iNeuron.ai

Credit Card Default Prediction

Low Level Design (LLD) Documentation

Shantanu Kukkar

25-08-2024

Document Version Control

Date Issued	Version	Description	Author
25/08/2024	1	Initial LLD – V1.0	Shantanu Kukkar

Table of Contents

1. Introduction

1.1. Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the credit card default system. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, this document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

1.2. Scope

This software system will be a web application, and this system will be designed to predict whether the credit card holder will default the payment in the upcoming month or not.

1.3. Risk

Document specific risks that have been identified or that should be considered.

1.4. Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

2. Technical Specifications

2.1. Dataset Information

The data required for the Thyroid Disease Detection solution is crucial to address the problem statement effectively. We need information from individuals who have previously undergone thyroid blood tests to determine whether they are afflicted with thyroid disease and, if so, what specific type of thyroid condition they have. The following attributes are essential for our data collection:

1. Age: Age is an important factor as thyroid conditions may be more prevalent in individuals older than 60, particularly in women.
2. Gender: Gender is significant, as women are statistically more prone to being diagnosed with thyroid disorders compared to men.
3. Current Thyroxine Treatment: We need to know if individuals are already undergoing thyroxine treatment for their thyroid condition.
4. Current Anti-Thyroid Medication: The information on whether individuals are currently taking anti-thyroid medications is essential.
5. Pregnancy (for females): Pregnancy status is crucial, as postpartum thyroiditis can occur in a percentage of women after childbirth.
6. Health Condition during Diagnosis: We need to record whether individuals were sick or unwell at the time of diagnosis.
7. Iodine Test: Both excess and insufficient iodine levels can lead to thyroid disorders.
8. Lithium Test: Since lithium affects thyroidal iodine uptake, this test is significant.
9. Goitre Test: The presence of goitre indicates a potential issue with thyroid hormone production (hyperthyroidism).
10. Tumour Test: Identifying thyroid cancer requires analyzing genetic changes in thyroid cells.
11. TSH Level Measurement: TSH level monitoring helps assess thyroid gland function. The normal TSH range for adults is 0.40 - 4.50 mIU/mL.
12. T3 Level Measurement: T3 is a thyroid hormone that should be within the normal range.
13. T4 Level Measurement: T4 levels are relevant for diagnosing hypothyroidism (low T4) or hyperthyroidism (high T4). The normal T4 range for adults is 5.0 – 11.0 ug/dL.
14. FTI (Free T4 or Free Thyroxine Index): The FTI is calculated by multiplying Total T4 and T3 Uptake. It helps diagnose thyroid disorders. The normal FT3 range is 2.3 - 4.1 pg/mL.
15. Thyroxine-binding globulin (TBG): The TBG blood test measures the level of a protein that moves thyroid hormone throughout your body.

Collecting data on these attributes will enable the development of a robust Thyroid Disease Detection model, offering valuable insights for effective diagnosis and treatment decisions.

