***HEART FAILURE PREDICTION USING AUTO AI EXPERIMENT***

***By***

***Bhargavi Allam***

***bhargavi333bavi@gmail.com***

***Introduction***

Heart failure (HF) is a complex clinical syndrome and not a disease. It prevents the heart from fulfilling the circulatory demands of the body, since it impairs the ability of the ventricle to fill or eject blood. It is characterized by symptoms, such as breathlessness, ankle swelling and fatigue that may be accompanied by signs, for example elevated [jugular venous pressure](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/central-venous-pressure), pulmonary crackles, and peripheral edema, caused by structural and/or functional cardiac or non-cardiac abnormalities. HF is a serious condition associated with high morbidity and mortality rates. According to the European Society of Cardiology (ESC), 26 million adults globally are diagnosed with HF, while 3.6 million are newly diagnosed every year. 17–45% of the patients suffering from HF die within the first year and the remaining die within 5 years. The related to HF management costs are approximately 1–2% of all healthcare expenditure, with most of them linked with recurrent hospital admissions

***a. Overview***

Cardiovascular diseases kill approximately 17 million people globally every year, and they mainly exhibit as myocardial infarctions and heart failures. Heart failure (HF) occurs when the heart cannot pump enough blood to meet the needs of the body. Available electronic medical records of patients quantify symptoms, body features, and clinical laboratory test values, which can be used to perform biostatistics analysis aimed at highlighting patterns and correlations otherwise undetectable by medical doctors. Machine learning, in particular, can predict patients’ survival from their data and can individuate the most important features among those included in their medical records.

***b. Purpose***

Heart failure is a serious condition with high prevalence (about 2% in the adult population in developed countries, and more than 8% in patients older than 75 years). About 3–5% of hospital admissions are linked with heart failure incidents. Heart failure is the first cause of admission by healthcare professionals in their clinical practice. The costs are very high, reaching up to 2% of the total health costs in the developed countries. Building an effective disease management strategy requires analysis of large amount of data, early detection of the disease, assessment of the severity and early prediction of adverse events. This will inhibit the progression of the disease, will improve the quality of life of the patients and will reduce the associated medical costs.

***Literature Survey***

***a. Existing problem***

The heart is a vital organ in the human body that is liable for blood circulation. The heart is responsible for oxygen and energy supply to all organs of the body including itself. Heart disease causes the abnormal blood circulation in the body that might be fatal for human life. Hence, if the heart stops its normal functionality, the whole system will be dead. From the literature, various risk factors are identified that cause the heart disease. The risk factors of heart diseases are classified into two major types such as the risk factors that can alter, e.g., smoking and physical exercise, and the risk factors that cannot be alter, e.g., gender, age, and patient’s family history

***b. Proposed solution***

In this application, we study the effects of age, smoking, BMI, gender, family history, cholesterol, exercise per week, palpitations per day and heartbeats per minute determine how much of a difference these factors can make on heart failure. Here, we build a model using Auto AI and build a web application where we can showcase the prediction of heart failure. By using our application, customers see the radical difference their lifestyle choices make on their life. The following are the methodologies

adopted for data analysis, prediction and display of results by integrating with the system:

● IBM Auto AI

● Node-red

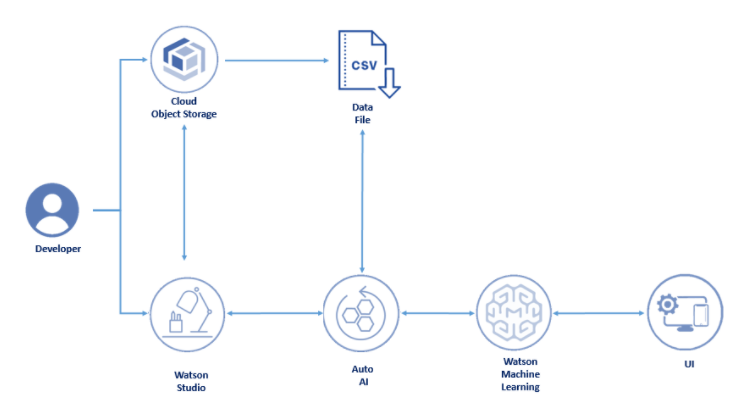
Created a model from a dataset that includes the age, smoking, BMI, gender, family history, cholesterol, exercise per week, palpitations per day and heartbeats per minute to predict the heart failure. Using IBM Auto AI, IBM cloud services, Machine Learning algorithms and Node-red, automated all of the tasks involved in building predictive models for different requirements. Node-red is a framework that helps build web apps that could act as the interface to the user

for input and output in the front and integration of values in the backend.

***Theoretical Analysis***

1. ***Block diagram***

Prediction using IBM Auto AI Experiment:



**Services Used:**

* IBM Watson Studio
* IBM Watson Machine Learning
* Node-RED
* IBM Cloud Object Storage

1. ***Hardware / Software designing***

***Hardware:***

● HP laptop/8Gb RAM/64 bit/Windows 10.

***Software:***

●IBM Cloud

***Experimental Investigations***

***Dataset:***

SmartInterns provided the data source link. The dataset is comprised of 10800 records with 10 attributes. Attributes are as follow age, smoking, BMI, gender, family history, cholesterol, exercise per week, palpitations per day, heart failure and heartbeats per minute. The data was in a structured format and was stores in a CSV file. In a dataset, not every attribute has an impact on the prediction. Whereas some attributes even decline the accuracy, so it becomes necessary to remove these attributes from the features of the code. Removing such attributes not only help in improving accuracy but also the overall performance and speed.

***Machine Learning***

Machine learning can be defined as the process of teaching a computer system that allows it to make accurate predictions after the data is fed.

***Classification:***

Since the data in the dataset is categorical type of data, the best suited Machine Learning model is classification. So, cleaning of dataset becomes vital for using the data under various

regression algorithms. Classification analysis is a data analysis task, that identifies and assigns categories to a collection of data to allow for more accurate analysis. The classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics.

