International Journal of Advanced Culture Technology Vol.8 No.2 297-304 (2020)
DOI https://doi.org/10.17703/IJACT.2020.8.2.297

IJACT 20-6-38

Analysis of Automatic Machine Learning Solution Trends of Startups

Yo-Seob Lee

Professor, School of ICT Convergence, Pyeongtaek University
E-mail: yslee@ptu.ac.kr

Abstract

Recently, open source automatic machine learning solutions have been applied in many fields. To apply open source automated machine learning to real world problems, you need to write code with expertise in machine learning. Writing code without machine learning knowledge is challenging. To solve this problem, the automatic machine learning solutions provided by startups are made easy to use with a clean user interface. In this paper, we review automatic machine learning solutions of startups.

Keywords: *Machine Learning, Automatic Machine Learning, AutoML*

1. INTRODUCTION

From an application point of view, the demand for machine learning systems has increased over the past few years. Machine Learning (ML) has been adopted in a wide range of applications. However, although machine learning has proven to provide better support for some companies, many companies are still struggling to implement ML model deployments.[1]

To deploy AI models, you must first have a team of experienced data scientists who expect high salaries for your enterprise. Even if you have a good team in your business, deciding which model is best for your business usually requires more experience than AI knowledge. The success of machine learning in a variety of applications has led to a growing demand for machine learning systems, which should be readily available to non-professionals. Automatic Machine Learning tends to automate as many steps as possible in the ML pipeline and maintain good model performance with minimal manpower.[1]

2. AUTO ML

Automated Machine Learning (AutoML) is a process of applying full machine learning pipeline in automatic way. The AutoML solution can do feature preprocessing and eningeering, algorithm training and hyperparameters selection.[1]

AutoML has three main advantages:

- Improve efficiency by automatically running repetitive tasks. This allows data scientists to focus more on problems instead of models.
- The automated ML pipeline helps to avoid potential errors due to manual work.

Manuscript received: April 30, 2020/ revised: May 21, 2020/Accepted: May 28, 2020
Corresponding Author: yslee@ptu.ac.kr
Tel:+82-31-659-8369, Fax: +82-31-659-8011
Professor, School of ICT Convergence, Pyeongtaek University, Korea

- AutoML is a big step towards democratization of machine learning and makes ML functionality available to everyone.

AutoML solutions can be divided into open source products that can be used free of charge and commercial products provided by startups. Open source tools should be written in a few lines of code.The products offered by startups aim to provide a tool that allows users, not technology, to get started with a clean user interface. Some tools also provide visualization capabilities to describe the results and result models. Table 1 shows a list of AutoML Solution Providers.[2]

Table 1. List of AutoML Solutions

Software	Category	Launched in*
Auto-WEKA	Open Source	2013
BigML OptiML	Open Source	
Auto-sklearn	Open Source	
TPOT	Open Source	
Caret	Open Source	
Auger.ai	Startup	2017
TIMi Suite	Startup	
Compellon	Startup	2011
DataRobot	Startup	2012
DMWay Analytics	Startup	2014
Enhencer	Startup	2017
H20.ai	Startup	2012
Kortical	Startup	2016
MLJAR	Startup	2016
PurePredictive	Startup	2011
Tazi.ai	Startup	2015
Xpanse Analytics	Startup	2014

* Date when tool was launched not the day the company was launched.

3. AUTO ML SOLUTIONS OF STARTUPS

3.1 Auger.AI

Auger’s patented Bayesian optimization search of ML algorithm/hyperparameter combinations build the best possible predictive models faster. Auger also provides the most powerful API for AutoML, allowing any developer to build predictive models from their data with no data science background.[3]

The Optimizer Service allows you access to Auger's patented Bayesian optimization search exposed as a web service. Once you have decided on a library or specific model/s you would like to optimize for you will need to define the search space that Auger will traverse. This will include 1 or more algorithms and their hyperparameters. After you've setup your search space you will want to post to the trial_searches endpoint to get your search id. This will be the id you will use for subsequent requests to get the next batch of trials to try.

Now that I have my search id I can begin evaluating trials to find the most accurate model, but before we do that lets load our dataset.[3] Figure 1 shows the process of the Optimizer Service.