age	sex	on_thyro	query_or	on_antith	sick	pregnant	thyroid_s	I131_trea	query_hy	query_hy	lithium	goitre	tumor	hypopitui	psych	TSH_mea	TSH	T3_meas	T3	TT4_mea	TT4	T4U_mea	T4U	FTI_meas	FTI	TBG_mea	TBG	referral_	Class
41	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	1.3	t	2.5	t	125	t	1.14	t	109	f	?	SVHC	negative
23	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	4.1	t	2	t	102	f	?	f	?	f	?	other	negative
46	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.98	f	?	t	109	t	0.91	t	120	f	?	other	negative
70	F	t	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.16	t	1.9	t	175	f	?	f	?	f	?	other	negative
70	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.72	t	1.2	t	61	t	0.87	t	70	f	?	SVI	negative
18	F	t	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.03	f	?	t	183	t	1.3	t	141	f	?	other	negative
59	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	?	f	?	t	72	t	0.92	t	78	f	?	other	negative
80	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	2.2	t	0.6	t	80	t	0.7	t	115	f	?	SVI	negative
66	F	f	f	f	f	f	f	f	f	f	f	f	t	f	f	t	0.6	t	2.2	t	123	t	0.93	t	132	f	?	SVI	negative
68	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	2.4	t	1.6	t	83	t	0.89	t	93	f	?	SVI	negative
84	F	f	f	f	f	f	f	f	f	f	f	f	t	f	f	t	1.1	t	2.2	t	115	t	0.95	t	121	f	?	SVI	negative
67	F	t	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.03	f	?	t	152	t	0.99	t	153	f	?	other	negative
71	F	f	f	f	t	f	f	f	f	t	f	f	f	f	f	t	0.03	t	3.8	t	171	t	1.13	t	151	f	?	other	negative
59	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	2.8	t	1.7	t	97	t	0.91	t	107	f	?	SVI	negative
28	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	3.3	t	1.8	t	109	t	0.91	t	119	f	?	SVHC	negative
65	F	f	f	f	f	f	f	f	t	f	f	f	f	f	f	t	12	f	?	t	99	t	1.14	t	87	f	?	other	compensated_hypothyroid
42	?	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	1.2	t	1.8	t	70	t	0.86	t	81	f	?	other	negative
63	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	1.5	t	1.2	t	117	t	0.96	t	121	f	?	SVI	negative
80	F	f	f	f	f	f	f	f	f	f	f	f	f	f	t	t	6	t	1.6	t	99	t	0.95	t	104	f	?	SVI	negative
28	M	f	f	f	f	f	f	f	f	f	f	f	f	f	t	t	2.1	t	2.6	t	121	t	0.94	t	130	f	?	SVHC	negative
51	F	t	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.1	f	?	t	130	t	0.86	t	151	f	?	other	negative
46	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.8	t	2.1	t	108	t	0.91	t	119	f	?	other	negative
81	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	1.9	t	0.3	t	102	t	0.96	t	106	f	?	SVI	negative
54	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	3.1	f	?	t	104	t	0.9	t	116	f	?	other	negative
55	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.2	t	1.8	t	134	t	1.02	t	131	f	?	SVI	negative
63	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.03	t	5.5	t	199	t	1.05	t	190	f	?	other	negative
60	M	t	f	f	f	f	f	f	f	f	f	f	f	f	f	t	13	t	1.4	t	57	t	0.62	t	92	f	?	other	negative
25	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.3	t	3.1	t	129	f	?	f	?	f	?	other	negative
73	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	1.9	t	1.5	t	113	t	1.06	t	106	f	?	SVI	negative
67	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	?	f	?	t	97	t	0.95	t	102	f	?	other	negative
34	F	f	f	f	f	t	f	f	f	t	f	f	f	f	f	t	0.035	t	2.5	t	119	t	1.55	t	76	f	?	other	negative
70	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	2.5	t	2.3	t	84	t	0.92	t	92	f	?	other	negative
78	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.5	t	1.9	t	81	t	0.83	t	98	f	?	SVI	negative
37	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	1.7	t	1.9	t	95	t	1.05	t	90	f	?	SVI	negative
85	F	f	f	f	f	f	f	f	t	f	f	f	f	f	f	t	7.3	t	2.4	t	66	t	1.09	t	61	f	?	other	negative
25	F	f	f	f	f	f	f	f	f	f	f	f	t	f	f	t	1.1	f	?	t	101	t	1.07	t	94	f	?	other	negative
26	F	f	f	f	f	f	f	f	f	f	f	f	f	f	t	t	1.8	t	2.5	t	147	t	1.13	t	129	f	?	SVHC	negative
58	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.26	t	2.7	t	120	t	1.27	t	95	f	?	other	negative
51	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	2.8	f	?	t	69	t	0.76	t	91	f	?	other	negative
64	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	?	f	?	f	?	f	?	f	?	f	?	other	negative
44	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	45	t	1.4	t	39	t	1.16	t	33	f	?	SVI	primary_hypothyroid
48	M	f	t	f	f	f	f	f	f	t	f	f	f	f	f	t	5.4	t	1.9	t	87	t	1	t	87	f	?	other	negative
61	F	f	f	f	f	f	f	f	f	f	f	t	f	f	f	t	0.99	t	1.5	t	63	t	0.56	t	113	f	?	SVI	negative
35	F	t	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.25	t	2	t	121	t	0.81	t	148	f	?	other	negative
70	F	f	f	f	t	f	f	f	f	f	f	f	f	f	f	t	1.1	t	1.7	t	95	t	0.68	t	140	f	?	SVI	negative
81	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.2	t	2.2	t	133	t	0.78	t	171	f	?	other	negative
83	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	0.92	t	0.9	t	86	t	0.76	t	113	f	?	SVI	negative

hypothyroid

3. Technology Stack

Front end	HTML/CSS
Back end	Flask

4. Architecture Description

4.1. Data description:

We will use the Thyroid Disease Data Set from the UCI Machine Learning Repository, which contains 7200 instances in different batches of data.

4.2. Data Preprocessing:

In this step we will import the necessary Python libraries such as Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn etc.

And importing the dataset as pandas DataFrame.

4.3. Exploratory Data Analysis:

In this step we handled null values, changed the columns names, plotted multiple graphs & charts in Seaborn and Matplotlib to understand the data properly and also the distribution of the data.

As there were no missing values in the data so we proceed with the visualization and analysis. For each specific feature, by analysing the data we got to know about some key points which can impact the final predictions.

4.4. Data Ingestion:

In this step, we divided the data into 3 CSV files, raw.csv, train.csv & test.csv. with the help of Train Test Split, we divided the data into train and test set, in the ratio of 80-20%, where 80% data got for training the model(train.csv) and 20% is for testing the model(test.csv).

4.5. Data Transformation:

In this step, we performed feature scaling using scikit-learn.

First, we divided the both train & test dataset into 2 categories, categorical data & numerical data. Then we apply the scaling by using the fit-transformed method. Also, we have read the train and test data and changed them into arrays. Then saved this as rf_model.pkl file for further steps.

4.6. Model Trainer:

In this step, we train the model using multiple algorithms and find the best algorithm with highest accuracy. We used Logistic Regression, Decision Tree, Random Forest, SVM, Navi Bayes etc. algorithms to train the model.

4.7. Prediction:

Random Forest got the highest accuracy score 89.19%

4.8. Saving the Model:

Here we saved the model using pickle library, which

4.9. Deploy In Localhost:

We have created an HTML template and deployed the model using streamlit