Phase 5 Documentation

Date	**28-10-2023**
Team ID	**4146**
Project Name	**Machine learning model deployment with IBM Watson studio**

Table of Contents

1.Introduction

This document provides an overview of our comprehensive project, which encompassed the development and deployment of a heart disease prediction model. The project was executed in three key phases: Model Development, IBM Cloud Watson Studio Deployment, and Flask Integration.

2.Project Overview

In this project, we embarked on a comprehensive journey to develop and deploy a heart disease prediction model using machine learning techniques. Our primary objectives were to leverage the power of IBM Cloud Watson Studio for both model development and deployment and integrate the model into a Flask application. The project encompassed three key phases: model development, deployment, and integration.

3.Objective

The central goal of our project was to build an accurate heart disease prediction model and make it accessible to a wide audience through a user-friendly web application.

4.Design Thinking Process

Our project was guided by a design thinking process, which involved iterative problem-solving and a focus on user needs:

Empathize: We started by understanding the need for a heart disease prediction tool. We empathized with potential users and identified the desire for an accessible and accurate prediction model.

Define: The project scope was defined, outlining the goals and deliverables. We set clear objectives for model development and deployment.

Ideate: We considered various machine learning classifiers for the model. After careful consideration, Gradient Boosting, SVM, and Random Forest were selected as potential candidates.

Prototype: The selected classifiers were implemented and tested in a development environment to create prototypes. We used a heart disease dataset for training and evaluation.

Test and Refine: We rigorously tested the prototypes and evaluated their performance using various metrics. Gradient Boosting emerged as the most accurate model during testing.

Implement and Deploy: After selecting Gradient Boosting as the optimal model, we proceeded with model deployment on IBM Cloud Watson Studio. This involved configuration, scaling, and security considerations.

Integrate Flask Application: A Flask web application was created to serve as the user interface for the deployed model. API endpoints were defined to handle user interactions and model predictions.

5.Development Phases

The project progressed through the following key phases:

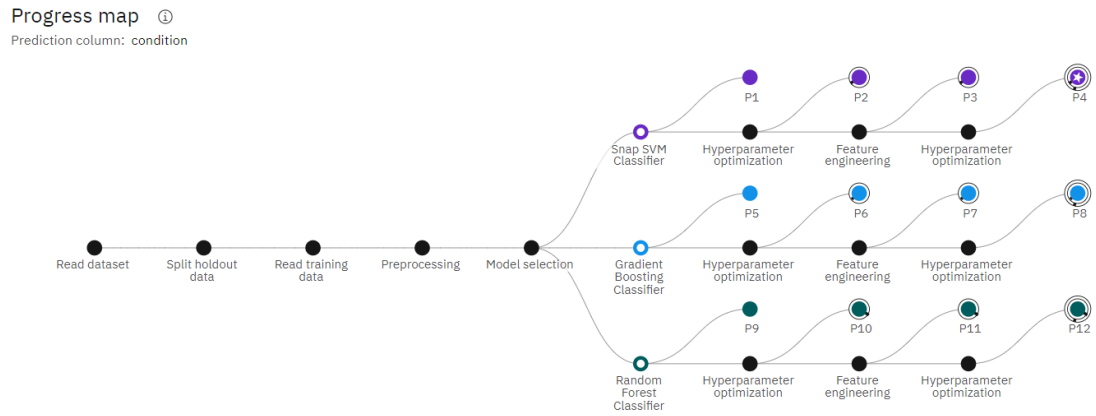
5.1.Data Collection and Preprocessing:

- Acquired a dataset containing health-related features and heart disease labels.
- Preprocessed the data by handling missing values, encoding categorical variables, and scaling numeric features.

5.2.Model Development:

5.2.1.Model Training

- We trained each classifier using the preprocessed dataset and conducted cross-validation to optimize hyperparameters.
- The models were evaluated based on performance metrics such as accuracy, precision, recall, and F1-score.



5.2.2.Model Evaluation:

- We used various evaluation metrics such as accuracy, precision, recall, F1-score, and ROC AUC to assess the performance of each model.
- The models were tested rigorously using cross-validation and an independent test dataset.

