ABSTRACT

Auto-ML sets out as connection joining the different levels of competence when building Machine learning pipelines or systems and achieve the data science processes more quickly. We present the common Automl tool which works on cleaned datasets using normal Ready-made algorithms provided by sklearn to run against regression and classification datasets and we also use the open source Automl libraries like auto-sklearn,hyperpot,Tpot etc and found that TPOT is best suitable for the regression datasets and auto-sklearn is best suitable for the classification datasets .The auto sklearn library is basically used for running the auto-ml . By using the scikit-learn machine learning library we can train the high level of machine learning algorithms and can find best accuracy easily for the given dataset.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER-1
INTRODUCTION

1.1 OBJECTIVE OF THE PROJECT:

To build an Auto-ML tool for any cleaned dataset so that it applies machine learning algorithm and predicts the accuracy scores for all applied algorithms and also give top three result.

The recent substantial progress in machine learning has led to a growing demand for hands-free ML systems that can support developers and ML novices in efficiently creating the new ML applications. Since different datasets require different ML pipelines, this demand has given rise to the area of automated machine learning.

1.1.1 Necessity: This website helps in overcoming the time management. This Application is very easy to use. It can work accurately and very smoothly in a different scenario. It reduces the effort workload and increases efficiency in work. In aspects of time value, it is worthy.

In this website the user can easily use our auto-ml tool for choosing the best algorithm for the given supervised machine learning data. Our auto-ml tool provides the accuracy scores of all algorithms for classification/Regression data. Then it displays the best three accuracy models for the uploaded dataset. So, by using our auto-ml tool the user can easily find the best suitable model for the data

Hence even the non-coders also easily done machine learning by using our tool.

1.1.2 Software development method:

In many software applications program different methods and cases are followed such as, Waterfall model, Iterative model, Spiral model, V-model and Big Bang model. I used waterfall model in this application. I tried to use test case and case software approaches.

1.1.3 Layout of the document:

This documentation starts with formal introduction. After introduction analysis and design of the project are described. In analysis and design of the project have many parts such as project proposal, mission, goal, target audience, environment. After that design and table diagram will be found. Use cases and test cases are in chapter 2 and chapter 3 respectively. Finally this documentation finished with result and Conclusion part.

1.2   Overview of the Designed Project:

At first we take the dataset from out resource then we have to perform data-preprocessing, visualization methods for cleaning and visualizing the dataset respectively and we upload the cleaned dataset and can run all algorithms easily by clicking button for getting accuracy scores and then it will give best three accuracy scores algorithms and flask is used for user interface.

CHAPTER-2

LITERATURE SURVEY

2.1 LITERATURE SURVEY:

General

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is secondary sources and discuss published information in a particular subject area and sometimes information in a particular subject area within a certain time period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that may be needed in the area and precedes a research proposal and may be just a simple summary of sources. Usually, it has an organizational pattern and combines both summary and synthesis.

A summary is a recap of important information about the source, but a synthesis is a re-organization, reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent or relevant of them.

Review of Literature Survey

Title : Benchmarking Automatic Machine Learning Frameworks

Author: Adithya Balaji, Alexander Allen

Year : 2018

We test auto-sklearn, TPOT, auto_ml, and H2O's AutoML solution against a compiled set of regression and classification datasets sourced from OpenML and found that TPOT is best suitable for the regression type datasets and auto-sklearn is best suitable for the classification type datasets

Title : Efficient and Robust Automated Machine Learning

Author: Matthias Feurer, Aaron Klein, Katharina Eggensperger

Year : 2015

The demand for the machine learning has been increased due to the success of machine learning in various range of applications .For good effective, such systems need to automatically chooses the algorithm and data pre-processing steps for a new dataset and should set their respective hyper-parameters.

Title : Hyper-parameter Optimization of Machine Learning Algorithms
Author: Li Yang, Abdallah Shami
Year : 2020

Currently most of us using machine learning algorithms in different applications and in different areas. For fitting into different problems the machine learning algorithms have to do hyper-parameters must be tuned. There is an direct impact on models performances by selecting the best hyperparameter configuration for ML models. It even also requires more knowledge on ML algorithms and hyper-parameter optimization.

Title : Hyperopt-Sklearn: Automatic Hyperparameter Configuration for Scikit-learn
Author: Brent Komer, James Bergstra, Chris Eliasmith
Year : 2014

Hyperopt-sklearn is the project which gives the auto algorithm configuration for the sklearn ML library. Following Auto-Weka, To represent a single large hyper-parameter optimization problem we can take the choice of large pre-processing modules and the choice of classifier together.

