



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 3, March 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Development of Auto-ML Tool

Chahal Vadalmiya, Nisha Gandhi, Gauri Khotele, Ansh Mishra, Prof. Rohini Jadhav

Department of Computer Engineering, Smt.Kashibai Navale College of Engineering, Pune, India

Department of Computer Engineering, Smt.Kashibai Navale College of Engineering, Pune, India

Department of Computer Engineering, Smt.Kashibai Navale College of Engineering, Pune, India

Department of Computer Engineering, Smt.Kashibai Navale College of Engineering, Pune, India

Associate Professor, Department of Computer Engineering, Smt.Kashibai Navale College of Engineering, Pune, India

ABSTRACT: Automated machine learning (AutoML) spans the fairly wide chasm of tasks which could reasonably be thought of as being included within a machine learning pipeline. An AutoML "solution" could include the tasks of data preprocessing, feature engineering, algorithm selection, algorithm architecture search, and hyperparameter tuning, or some subset or variation of these distinct tasks. Thus, automated machine learning can now be thought of as anything from solely performing a single task, such as automated feature engineering, all the way through to a fully-automated pipeline, from data preprocessing, to feature engineering, to algorithm selection, and so on.

KEYWORDS: AutoMachine Learning, Support Vector Machine, Auto-ML Tool, Hyperparameter Tuning, Data preprocessing, Feature Engineering, Data augmentation, Scalability, Regression.

I.INTRODUCTION

AutoML can be defined as a set of tools that can automate the process of solving problems with Machine Learning. The "Automated" here apparently means that the normal stages of the machine learning process are built into a fully programmed sequence of tasks so that the layman can do it without the need to understand the complicated steps.

The goal behind the use of AutoML is that it would allow anyone to apply machine learning models and techniques even if they don't have any expertise in machine learning or coding which doesn't mean it can't be used by an experienced data scientist; AutoML can make it easier and faster for them to do their work, enabling them to accurately scale their efforts while focusing on what matters.

The process of developing a machine learning model can be complex and time-consuming, involving multiple steps, such as data cleaning, feature engineering, model selection, hyperparameter tuning, and evaluation. AutoML simplifies this process by automatically selecting the best model, tuning the hyperparameters, and evaluating the performance metrics. This enables data scientists and other professionals to focus on higher-level tasks such as problem definition, data collection, and data analysis.

ADVANTAGES

There are several advantages of using an AutoML (Automated Machine Learning) tool, including:

1. **Efficiency:** AutoML tools automate the machine learning pipeline, reducing the time and effort required to develop accurate and effective models. This enables data scientists and other professionals to focus on higher-level tasks such as problem definition, data collection, and data analysis.
2. **Scalability:** AutoML tools can process and analyze vast amounts of data with greater accuracy and efficiency, making it possible to build models that can handle large datasets and complex problems.
3. **Reduced Human Error:** AutoML tools reduce the risk of human error by automating the machine learning pipeline, minimizing the potential for mistakes in data cleaning, feature engineering, model selection, and hyperparameter tuning.
4. **Improved Model Performance:** AutoML tools can automatically select the best model, tune hyperparameters, and evaluate performance metrics, leading to more accurate and effective models.

5. **Cost-Effective:** AutoML tools can reduce the cost of machine learning development by automating time-consuming and repetitive tasks, enabling data scientists and other professionals to focus on higher-level tasks that require their expertise.
6. **Flexibility:** AutoML tools can be customized to meet the specific needs of different industries and applications, enabling organizations to tailor the machine learning pipeline to their unique requirements.

DISADVANTAGES

While there are many advantages to using an AutoML (Automated Machine Learning) tool, there are also some potential disadvantages to be aware of:

1. **Limited Control:** AutoML tools automate many aspects of the machine learning pipeline, which means that users may have limited control over certain parameters, such as feature selection and model architecture. This can make it difficult to fully customize the machine learning process to meet specific needs.
2. **Lack of Transparency:** AutoML tools can be complex and opaque, which may make it difficult to understand the underlying algorithms and decisions that the tool is making. This can make it harder to debug and troubleshoot problems that may arise during the machine learning process.
3. **Overfitting:** AutoML tools can sometimes overfit the data, meaning that the model performs well on the training data but does not generalize well to new data. This can be especially problematic when working with small datasets.
4. **Dependency on Data Quality:** AutoML tools rely heavily on the quality and cleanliness of the input data. If the data is noisy, incomplete, or biased, the resulting model may be less accurate or effective.
5. **Limited Compatibility:** AutoML tools may not be compatible with all types of data or models. Some tools may only support specific types of machine learning algorithms or require certain data formats, which can limit their usefulness in certain applications.
6. **Lack of Expertise:** AutoML tools may give users a false sense of expertise, leading them to rely on the tool without fully understanding the underlying concepts and principles of machine learning.