Many techniques for performing classifier predictions have been developed, but, in this project, two models

* **Random Forest classifier** and
* **Decision tree classifier** were tested and compared.

***Decision Tree Classifier***

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

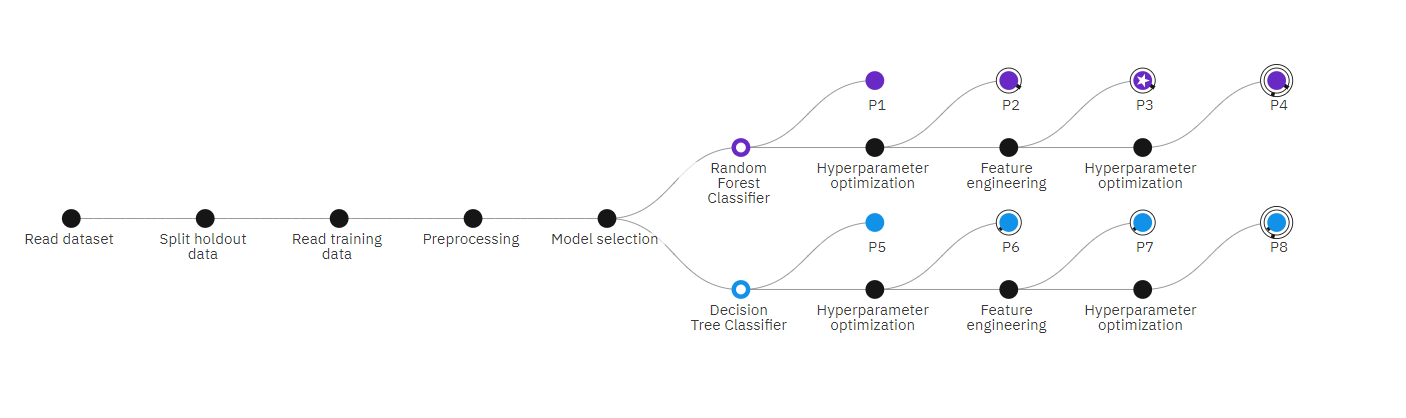
***Random Forest Classifier***

The core unit of random forest classifiers is the decision tree. The decision tree is a hierarchical structure that is built using the features (or the independent variables) of a data set. Each node of the decision tree is split according to a measure associated with a subset of the features. The random forest is a collection of decision trees that are associated with a set of [bootstrap samples](https://www.sciencedirect.com/topics/mathematics/bootstrap-sample) that are generated from the original data set. The nodes are split based on the entropy of a selected subset of the features. The subsets that are created from the original data set, using bootstrapping, are of the same size as the original data set.

## ***Random Forests vs Decision Trees***

* Random forest is a set of multiple decision trees.
* Deep decision trees may suffer from overfitting, but random forests prevent overfitting by creating trees on random subsets.
* Decision trees are computationally faster.
* Random forest is difficult to interpret, while a decision tree is easily interpretable and can be converted to rules.

***Progress Map***

******

***Flowchart***

Create Watson service

Associate Cloud Object Storage service

Add Auto AI Service and Associate ML to it

Add dataset to Auto AI experiment

Run the experiment

Save the model

Deploy the model

Build Node-red Application

Integrate Node-red with Auto AI model

Open Node-red dashboard

Fill the form

Get prediction by IBM Auto AI experiment

***Result***

***Prediction using IBM Auto AI:***

With Random Forest Classifier model using age, smoking, BMI, gender, family history, cholesterol, exercise per week, palpitations per day, and heartbeats per minute as independent variables and heart failure as dependent variable, with Accuracy 86.2% and Build time 23 seconds “Pipeline 3” is the top performer, which is the best model ranked by IBM Auto AI.

***Advantages & Disadvantages***

**Advantages:**

* Machine learning applied to medical records, in particular, can be an effective tool both to predict the survival of each patient having heart failure symptoms, and to detect the most important clinical features (or risk factors) that may lead to heart failure.
* Scientists can take advantage of machine learning not only for clinical prediction, but also for feature ranking.
* Computational intelligence, especially, shows its predictive power when applied to medical records, or coupled with imaging.

**Disadvantages:**

* Insufficient data can result in miss interpretation and wrong prediction.

For example, in the same dataset, family history, no clear idea about IBM, cholesterol or lying about smoking etc. are some other missing attributes that can contribute or make a change in accuracy and prediction.

* The most important in the building of models would be data pre-processing. If this initial stage is not carried out or if the data is not pre-processed properly by a developer, this can

lead to a poor prediction model.

***Applications***

* Given the importance of a vital organ such as the heart, predicting heart failure has become a priority for medical doctors and physicians, but to date forecasting heart failure-related events in clinical practice usually has failed to reach high accuracy.
* Machine learning applied to medical records, in particular, can be an effective tool both to predict the survival of each patient having heart failure symptoms [[18](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5#ref-CR18), [19](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5#ref-CR19)], and to detect the most important clinical features (or risk factors) that may lead to heart failure

***Conclusion***

* Aim of this project is to predict whether or not a patient will develop heart disease. This research was done on supervised machine learning classification techniques using decision tree and, random forest. Various experiments using different classifier algorithms were conducted through the IBM Auto AI service.
* And that is achieved.

