International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

ABLE - Machine Learning for Everyone

Kanchan Katake[1], Manan Gaur[2], Manish Patel[3], Ritik Mahyavanshi[4], Yashwardhan *Singh* Chundawat[5]

[1]Professor, Dept. Information Technology, MIT School of Engineering, Maharashtra, India
[2-5]B.Tech student, Dept. Information Technology, MIT School of Engineering, Maharashtra, India

---

Abstract - *In this project we have worked on the emerging technology, Automated Machine Learning (AutoML). We believe this technology will change how we look at machines. We very well realize the effectiveness, and the positive change that this technology will bring to the market. well, this is also realized by some technology giants like Google and Amazon for them to follow or maybe make this a trend. We have certainly analysed some of the AutoML platforms developed by these giants and as well some the greatest platforms available today, and we have deeply learned how these platforms function and what their pros and cons are. AutoML for us means deploying Machine Learning models in an autonomous manner, where the tedious tasks in Machine Learning can be automated, so as to save time and costs. We have deeply studied some of the best research papers on AutoML and some very necessities of AutoML, so in order to get to a conclusion of what works the most efficiently, and to develop a greater understanding of AutoML, so for us to implement AutoML the best way possible.*

Key Words: **AutoML, Machine Learning, Deep Learning, Emerging Technology**

1.INTRODUCTION

Automated Machine Learning (AutoML) provides strategies and processes to create Machine Learning offered for non-Machine Learning consultants, to enhance the potency of Machine Learning, and to accelerate analysis on Machine Learning.

Machine Learning (ML) has achieved significant successes in recent years Associate in nursing an ever-growing variety of disciplines have faith in it. However, this success crucially depends on human-machine learning consultants to perform the subsequent tasks:

• Preprocess the information,

• Select applicable options,

• Select associate applicable model family,

• Optimize model hyperparameters,

• Postprocess machine learning models,

• Critically analyze the results obtained.

As the complexness of those tasks is usually on the far side of non-ML-experts, the rapid climb of machine learning applications has created a requirement for ready-to-wear machine learning ways which will be used simply and while not professional data. We tend to decision the ensuing analysis space that targets progressive automation of machine learning AutoML. As a replacement sub-area in machine learning, AutoML possesses additional attention not solely in machine learning however conjointly in pc vision, tongue process, and graph computing.

2. Literature Survey

2.1 A Survey of the State-of-the-Art

In the growing market of AutoML, this paper focuses on the in-depth methodologies of AutoML. It relies on the best available practices of AutoML and its deep mechanics and further accustoms to a specific technique called NAS (Neural Architecture Search). This paper then lays out the open problems of the existing AutoML methods and the scope for future research. This paper does not include any solution to a problem per se, it rather lays down the preferred steps being used in AutoML today, (example, Data Preparation, Feature Engineering, Model Generation, Model Estimation), then it summarized the performance of NAS (Neural Architecture Search) and argues how it is better.

2.2 Benchmark and Survey of Automated Machine Learning Framework

In this paper, AutoML is referred to as something that aims to reduce the demand of Data Scientists, by enabling domain experts to automatically build machine learning applications without extensive knowledge of statistics and machine learning. This paper includes a collection of surveys on the most popular platforms of AutoML on real datasets. The paper then summarizes important AutoML techniques and methods counting every step in developing an ML pipeline. The selected AutoML frameworks are evaluated on 137 data sets. This paper compares several different popular AutoML frameworks and then compares them all on different metrics like accuracy, and time is taken and averaged pair-wise Levenshtein ratio. This paper claims to provide the first empirical evaluation of CASH algorithms on 114 publicly available real-world data sets. This paper has conducted the largest evaluation of AutoML frameworks in terms of considered frameworks as well as a number of data sets. Important techniques used by those frameworks are theoretically introduced and summarized. This way, this paper presented the most important research for automating each step of creating an ML pipeline. Finally, extended current problem formulations to cover the complete process of building ML pipelines.