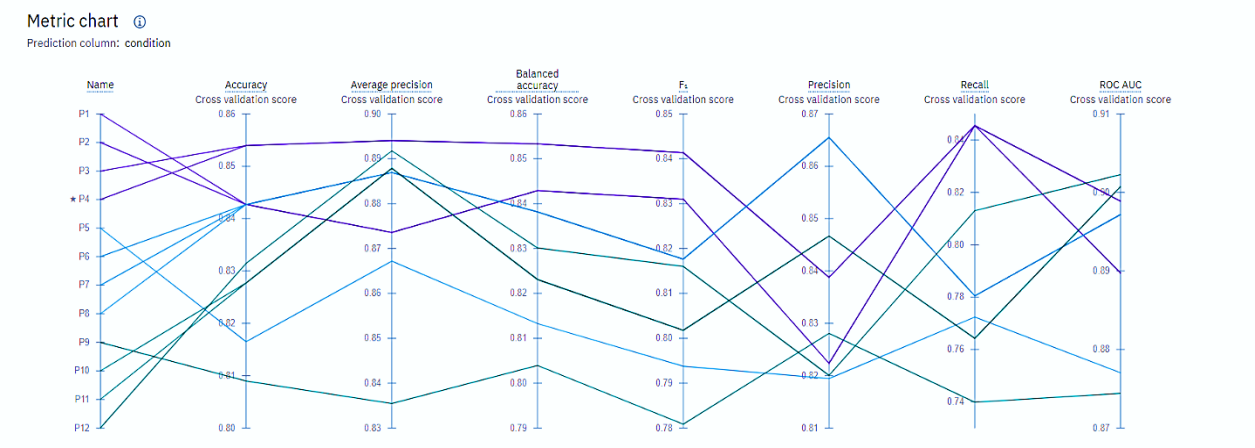
Projects / heart-disease-pred / Heart-Disease-Prediction

Experiment summary | Pipeline comparison | ★ Rank by: Accuracy (Optimized) | Cross validation score

Pipeline leaderboard

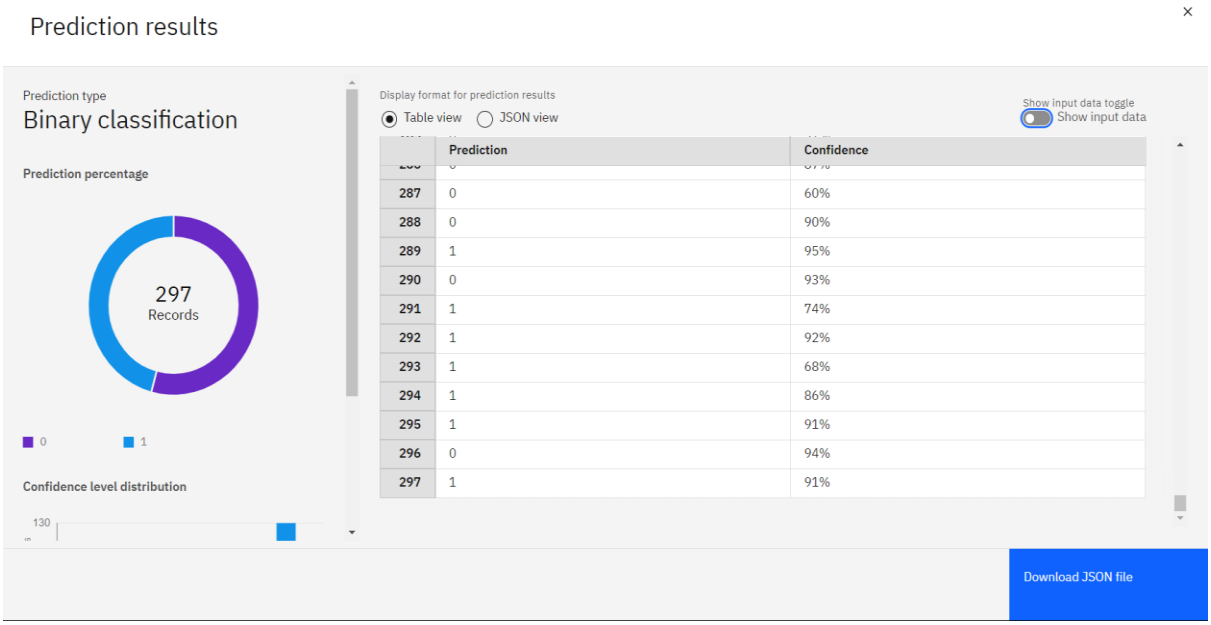
	Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 4	Snap SVM Classifier	0.854	HPO-1 FE HPO-2	00:00:16
	2	Pipeline 3	Snap SVM Classifier	0.854	HPO-1 FE	00:00:14
	3	Pipeline 8	Gradient Boosting Classifier	0.843	HPO-1 FE HPO-2	00:00:28
	4	Pipeline 7	Gradient Boosting Classifier	0.843	HPO-1 FE	00:00:18
	5	Pipeline 6	Gradient Boosting Classifier	0.843	HPO-1	00:00:01
	6	Pipeline 2	Snap SVM Classifier	0.843	HPO-1	00:00:01
	7	Pipeline 1	Snap SVM Classifier	0.843	*None*	00:00:01
	8	Pipeline 12	Random Forest Classifier	0.831	HPO-1 FE HPO-2	00:00:29