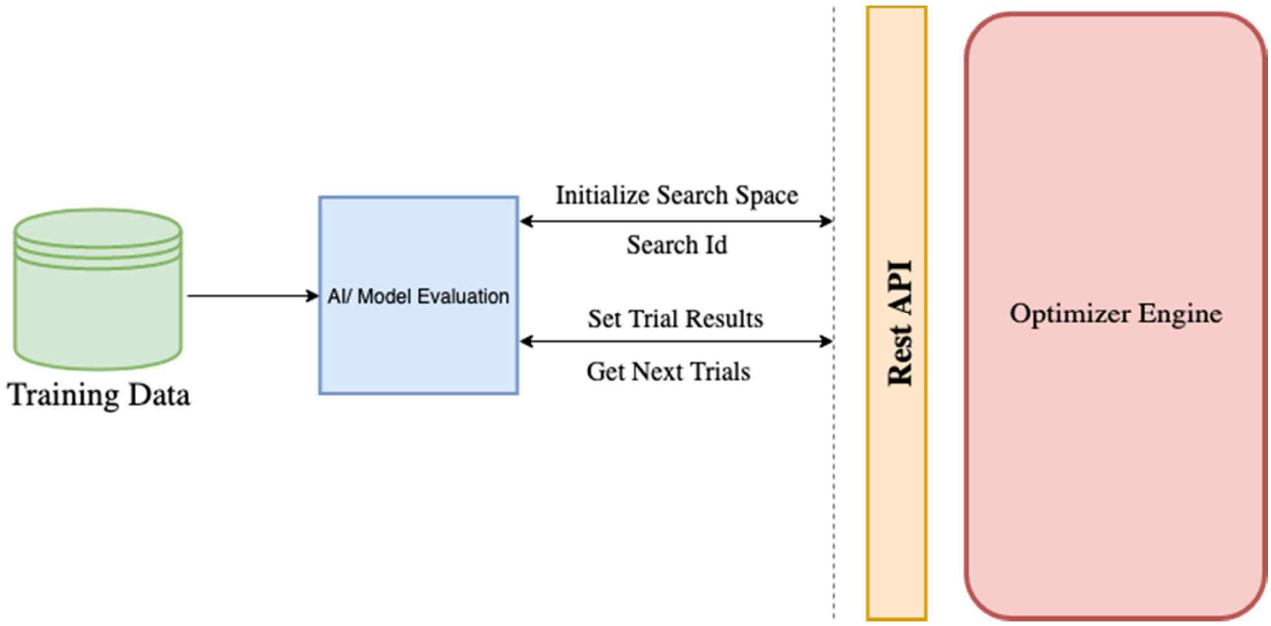

Figure 1. Optimizer Service of Auger.AI

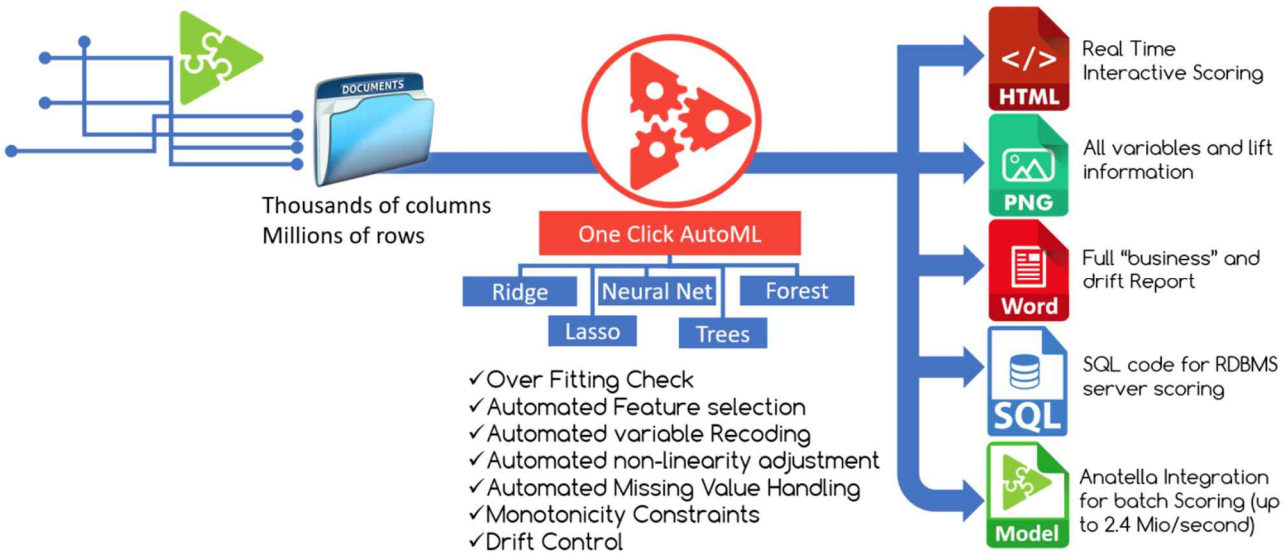

Figure 2. Modeler of TIMi Suite

3.2 TIMi Suite

TIMi Suite consists of Anatella, Modeler, Stardust and Kibella. Anatella makes data management easy and intuitive.[4] With just a few clicks, you can extract data, clean data, join several databases, aggregate data in meaningful KPI's, create new predictive models.

Figure 2 shows the flow diagram of the modeler of TIMi Suite. Modeler is one of the very few modeling tools that provides full automation when creating new predictive models. Everything is automated, from variable selection, sampling and validation, to data audit and model reports in Excel and Word.

Stardust offers you real-time 3D visualization of your whole population. The rendering of the 3D display is performed using hardware-accelerated OpenGL libraries. These high-performance libraries allow the smooth display of millions of customers in real time.

Kibella offers drill-down capabilities, connection to SQLite database and, optionally, connection to the leading SQL databases.

3.3 Compellon

Compellon is rapidly revolutionizing the analytics industry with our Plug and Play AI that brings business analysts the solutions they need. Compellon has transformed existing analytics from assumptions and from subjective processes to objective processes that continue to create value while reducing risk over time.
By reengineering the process, Compellon has eliminated the shortcomings of traditional analytics and has developed an approach that focuses on solving business problems rather than building too many projects and building a model.[5]

Compell's products include Iris Pro, Compellon 20 | 20, and 20 | 20 Survey. Iris Pro is an analytics advisory tool, Compellon 20|20 is a clear-box analytics platform and 20|20 Survey is an unprecedented survey analytics. Figure 3 shows the Iris Pro of Compellon.