***Future Scope***

Here the scope of the project is that integration of clinical decision support with computer-based

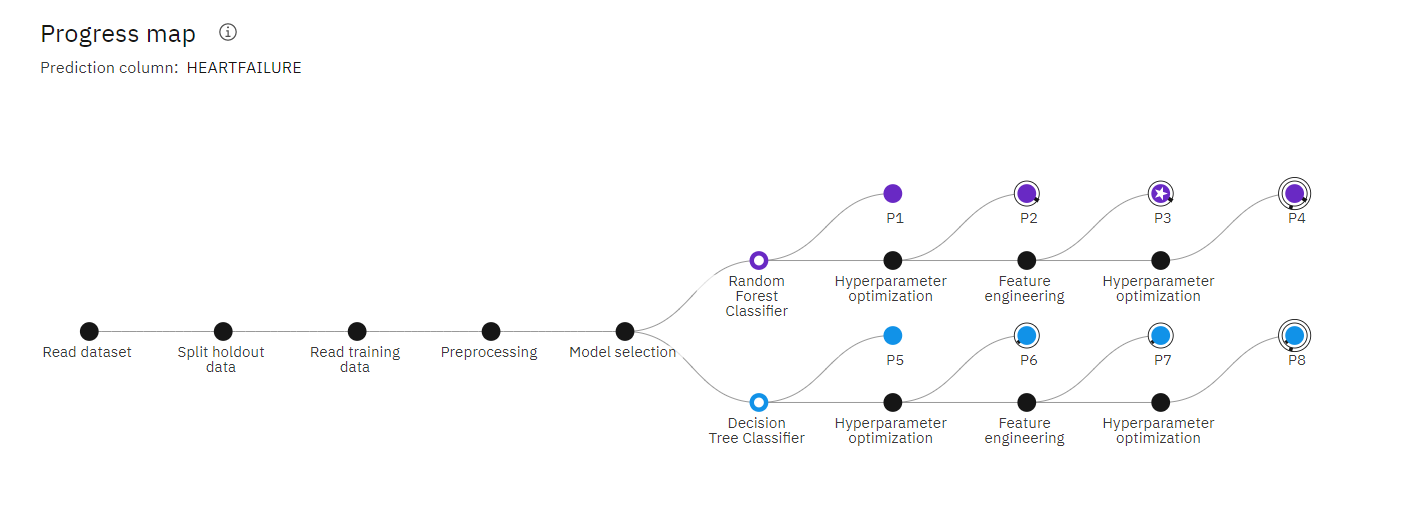
patient records could reduce medical errors, enhance patient safety, decrease unwanted practice

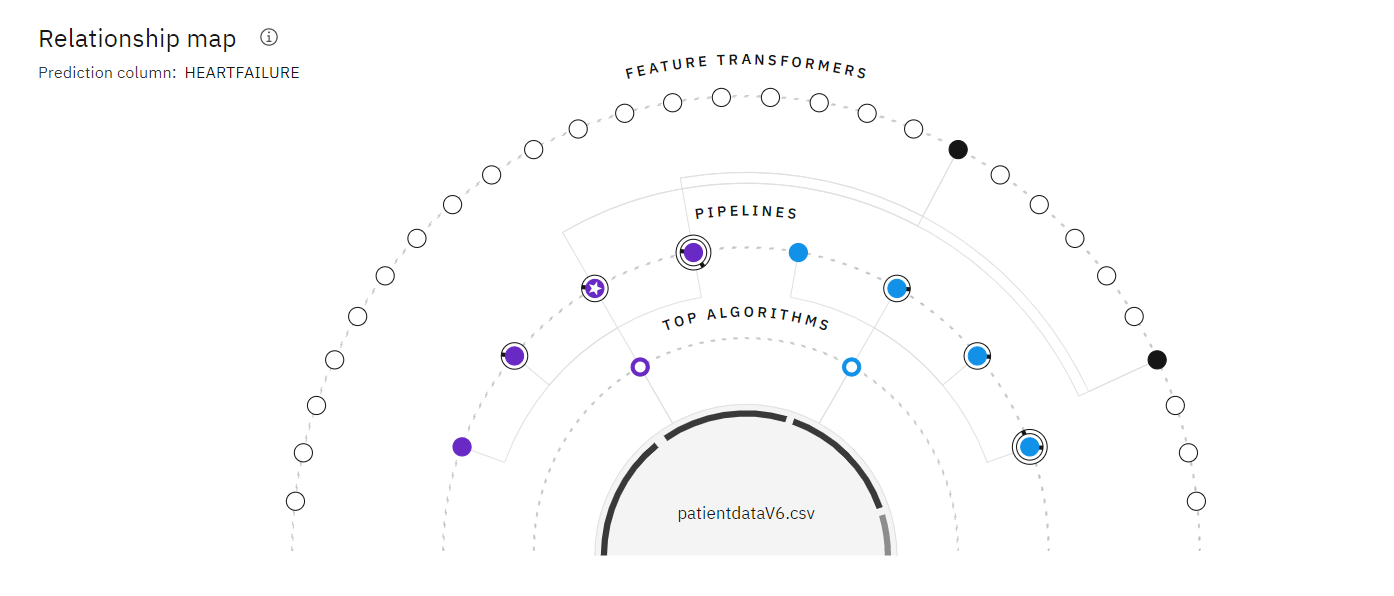
variation, and improve patient outcome.