9	**Pipeline 11**	Random Forest Classifier	**0.828**	HPO-1 FE	00:00:21
10	**Pipeline 10**	Random Forest Classifier	**0.828**	HPO-1	00:00:03
11	**Pipeline 5**	Gradient Boosting Classifier	**0.816**	*None*	00:00:01
12	**Pipeline 9**	Random Forest Classifier	**0.809**	*None*	00:00:01



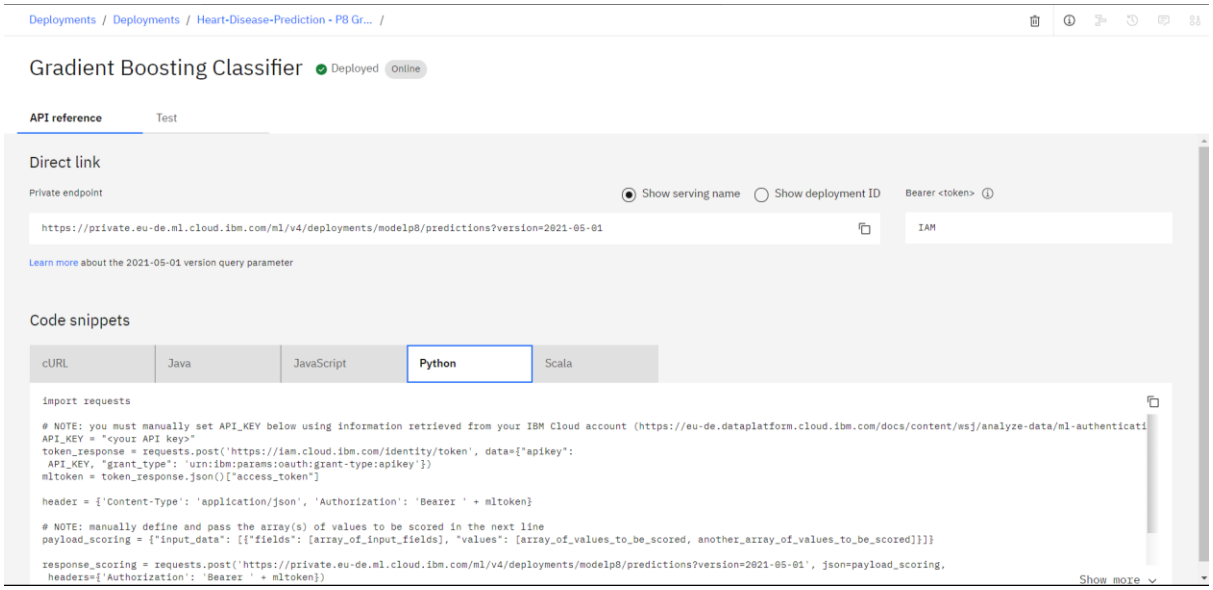
5.2.3.Results:

- Gradient Boosting outperformed the other algorithms in terms of predictive accuracy during testing.



