iNeuron.ai

Credit Card Default Prediction

High Level Design (HLD) Documentation

Shantanu Kukkar

25-08-2024

Document Version Control

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

Table of Contents

1. Introduction

1.1. Why this High-Level Design Document?

This purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspect and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project

1.2. Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), & technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators to the system.

1.3. Definitions

Term	***Description***
Database	Collection of all the information monitored by this system
IDE	Integrated-Development Environment

2. General Description

2.1. Product Perspective

The Thyroid Disease Detection solution system is a data science- based machine learning model which help us to detect the thyroid disease in people and take necessary action.

2.2. Problem Statement

Thyroid disease is a common cause of medical diagnosis and prediction, with an onset that is difficult to forecast in medical research. The thyroid gland is one of our body's most vital organs. Thyroid hormone releases are responsible for metabolic regulation. Hyperthyroidism and hypothyroidism are one of the two common diseases of the thyroid that releases thyroid hormones in regulating the rate of body's metabolism. The main goal is to predict the estimated risk on a patient's chance of obtaining thyroid disease or not.

2.3. PROPOSED SOLUTION

The solution is a classification-based machine learning model. This can be implemented by using the different classification algorithms (such as Logistic Regression, Random Forest, Decision Tree, SVM, Navi Bayes etc.)

To build this model we will perform Data Preprocessing which contains feature engineering, feature selection, feature transformation etc.

2.4. Technical Requirements

This document addresses the requirements for detecting who is defaulting the bill payments.

Here are some requirements for this project.

- Model should be exposed through API or User interface, so that anyone can test the model
- Model should be deployed on cloud (Azure, AWS, GCP, Heroku).

2.5. Data Requirements

Data requirement completely depends on our problem statement.
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.

2.6. Tools used

Python programming language and frameworks such as NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn & Flask are used to build the whole model.

- Jupyter Notebook is used as an IDE.
- For visualization of the plots Matplotlib & Seaborn are used.
- Front end development is done HTML/CSS.
- Flask is used for backend development.
- GitHub is used as a version control system.

2.7. Constraints

The Thyroid Disease Detection solution system must be correct enough that it not mislead any report and as automated as possible and users should not be required to know any of the workings.

2.8. Assumptions

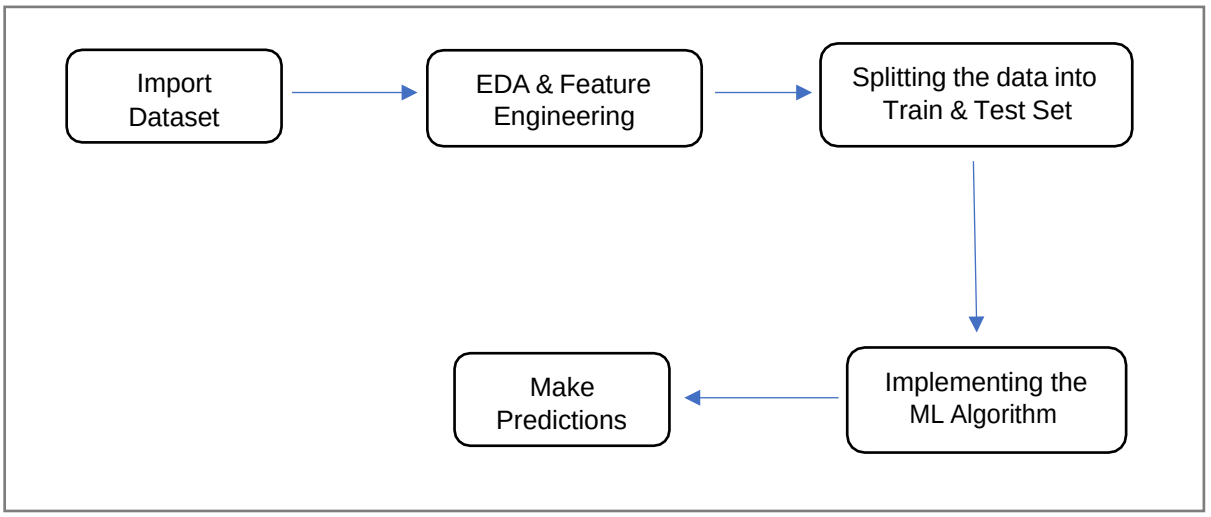
The main objective of the project is to implement the use cases as previously mentioned for new dataset that comes through Hospitals which has this solution install in their campus to capture people reports.