***Appendix***

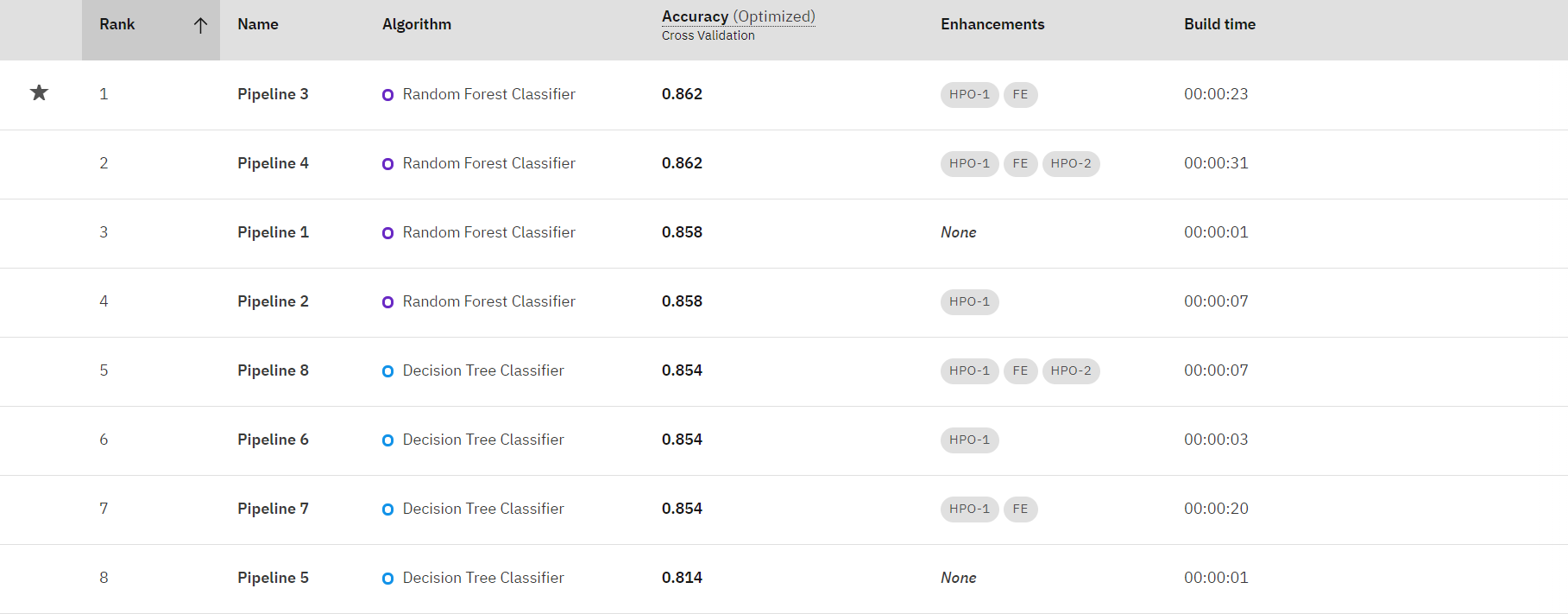
***Source code***

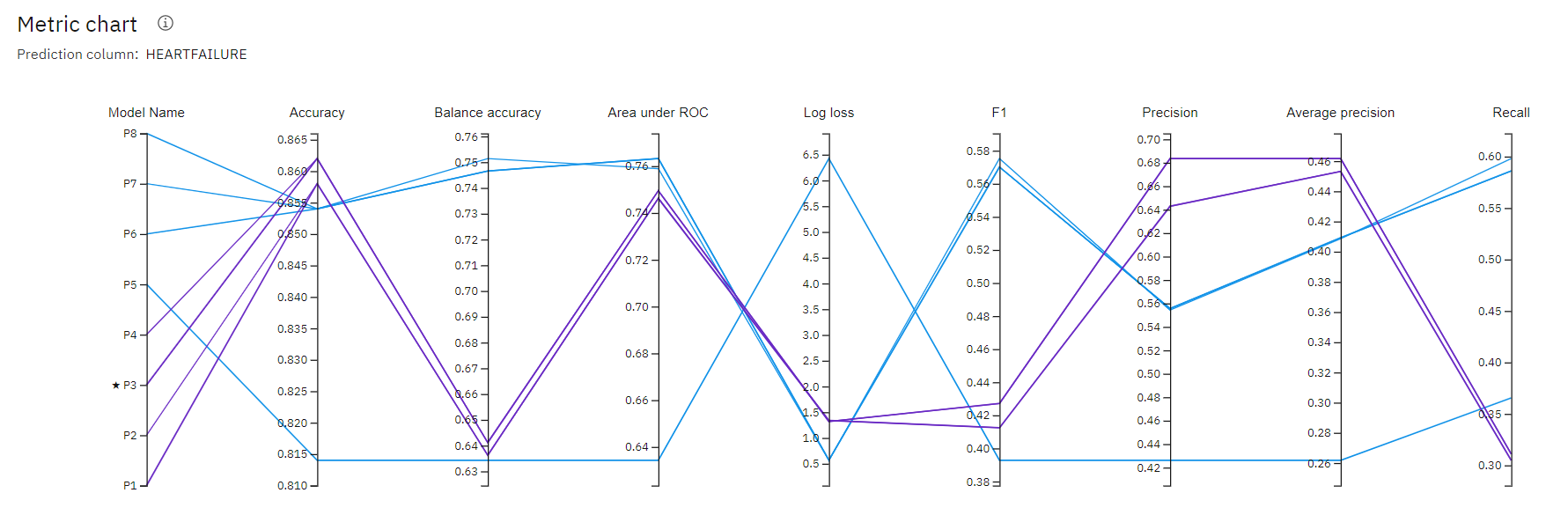
* ***Auto AI experiment***

