Phase 3 Documentation

Date	**26-10-2023**
Team ID	**4146**
Project Name	**Machine learning model deployment with IBM Watson studio**

Table of Contents

Heart Disease Prediction Model Evaluation

1.Introduction

In this document, we present the results of our project, which involved the training and evaluation of a heart disease prediction model using three different classifiers: Gradient Boosting, Snap Support Vector Machine (SVM), and Random Forest. Our project was conducted in IBM Cloud Watson Studio, a powerful platform for data science and machine learning.

2.Problem Statement

Our goal is to design and deploy a scalable and accessible web service that leverages the power of machine learning to predict the risk of heart disease in individuals based on their medical and lifestyle data. The deployment will use IBM Watson Studio, an integrated environment for data science and machine learning, to provide a user-friendly interface for healthcare professionals, researchers, and patients to assess heart disease risk quickly and accurately.

3.Steps Involved in Model Evaluation:

3.1.Data Collection and Preprocessing

- We obtained a dataset containing various health-related features and the presence/absence of heart disease.
- This dataset served as the foundation for our model.
- The dataset underwent preprocessing, which included handling missing values, encoding categorical variables, and scaling numeric features.
- This ensured that the data was suitable for machine learning.

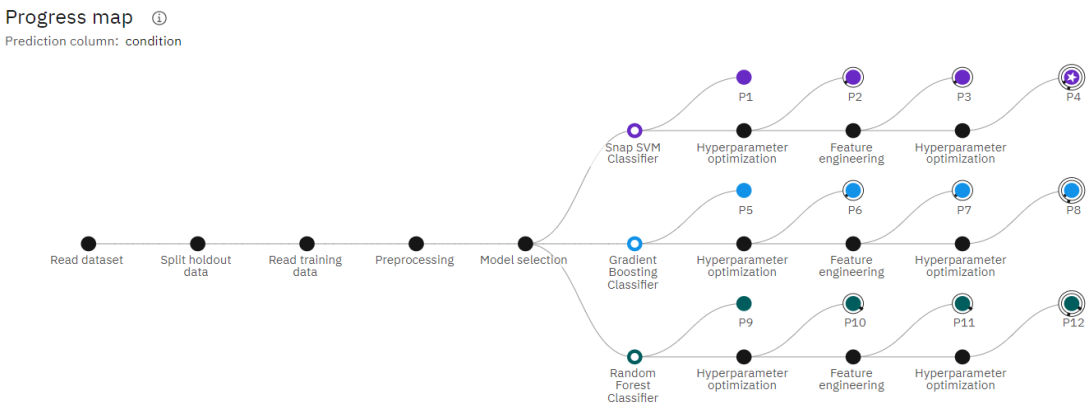
3.2.Classifier Selection

For our heart disease prediction model, we selected three powerful classifiers:

- **Gradient Boosting:** Gradient Boosting is an ensemble learning method that combines the predictions of multiple decision trees, iteratively improving model performance.
- **Snap Support Vector Machine (SVM):** SVM is a robust and versatile classifier that works well for both linear and non-linear classification problems.
- **Random Forest:** Random Forest is another ensemble method that combines multiple decision trees to make predictions. It is known for its ability to handle complex datasets and maintain good accuracy.

3.3.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.



3.4.Model Evaluation:

- We used various evaluation metrics such as accuracy, precision, recall, F1-score, and ROC AUC to assess the performance of each model.