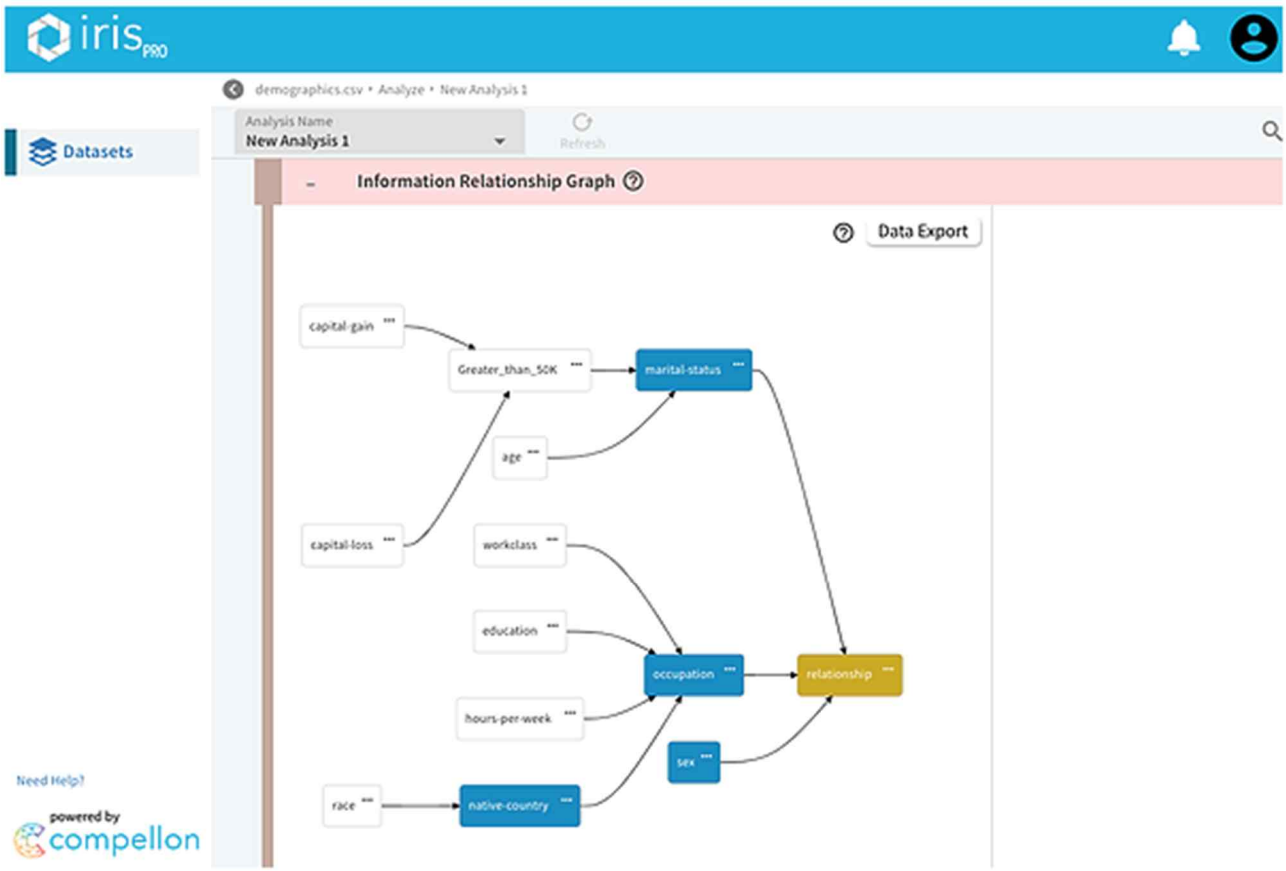

Figure 3. Iris Pro of Compellon

3.4 DataRobot

DataRobot is the only product to automate the entire data science lifecycle from raw data to value. Built-in guardrails at every step ensure best practices are consistently followed – from automated data preparation and feature engineering that gets your data ready, to automatic model selection, training, and testing to recommend the best option for your production environment. DataRobot even automates model deployment, monitoring, and management, so you always get the most accurate predictions possible.[6] Figure 4 shows the process of DataRobot.