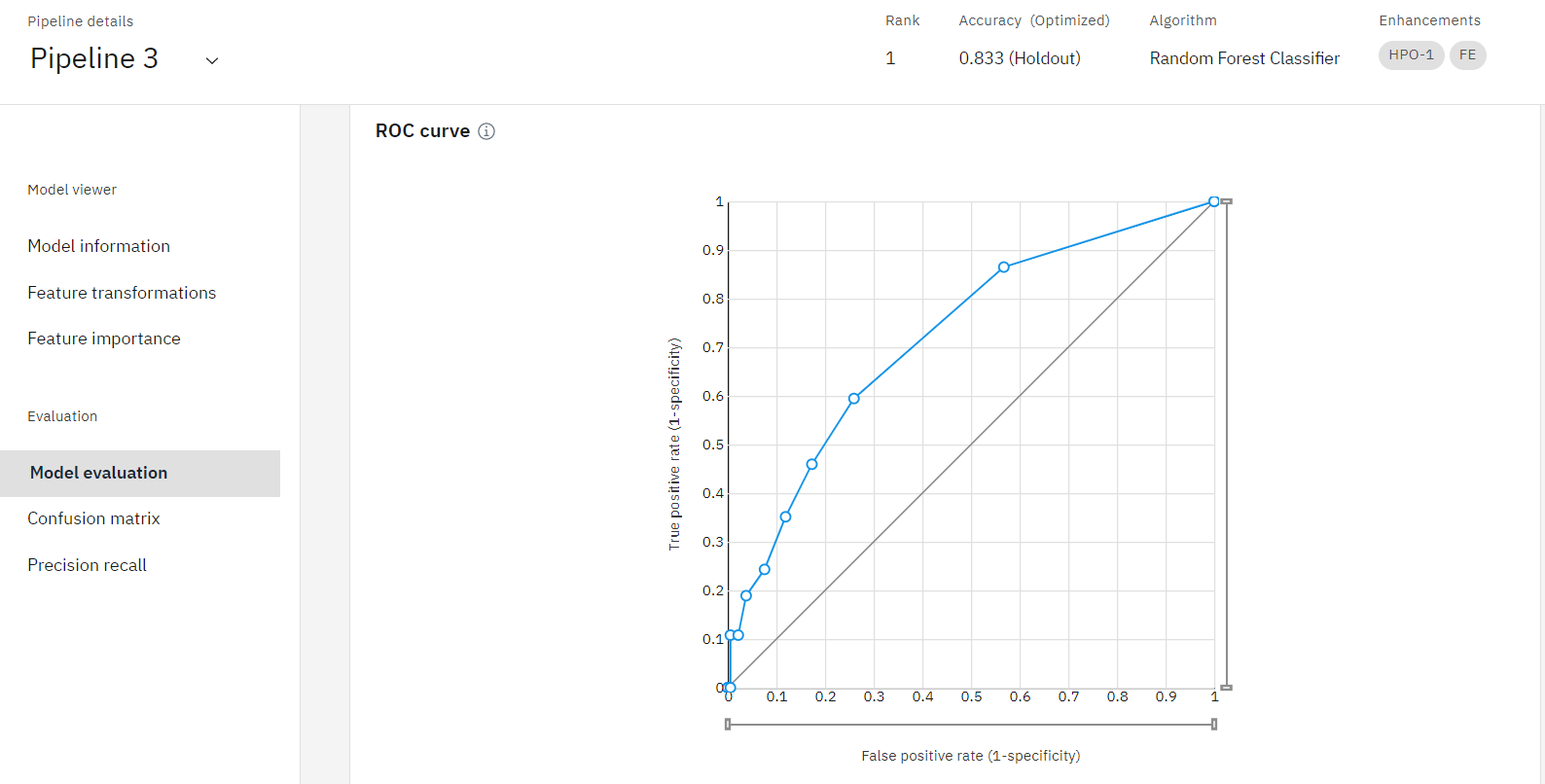


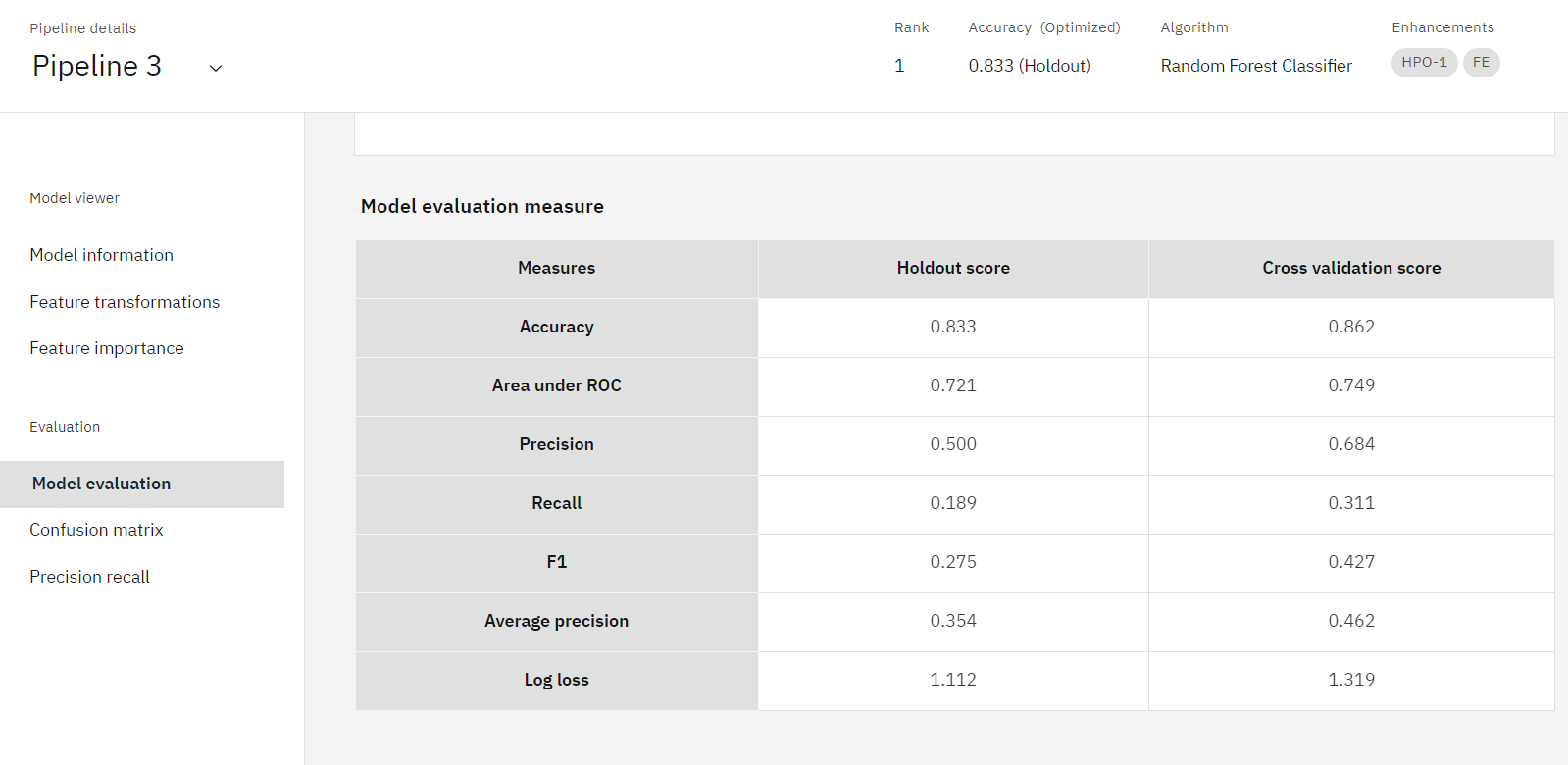


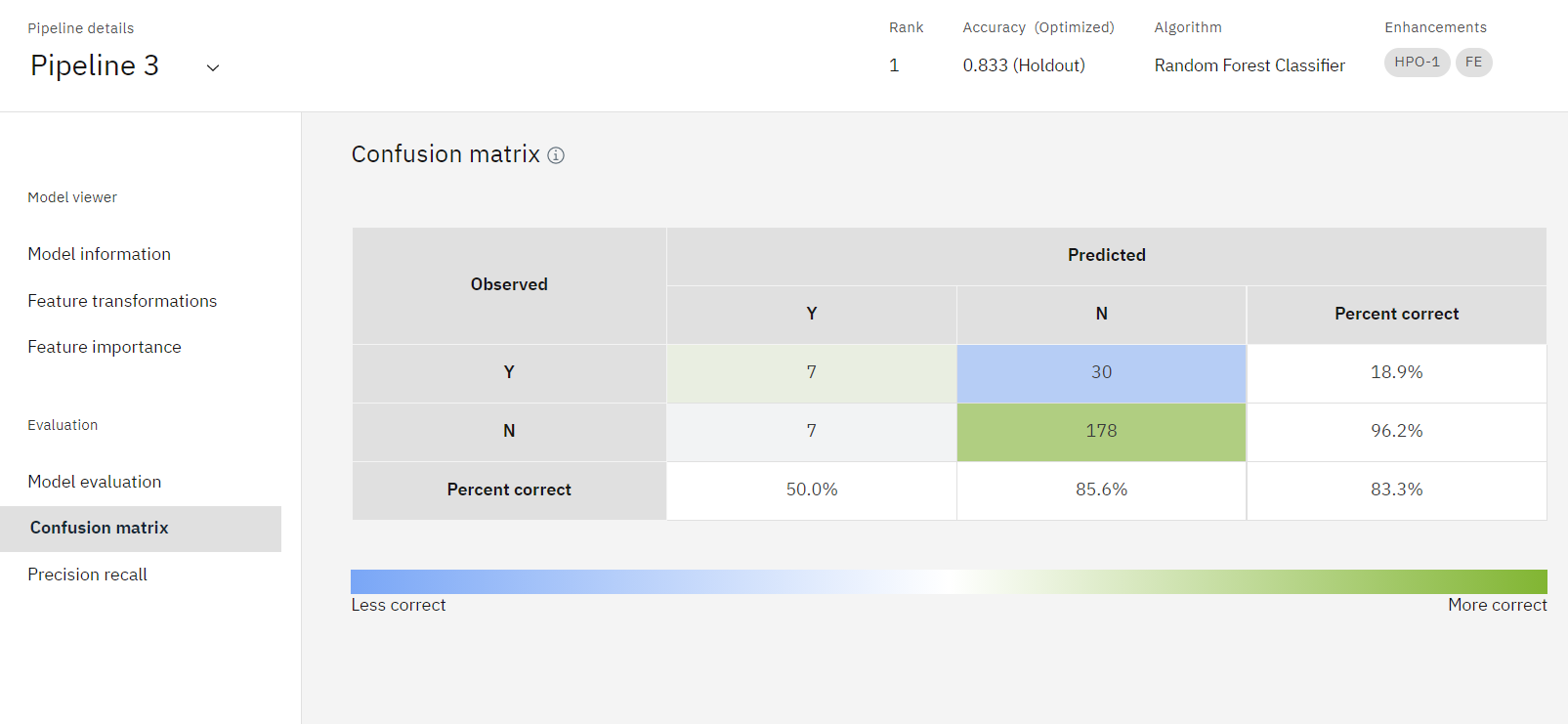