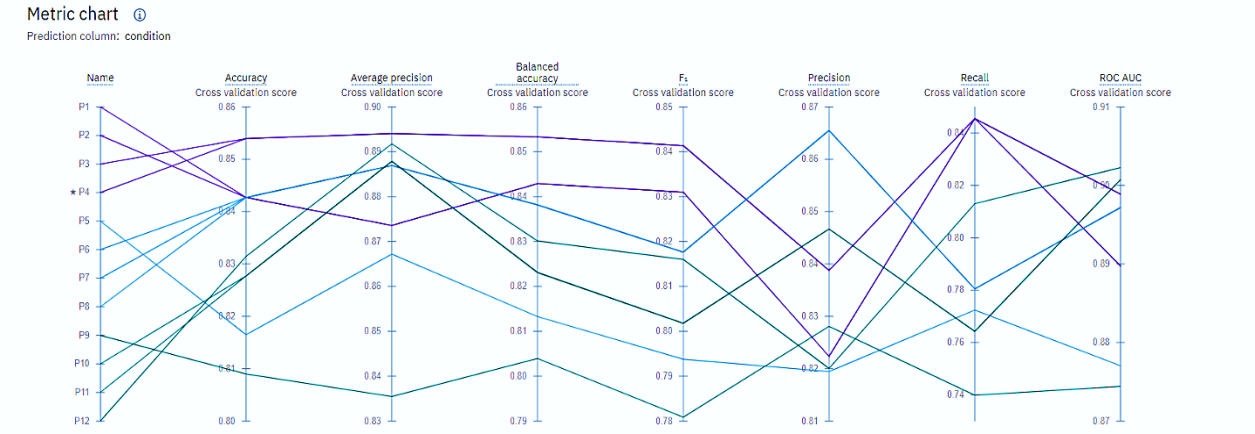
- The models were tested rigorously using cross-validation and an independent test dataset.

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

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

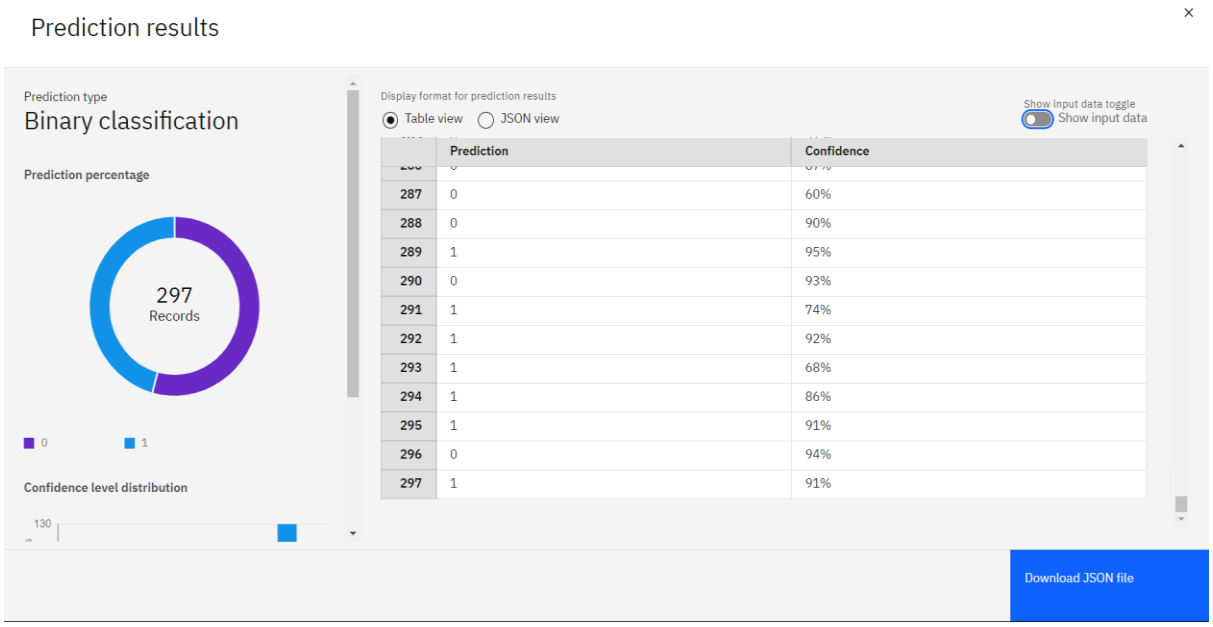
Pipeline leaderboard

	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



3.5.Results:

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



4.Conclusion:

This project has successfully developed and evaluated heart disease prediction models using Gradient Boosting, SVM, and Random Forest classifiers. The Gradient Boosting model showed promising results with high accuracy. By integrating this model into a web service using Flask, we aim to make it accessible to a broader audience and potentially contribute to better healthcare decision-making.