3. Design Details

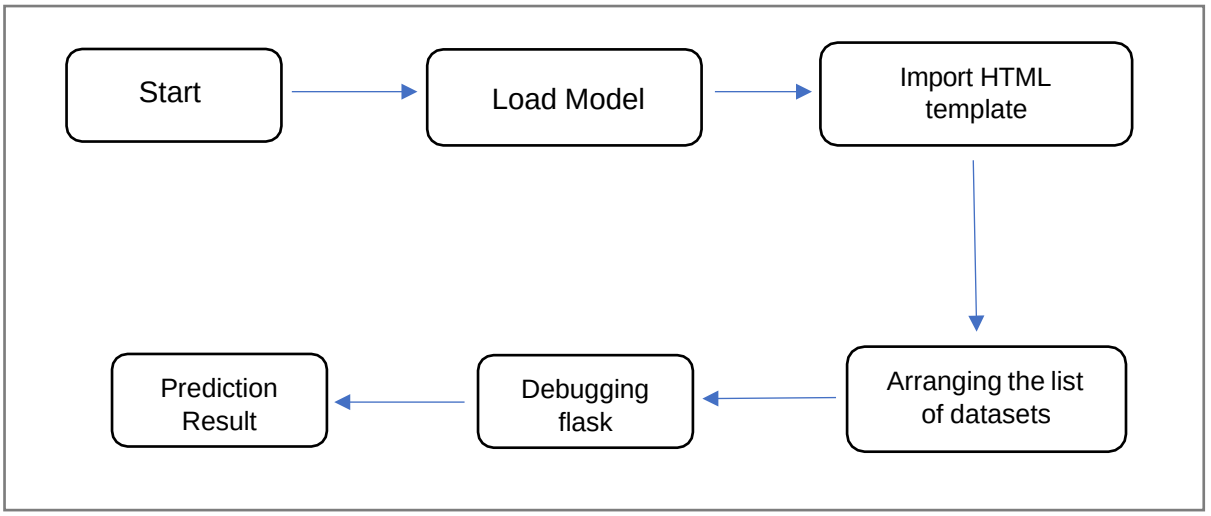
3.1. Process Flow

For identifying the defaulters, we will use a machine learning base model. Below is the process flow diagram as shown below.

Proposed Methodology:



3.1.1. Deployment Process



3.2. Evant log

The system should log every event so that the user will know what process is running internally.

Initial Step-by-Step Description:

1. The system identifies at what step logging required
2. The system should be able to log each and every system flow
3. Developers can choose logging methods. You can choose database logging/ File logging as well.
4. System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

3.3. Error Handling
Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

3.4. Performance
We are developing a machine learning-based Thyroid Disease Detection solution for identifying thyroid diseases in patients who exhibit symptoms related to thyroid issues. The main objective is to facilitate early detection and prompt intervention. To ensure the system's effectiveness, regular model retraining will be implemented to continuously improve its performance and accuracy over time. This will enable us to take necessary actions promptly and provide appropriate medical attention to individuals diagnosed positively for thyroid disease.

3.5. Reusability
The code written and the components used should have the ability to be reused with no problems.

3.6. Application Compatibility
The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

3.7. Resource Utilization
When any task is performed, it will likely use all the processing power available until that function is finished.

3.8. Deployment
*Localhost

4. Conclusion
We propose to develop a Thyroid Disease Detection solution utilizing machine learning. The system will be trained on health-care domain data from patients who have undergone thyroid diagnosis. The model's performance will be evaluated across various use cases. Subsequently, we will utilize the trained model to predict the presence of thyroid disease in individuals exhibiting symptoms. In case of a positive prediction, the system will promptly alert the individuals, ensuring they receive immediate medical attention and treatment. Our primary focus is on achieving high accuracy to minimize the risk of generating misleading reports.