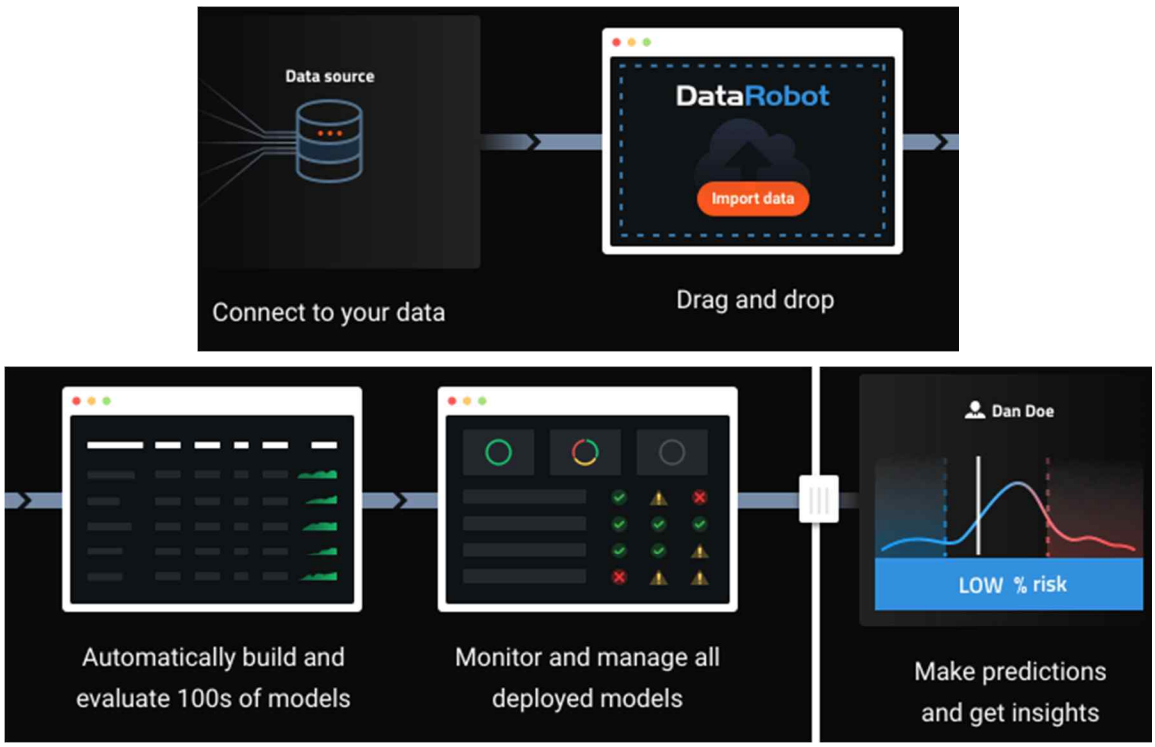

Figure 4. Data Flow of DataRobot

3.5 Kortical

Kortical's competition crushing, cloud scale, distributed AutoML will start finding the best machine learning solution for your dataset. It builds machine learning solutions from the ground up, data cleaning, preprocessing, feature creation, model selection, tuning and much more. Using distributed cloud scale AI it searches the solution space for the best possible machine learning model and surrounding solution.[7]

Kortical offers two methods, coded and codeless. Kortical has a simple high level language for data-science and it generates intuitive high level code that can be refined before running the AutoML. Because Kortical's AutoML is code based, it's really easy to fix parts of the solution and just iterate on the bits you're working on. Kortical makes it easy to separate code from data-science making model updating and management a breeze through a simple UI or API.[7]

3.7 MLJAR

MLJAR is a platform for rapid prototyping, developing and deploying machine learning models. MLJAR makes algorithm search and tuning painless. It checks many different algorithms for you. For each algorithm hyper-parameters are separately tuned. All computations run in parallel in MLJAR cloud, so you get your

results very quickly. At the end the ensemble of models is created, so your predictive model will be super accurate.[8]

Figure 5 shows the project setting screen of MLJAR.