CHAPTER-3

METHODOLOGY

3.1 Project Proposal:

The project proposal is the term of documents. A project can describe the project proposal. It is the set of all plans of a project. Like, how the software works, what are the steps to complete the entire projects, and what are the software requirements and analysis for this project. In my project, I am doing all the steps and also risk and reward and other project dependencies in the project proposal.

3.1.1 Mission:

An online Web based machine learning application is very popular and well known to everyone. We can train high level custom machine learning models with minimal effort and machine learning expertise. Hence even non coder also can do machine learning easily by using our Auto-ML tool. This simple Auto-ML tool gives fast and accurate results for choosing the best model for the given dataset.

3.1.2 Goal:

The goal is to build an Auto-ML tool for choosing the best accuracy model for any given cleaned dataset.

3.2 Scope of the Project:

The scope of this paper is to implement and investigate how different supervised binary classification methods impact default prediction. The model evaluation techniques used in this project are limited to precision, sensitivity, F1-score.

3.3 Overview of the Project:

The overview of the project is to build an Auto-ML tool for choosing the best accuracy model for any given cleaned dataset. So we can do any classification and regression type projects can predict the data and displays the best three accuracy model.
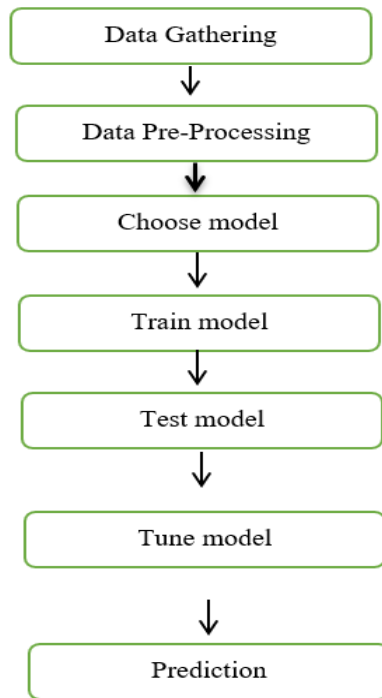
3.4 Flow Chart:

Fig:3.1: Machine Learning workflow diagram

The above flow diagram represents the flow of process from gathering data to predicting result.
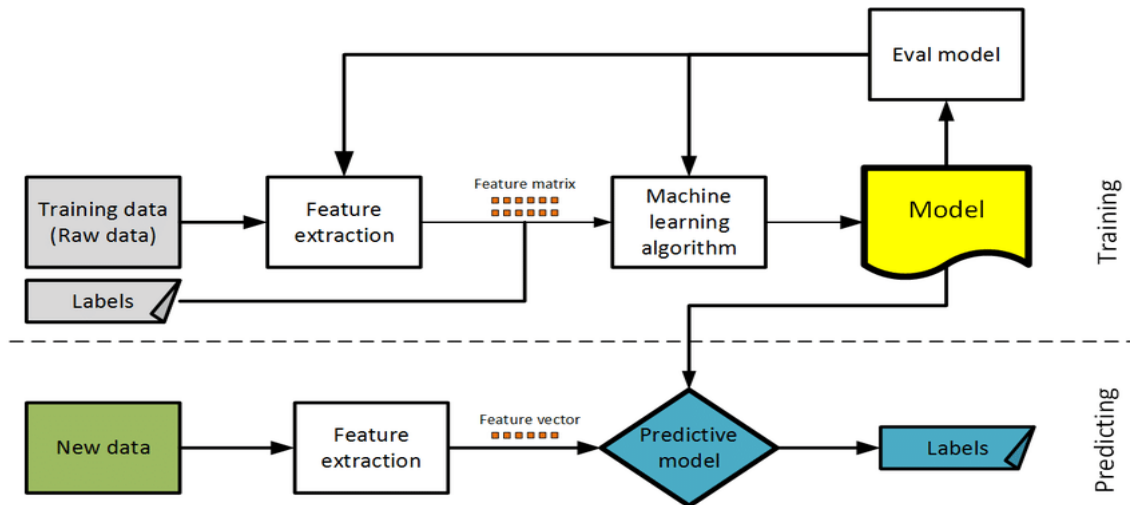

Fig:3.2: System Architecture

Clean the raw data then apply machine learning algorithms and finally predict the results.