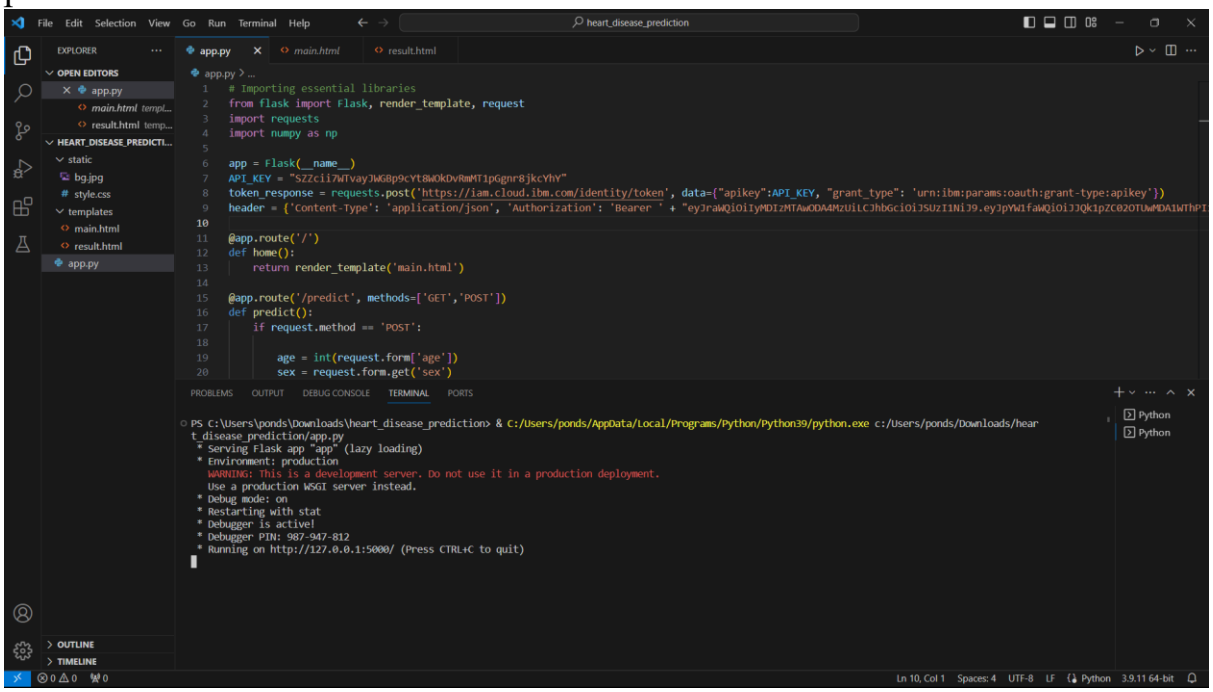
5.3.IBM Cloud Watson Studio Deployment:

- Exported the Gradient Boosting model and deployed it on IBM Cloud Watson Studio.
- Configured deployment settings, including scalability and security options.



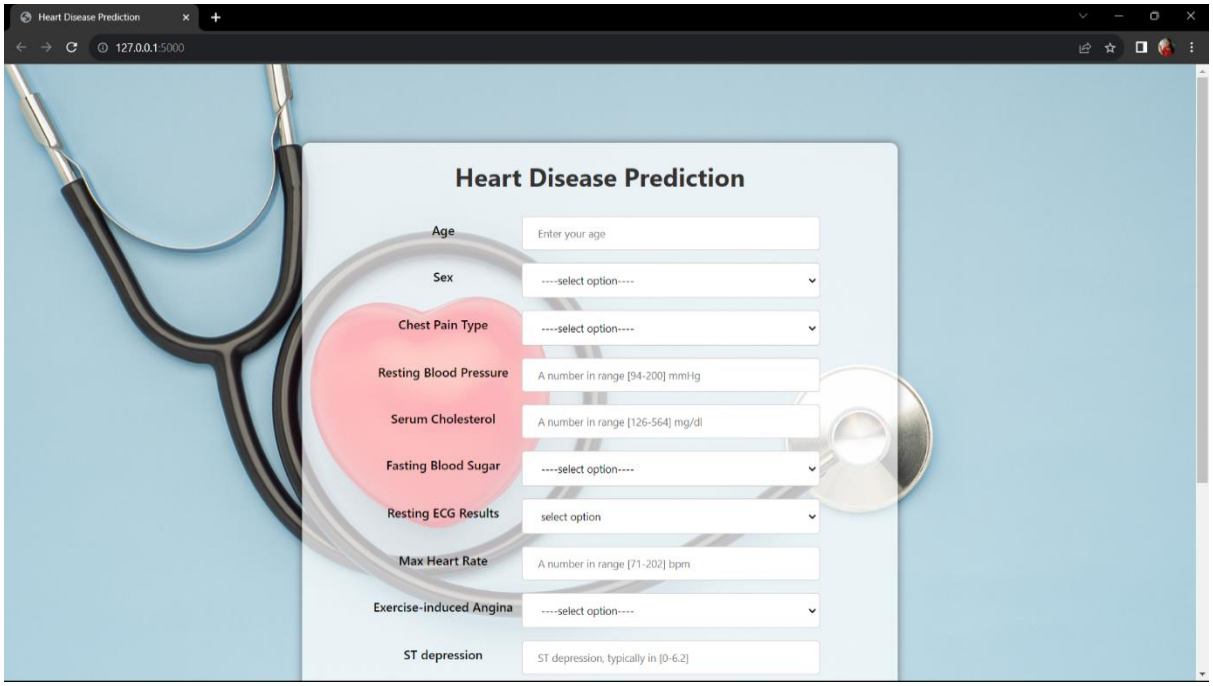
5.4.Flask Integration:

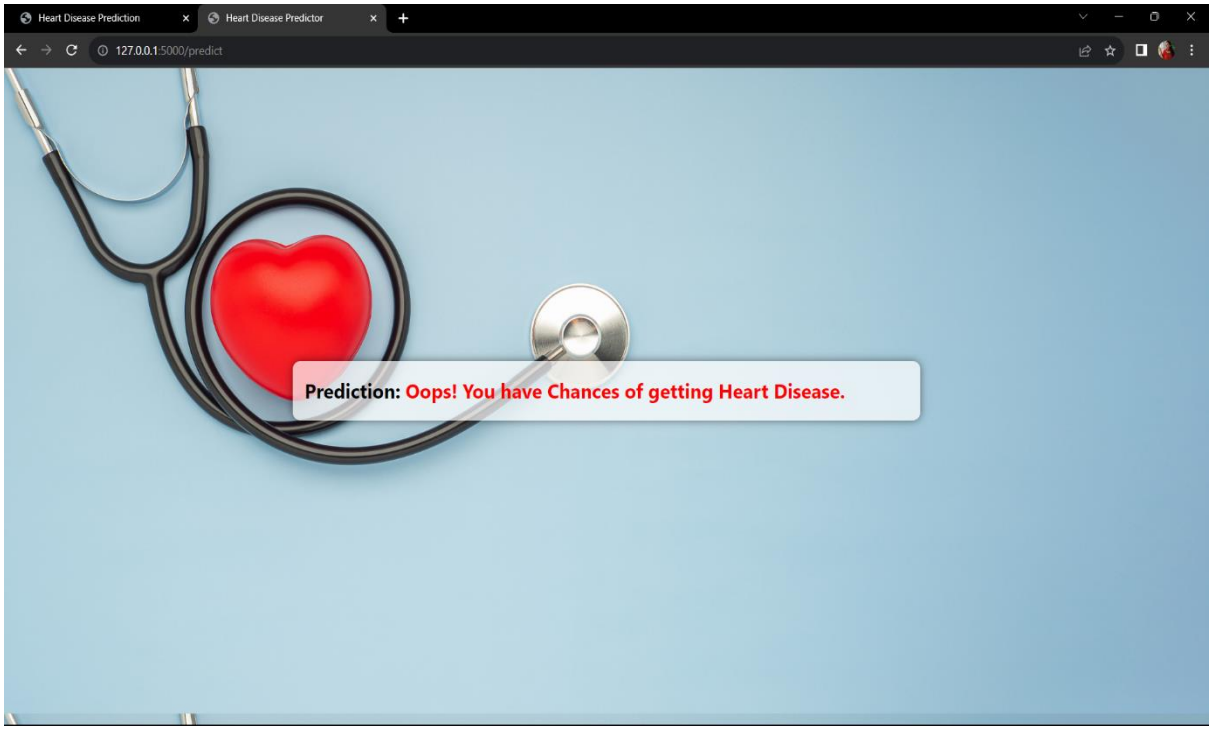
- Developed a Flask application with API endpoints to interact with the deployed model.
- Implemented data preprocessing and post-processing to ensure accurate predictions.



5.5.User Interface Design:

- Designed a user-friendly web interface using HTML and CSS.
- Created a form for users to input health-related data.
- Model prediction results displayed to the user in a clear and understandable format.





6.Platform Layout and Features

The platform encompasses two main components:

IBM Cloud Watson Studio: This serves as the backend for model deployment and management. It includes features for model deployment, scalability, and monitoring.

Flask Application: This frontend component provides a user-friendly web interface. Users can access the heart disease prediction tool through a form, input their health-related data, and receive predictions.

7.Technical Implementation Details

The technical aspects of the project included:

- Data preprocessing to ensure data quality and suitability for machine learning.
- Training and evaluation of three machine learning models (Gradient Boosting, SVM, and Random Forest).
- Exporting and deploying the selected Gradient Boosting model on IBM Cloud Watson Studio.
- Integration of the deployed model into a Flask application with API endpoints.
- User interface design using HTML and CSS.
- Data handling and result presentation within the Flask application.
- Model scalability and security configurations on IBM Cloud Watson Studio.

8.Conclusion

Our project successfully achieved its objectives, culminating in the deployment of the most accurate heart disease prediction model, Gradient Boosting, on IBM Cloud Watson Studio. The model is now accessible through a user-friendly Flask application, making it easier for users to make informed healthcare decisions. This project represents a significant step in leveraging machine learning for the benefit of healthcare and highlights the potential of deploying models in cloud environments for widespread use.