enter the heart rate of the person: 75
The person is unlikely to have thyroid

Enter the Age: 35
Is there any family history affected by thyroid (1 for yes, 0 for no): 1
Enter if the person has fatigue (1 for yes, 0 for no): 1
```

Enter weight changes: 1

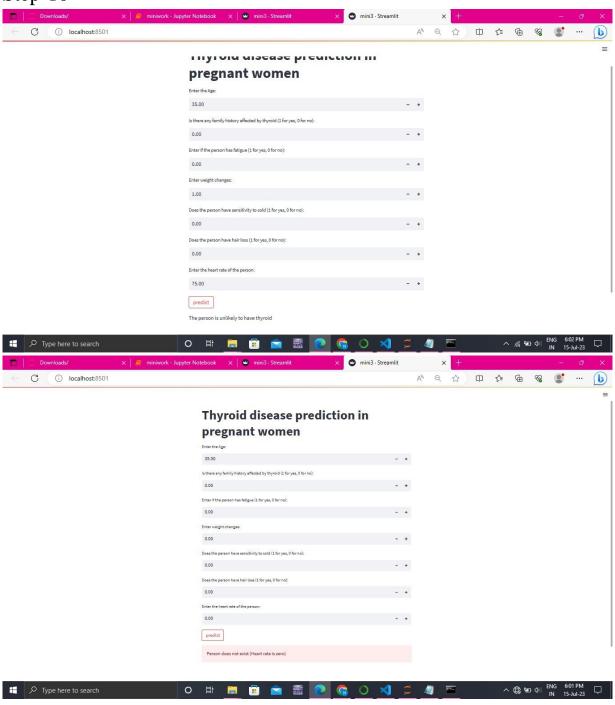
Does the person have sensitivity to cold (1 for yes, 0 for no): 1

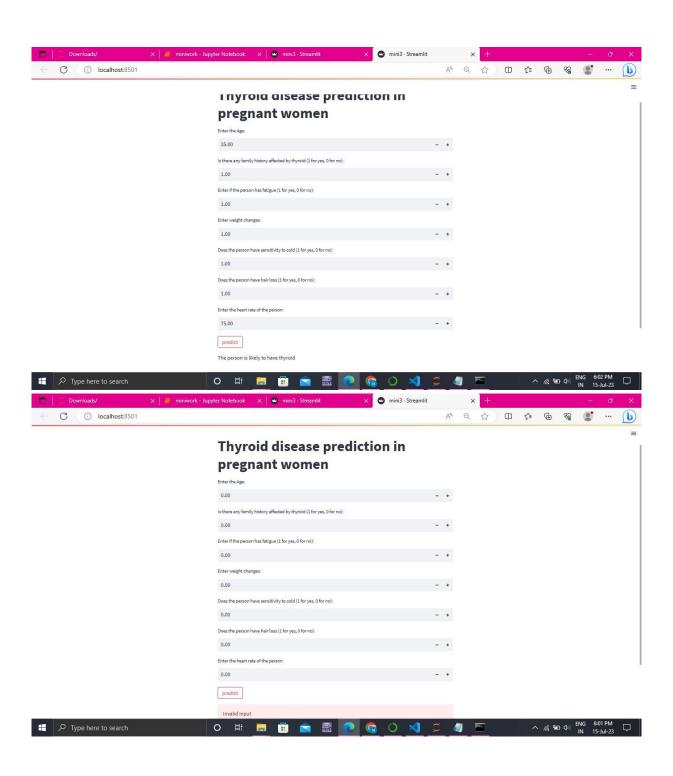
Does the person have hair loss (1 for yes, 0 for no): 1

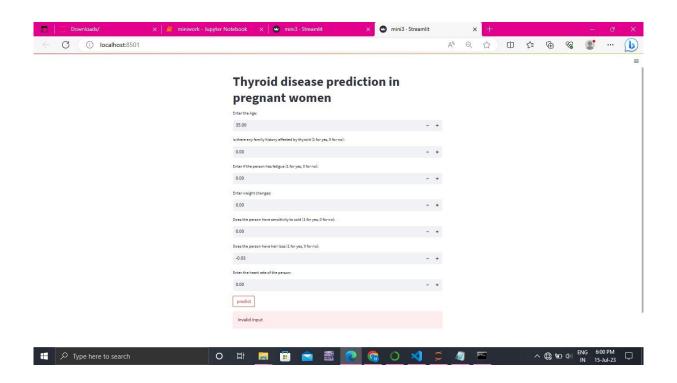
Enter the heart rate of the person: 75

Enter the heart rate of the person: 75 The person is likely to have thyroid

Step-10







CHAPTER-8

CONCLUSION

- 1. In today's highly payable treatment our prediction is very useful for people who cannot afford much money for thyroid tests
- 2. It can predict the results to take early precautions, it can help full to both mother and baby throughout the pregnancy
- 3. They can take early precautions during pregnancy

CHAPTER-9

REFERENCES

- 1. https://ieeexplore.ieee.org/document/8745910
- 2. https://ieeexplore.ieee.org/document/9811472