Pipelines

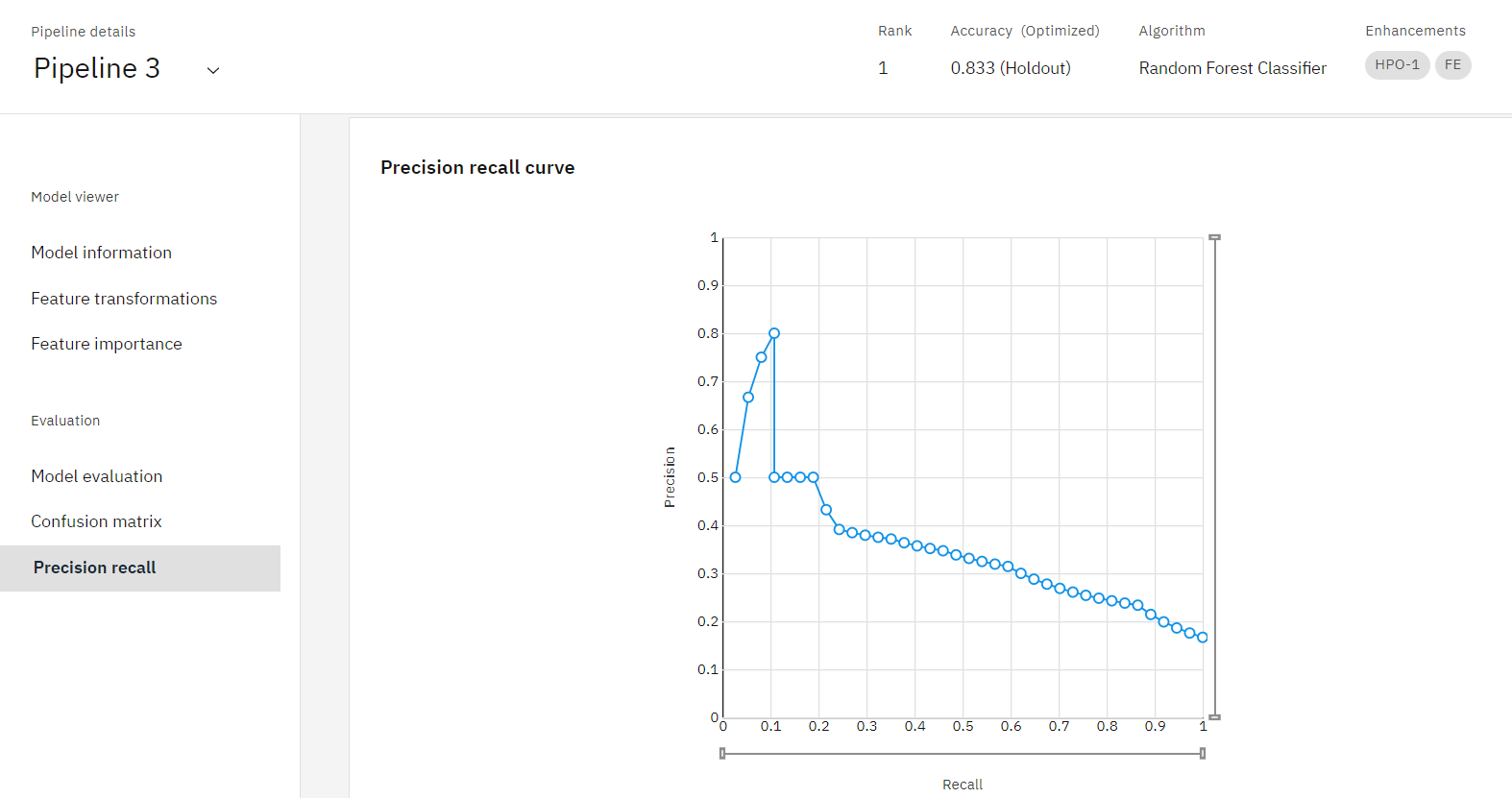


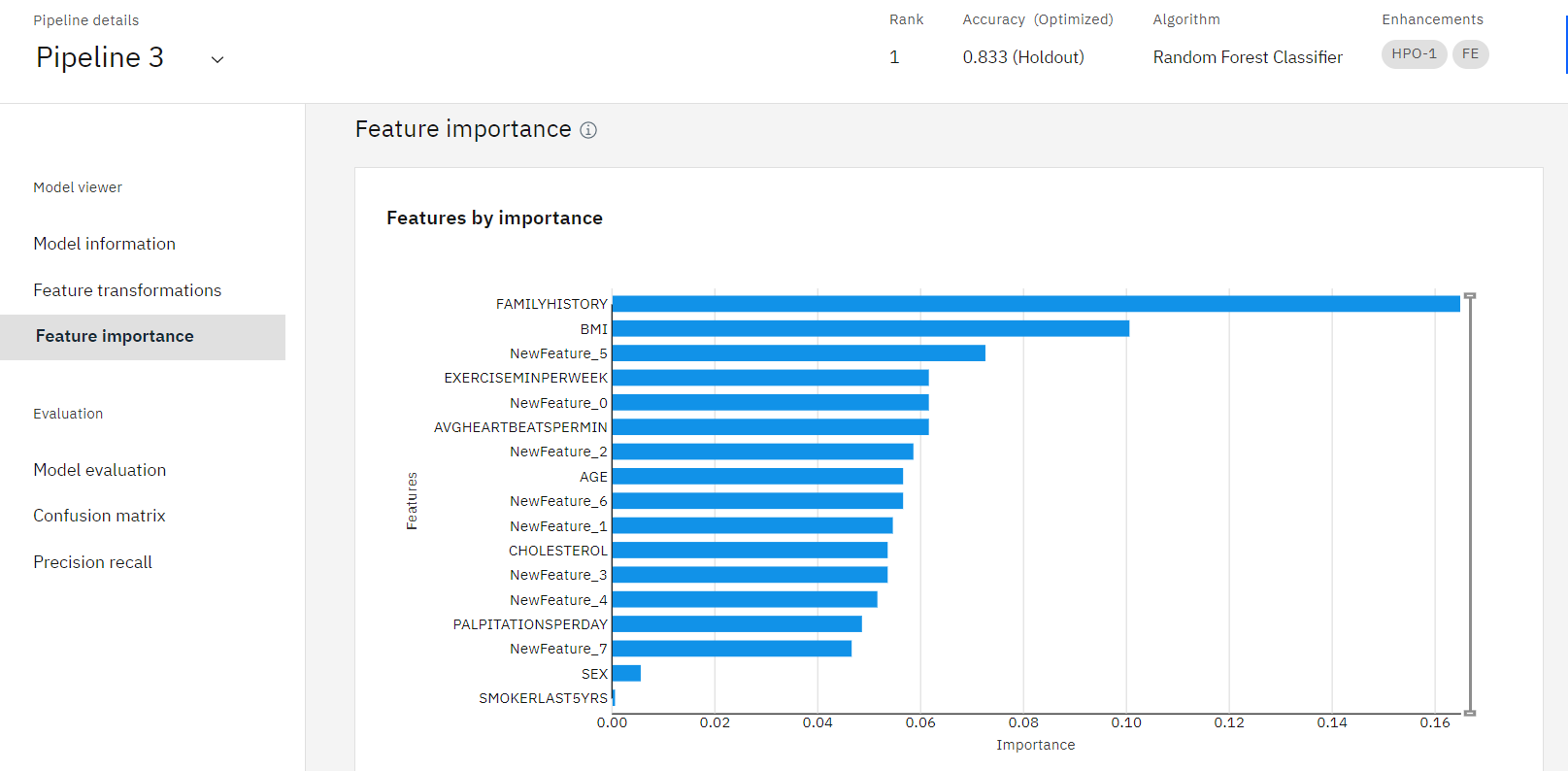




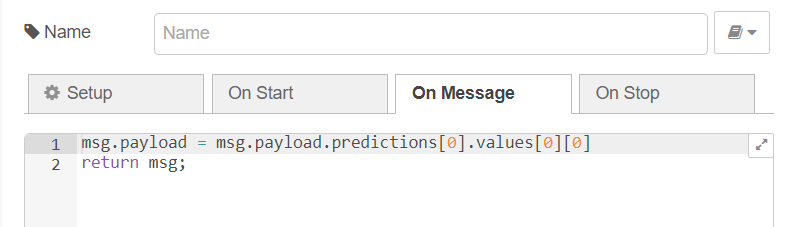