2.3 Analysis of the AutoML Challenge Series 2015–2018

This paper focuses on the AutoML challenge series that consisted of six rounds of a machine learning competition of progressive difficulty, subject to limited computational resources. It was followed by a one-round AutoML challenge (PAKDD 2018). This paper also analyses the results from the various challenges that has happened. This paper than describes the various methods that were used in the challenges. These methods enlists as:

- Ensembling: dealing with over-fitting and any-time learning model evaluation: cross-validation or simple validation
- Model space: Homogeneous vs. heterogeneous
- Search strategies: Filter, wrapper, and embedded methods
- Multi-level optimization
- Time management: Exploration vs. exploitation tradeoff
- Preprocessing and feature selection
- Unsupervised learning
- Transfer learning and meta learning
- Deep learning
- Task and metric optimization
- Engineering
- Parallelism

2.4 Evolutionary Neural AutoML for Deep Learning

This paper is in the context of Deep Neural Network (DNN), and it provides focus on the various problems in DNN methodologies like the success of DNNs depends on the proper configuration of its architecture and hyperparameters, and such a configuration is difficult and as a result, DNNs are often not used to their full potential. In addition, DNNs in commercial applications often need to satisfy real-world design constraints such as size or number of parameters. This solution proposed by the author is Automated Machine Learning (AutoML) itself. As to make configuration in handy, automatic machine learning (AutoML) systems for deep learning have been developed, mainly for the optimization of hyperparameters. This paper takes AutoML a step further. It initiates a progressive AutoML framework called LEAF that not only optimizes hyperparameters but also network architectures and the size of the network. This paper showed that LEAF can outperform existing state-of-the art AutoML systems and the best hand-designed solution.

2.5 Taking the Human out of Learning Applications: A Survey on Automated Machine Learning

The key problem addressed in this paper is to take out the human from the machine learning application. Automated model selection Auto-sklearn, neural architecture search Google's Cloud, automated feature engineering are some of the attempts to solve it as has been described in the paper. The author's idea of AutoML may be stated as -With the help of AutoML we can enable faster deployment of machine learning solutions across organizations, efficiently validate and benchmark the progress of deployed solutions, and make the centre focus more on problems with more application and business values. These would make machine learning much more comprehensive and accessible for real-world usages, leading to new levels of competence and contrive, of which the impression can be indeed dramatic. The paper features the existing work and serve as a good guideline not only for beginners' usage but also for future researches. AutoML focuses on how to do machine learning in a special way (Definition 2), the current trends in machine learning can also be seen as future works of AutoML. Examples are human interpretability and privacy in machine learning. 1, one important future work is to automatically create features from the data. 2, how much training data is sufficient? What general bounds can be found to relate the confidence of the learned hypotheses to the amount of training experience and characters of learner's hypothesis space?

3. Background/Motivation

AutoML, with its ability to perform data pre-processing, ETL tasks, and transformation, can seemingly become the foremost well-liked trend for the approaching years. With the arrival of massive knowledge, advanced analytics, and prognosticative models, data scientists are expected to possess a lot of talent and updated skills once it involves handling computing and machine learning.

However, bridging the abilities gap, the opposite facet of the herd has not solely been able to survive but are capable of building models, victimization the simplest diagnostic and prognostic analytics tools.

According to a report, the information explosion within the world goes to extend multiple, therefore the world of analytics, AI, machine learning and information science can see a wave of data training.

And, with the increasing amount of data, here's why AutoML might be the most used technology in the coming years. 'Able' aims to reduce/ eliminate the need for skilled data scientists to build machine learning models. The system allows layman to provide labelled training dataset as input and receive an optimized model as output.

The scope of Machine Learning is ENDLESS, we just need to figure out the channels of provision, and we will be leading the future!

4. ABLE (Our Solution)

We intend to bring basic ML to the doorstep of any individual who can operate a computer device.

It can be seamlessly extended to following use cases–

An online store facilitating you the powerful ML.

An offline store that provides the large, and medium scale corporations the high end ML technologies.

An offline store that provides small, and micro scale businesses the key to unleash the ML technology.