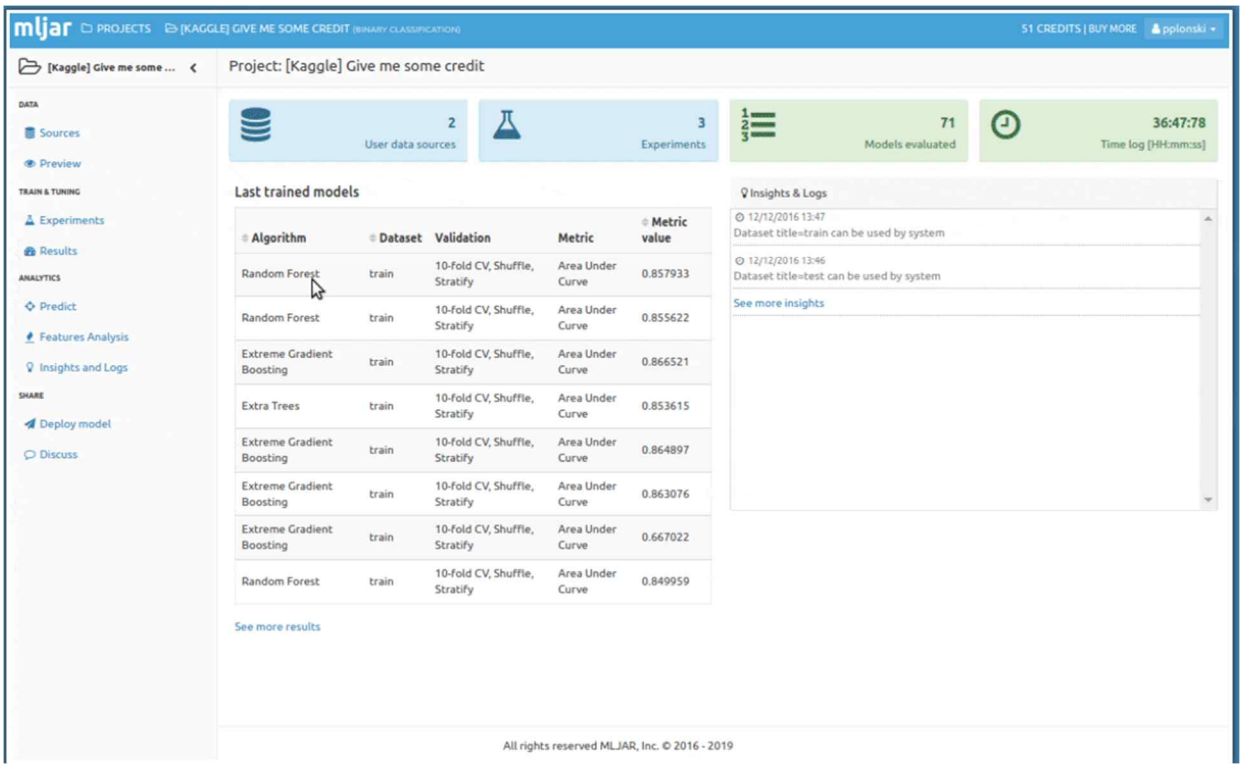

Figure 5. Project Setting of MLJAR

3.9 Tazi.ai

Tazi's Automated Machine Learning is understandable continuous machine learning from data and humans, enables business domain experts to use machine learning to make predictions and take actions. It also helps data analysts and scientists for their daily model creation and deployment.[9]

Data may change in time, Tazi AutoML uses multiple machine learning models, with supervised, unsupervised, semi-supervised algorithms, learning together and combined dynamically. While the batch ML performance degrades in time as data change, tazi continuous ML is able to get better in time. Compared batch ML, robust continuous machine learning results in better accuracy and business benefits.[9]

Tazi platform provides SaaS automated Machine Learning (AutoML) solutions for insurance, banking & finance, retail and telecoms industries. Tazi platform is composed of Tazi DEPLOY, Tazi HUNT and Tazi LIVE. Tazi DEPLOY with its easy to use interface allows users to create and deploy models. Tazi HUNT work together in a continuous way to generate superior models. Tazi LIVE allows users to visualize the model outputs, review model performance and explanations.[10]

3.10 Xpanse AI

Xpanse AI is an automated data science platform delivering Predictive Models from complex databases in minutes. Autonomous feature discovery and data integration engine of Xpanse AI automatically prepares

Modelling Datasets in minutes, saving you weeks of manual coding. It uses the integrated ML algorithms or exports the modelling dataset and experiment with ML in another tool.[11]

Figure 6 shows the automated feature engineering of Xpanse AI.