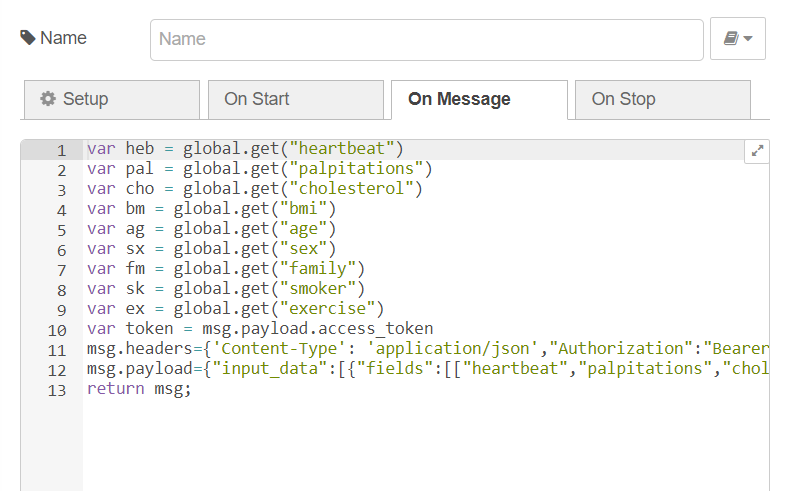
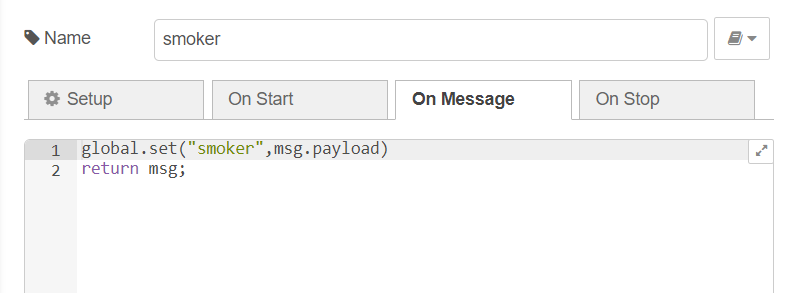
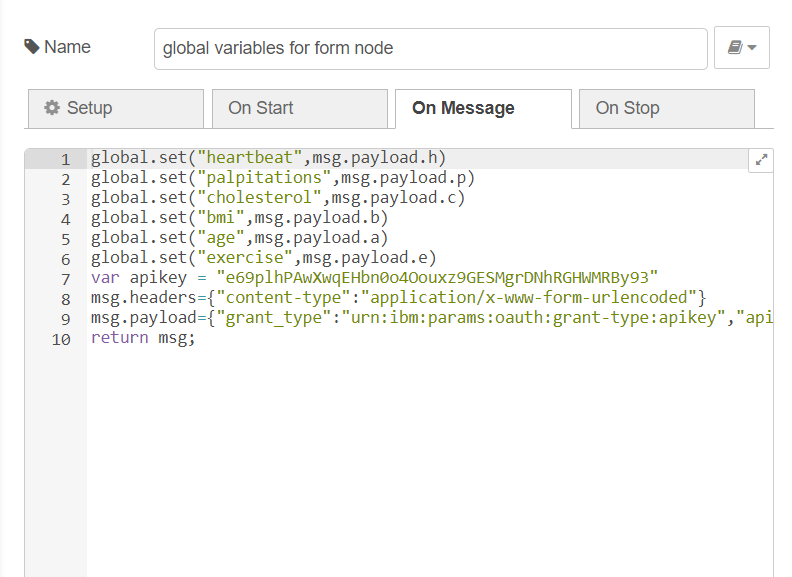