Here we discuss the key factors that should be kept in mind while working on a project:

4.1 Target Audience

While developing an AutoML project or any project for that matter, it is super important to know the target audience. To keep ourselves at their position and ask the questions from ourselves those the target audience may. Keeping this in mind all the time keeps the project in line with the requirements. And as we constantly say – we intend to bring basic ML to the doorstep of any individual who can operate a computer device.

It can be anyone – the school Teacher, the small store owner, a salesman anyone. So, we need to keep them in mind while creating the project.

4.2 Open-Source nature?

There exists already lots of AutoMLs in the market but how much of them are people generally aware of? – very few I'd say. Because they are either costly or they require the knowledge of ML for being operated.

4.3 User Friendliness

This is something that's missing in the existing AutoMLs. For a layman who doesn't even know about ML to use this tool – it requires a next level of user friendliness. With 'able', we intend to make it really intuitive and user friendly so that a layman can enjoy the benefits of using machine learning without the need to really learn about it.

5. Challenges

5.1 Data Pre-processing

This is something which will appear in the challenges section of almost every AutoML research paper. Automating Data Pre-processing has been really painful and it still can not be trusted upon. Human intervention is required to get the optimal results.

5.2 Feature Engineering

What we spoke about Data pre-processing is true for Feature engineering as well. It's been struggling to get Automated too and it's still the humans are better at than the machine doing itself. In order to get the optimal results and better performance – human intervention is still required.

6. Conclusion

The thorough analysis of the project ideation has been done completely, and the implementation of the project has been done more than a quarter. Deep and meaningful insights have been gained by the Literature Survey review process, and the implementation and several other tweaks in the project have been made accordingly, vigilantly. The high quality experience have been gained through working in an environment of AGILE and some other useful Software Development methodologies. As well, we had some phenomenal takeaways in learning. We got to learn and perceive more about machine learning, and how advanced is it going to become in the coming future, we can expect such a growth in machine learning that in the coming future we may not have to manually set any algorithms, and previously built algorithm would be well capable of setting and developing new algorithms on its own. We learned hugely about ensemble learning, and a various different procedures to follow the most efficient methods. Data cleaning has been one of the most important factors of our development. We got to spend enormous time figuring out the best techniques for data cleaning. We had been very observant about some of the best AutoML products in the market. We specifically, learned a great deal about H2O AutoML, and Microsoft's Azure, where we found out that, H2O is the most efficient AutoML till date and gives the best accuracy for any type of datasets, and Azure being the most popular AutoML platform and has the most wide variety of AutoML features. Implementation of several of the processes like, Feature selection, and Ensemble Learning has been done sincerely, with high quality results. The designing of the Front End Interactive Interface has been successfully done. The further implementation of the processes of Usability Design, development of the Front End, Integration, and Testing are to be done. As the project goes on, the further extension on the Literature Survey Review is to be done effective and the changes are to be adapted accordingly.

7. References:

1.) Title: AutoML: A Survey of the State-of-the-Art

https://arxiv.org/pdf/1908.00709.pdf

2.) Title: Benchmark and Survey of Automated Machine Learning Frameworks

https://arxiv.org/pdf/1904.12054.pdf

3.) Title: Analysis of the AutoML Challenge Series 2015–2018

https://library.oapen.org/bitstream/handle/20.500.12657/23012/1007149.pdf?sequence=1#page=180

4.) Title: Evolutionary Neural AutoML for Deep Learning

https://arxiv.org/pdf/1902.06827.pdf

5.) Practical Automated Machine Learning for the AutoML Challenge 2018

http://www.tnt.uni-hannover.de/papers/data/1407/18-AUTOML- AutoChallenge.pdf

6.) https://arxiv.org/pdf/2008.08516.pdf

7.) https://arxiv.org/pdf/1907.00909.pdf

8.)https://www.researchgate.net/publication/339263193_TPOT-SH_A_Faster_Optimization_Algorithm_to_Solve_the_AutoML_Problem_on_Large_Datasets

9.) https://arxiv.org/pdf/1906.02287.pdf

10.)https://arxiv.org/pdf/1810.13306.pdf