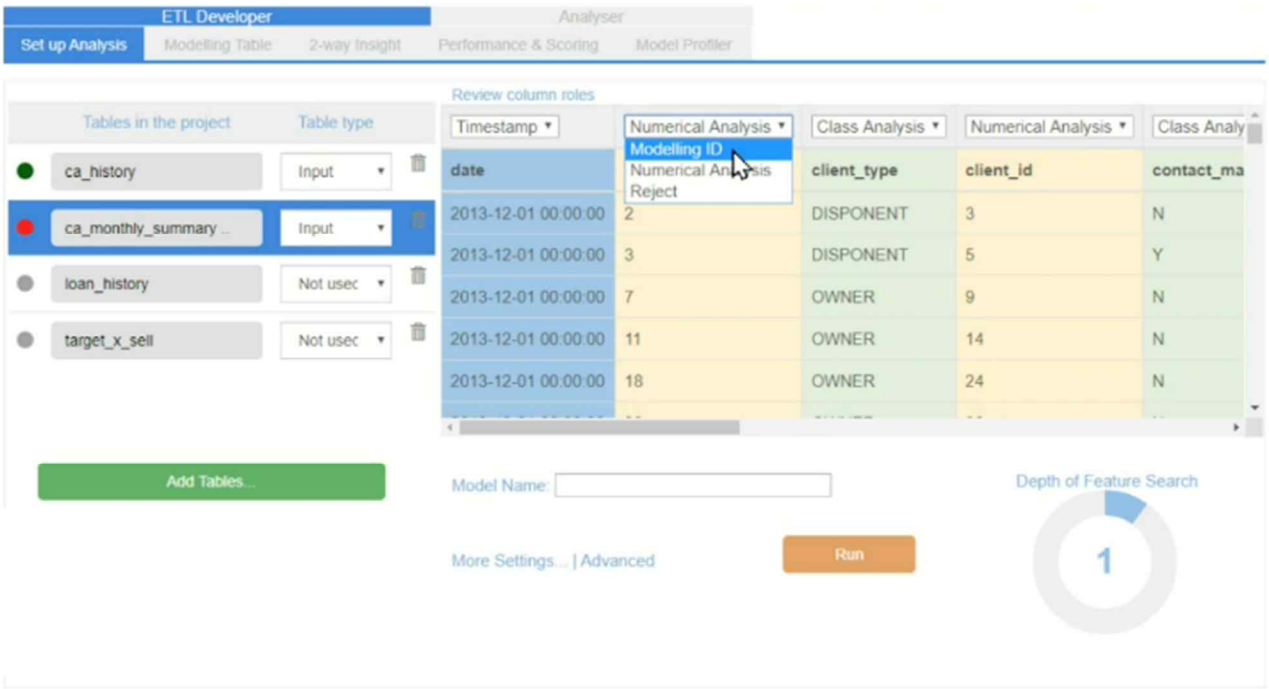

Figure 6. Automated Feature Engineering of Xpanse AI

4. ANALYSIS OF AUTO ML SOLUTIONS

In this section we compare the features of the Auto ML solutions by startups.

Table 2 shows the interface of the Auto ML solutions. Most solutions provide an easy-to-use interface. Most solutions provide a web-based interface for easy use, and allow you to write code through the Python API.

Table 2. Interface of AutoML Solution

Name	Programming Interface	User Interface
Auger.AI	Python	Web
TIMi Suite	R, Python, JavaScript	
Compellon		Standalone app
DataRobot		Standalone app
Kortical	R, Python	Standalone app
MLJAR	R, Python	Web
Tazi.ai		Standalone app
Xpanse AI		Standalone app

Table 3 shows the features of AutoML solutions. Most solutions provide function that are easy to use.

With just a few clicks in TIMi Suite, you can extract data, clean data, join several databases, aggregate data in meaningful KPI's, create new predictive models. Compellon works through Smart Plug and Play AI. DataRobot is easily used through Machine Learning Operations (MLOps).

Some solutions offer a variety of services. Auger provides an open source A2ML framework which automates all essential parts of the machine learning process: from training to deployment to prediction. Tazi.ai provides AutoML solutions as SaaS.

Table 3. Features of AutoML Solution

Name	Feature
Auger.AI	Optimizer Service, Open source A2ML framework
TIMi Suite	Easy and Intuitive Data Management Real-time 3D visualization
Compellon	Smart Plug and Play AI
DataRobot	Time Series Automation, Machine Learning Operations (MLOps)
DMWay Analytics	Upload your data files, set your desired goals, run - in just 3 clicks
MLJAR	binary classification, regression
Tazi.ai	Continuous Machine Learning, SaaS AutoML solutions
Xpanse AI	Autonomous Feature Discovery and Data Integration

5. CONCLUSION

In this paper, we review AutoML solutions of startups, and compare various characteristics of AutoML solutions of startups.

It has been found that Startup's AutoML solutions provide a variety of clean user interfaces that can be easily used without knowledge of machine learning, and various programming interfaces that can be easily adjusted and used by experts.

Through review of AutoML solutions of startups, we expect features and methods that can be easily used without expert knowledge, and will be able to apply AutoML solution to more fields.

REFERENCES

[1] 6 Top AutoML Frameworks for Machine Learning Applications, *https://www.alibabacloud.com/blog/6-top-automl-frameworks-for-machine-learning-applications-may-2019_595317*
[2] AutoML Software / Tools in 2020: In-depth Guide, *https://blog.aimultiple.com/auto-ml-software/*
[3] Auger.AI, *https://auger.ai/*
[4] Business Insight's TIMi Suite, *https://timi.eu/*
[5] Compellon, *https://www.compellon.com/*
[6] DataRobot, *https://www.datarobot.com/*
[7] Kortical, *https://kortical.com/*
[8] MLJAR, *https://mljar.com/*
[9] Tazi.ai, *https://www.tazi.ai/*
[10] Tazi AutoML Platform, *https://www.tazi.ai/automl-platform/*
[11] Xpanse Analytics, *https://xpanse.ai/*