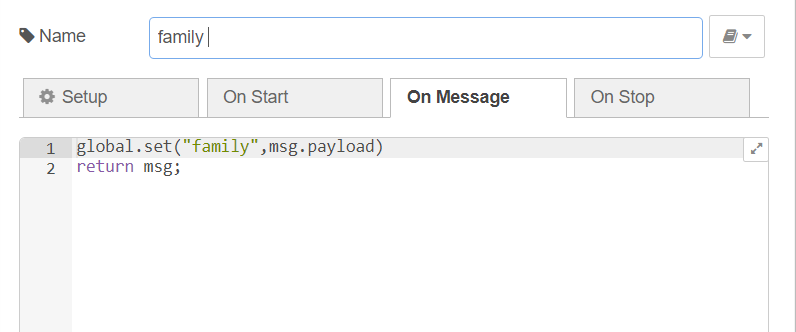
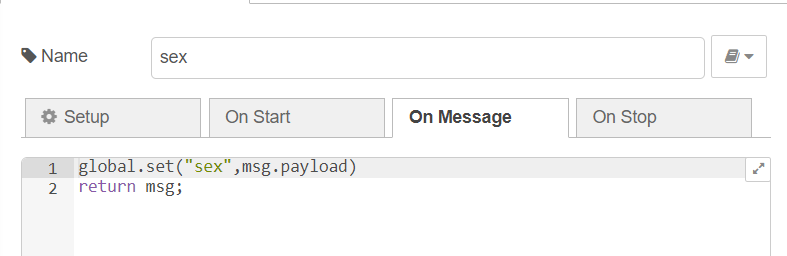






* ***Node-Red***





***UI output Screenshot***