II.PROPOSED MECHANISM

The proposed mechanism of an AutoML tool generally involves the following steps:

1. **Data Preprocessing:** The first step is to preprocess the data to clean, transform, and normalize it for machine learning. This step involves feature engineering, data cleaning, data normalization, and other data transformations.
2. **Model Selection:** The next step is to select an appropriate machine learning model that is best suited for the given dataset. The tool can provide suggestions based on the type of problem and dataset characteristics.
3. **Hyperparameter Tuning:** The machine learning model must be tuned for optimal performance by adjusting its hyperparameters. This step is critical to improving the accuracy of the model and can be time-consuming if done manually.
4. **Model Evaluation:** The model must be evaluated on various performance metrics such as accuracy, precision, recall, and F1 score. This step is necessary to validate the effectiveness of the model and to identify areas for improvement.
5. **Deployment:** The final step is to deploy the trained model in a production environment. The tool can provide options for exporting the model to various formats and integrating it into other software systems.

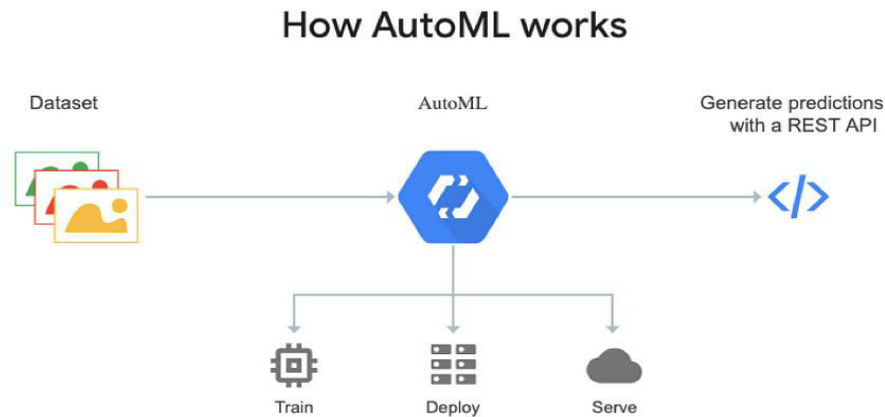


Fig. 1: Auto-ML

III.FUTURE SCOPE

The future scope of AutoML (Automated Machine Learning) tools is highly promising as they continue to evolve and improve in several key areas. Firstly, these tools are becoming more user-friendly with intuitive interfaces that allow users to build and deploy machine learning models without extensive programming or data science knowledge. Secondly, AutoML tools are increasingly being integrated with cloud services such as AWS, Azure, and Google Cloud Platform, which makes it easier to deploy and scale machine learning models on the cloud.

Another significant improvement in AutoML is the incorporation of automated feature engineering. This can help in identifying and extracting relevant features from raw data, reducing the time and effort required to build effective machine learning models. AutoML tools also incorporate automated data cleaning, which can identify and correct errors, remove outliers, and impute missing values, further streamlining the machine learning pipeline.

Furthermore, there is growing interest in developing AutoML tools that are more transparent and explainable, enabling users to better understand the underlying algorithms and decisions made by the tool. This would help in developing a deeper understanding of machine learning and provide better insights.

Overall, the future of AutoML looks bright, with continued improvements in usability, scalability, and performance. As these tools become more advanced and accessible, we can expect to see even greater adoption of machine learning across a wide range of industries and applications.

IV.CONCLUSION

The practice of data science is rapidly becoming automated. New techniques developed by the artificial intelligence and machine learning communities are able to perform data science work such as selecting models, engineering features, and tuning hyperparameters. Sometimes, these automated methods are able to produce better results than people. Given the current shortage of data scientists in the profession, automated techniques hold much promise for either improving the productivity of current data scientists or replacing them outright.

AutoML tools have emerged as a powerful solution for building and deploying machine learning models. They offer several advantages such as improved efficiency, scalability, and reduced human error. However, they also have some potential limitations, such as limited control and lack of transparency, that must be taken into account. It is important to carefully evaluate the strengths and weaknesses of AutoML tools and to choose the right approach for each specific problem. The future of AutoML looks promising, with continued advancements in usability, scalability, and



performance. As these tools become more accessible and intuitive, we can expect to see even greater adoption of machine learning across a wide range of industries and applications.

REFERENCES

1. "Automated Machine Learning: A Survey" by Sharan Sankar, et al. - IEEE Access, 2020. DOI: 10.1109/ACCESS.2020.3048192
2. "AutoML for High-Dimensional Data: A Case Study in Bioinformatics" by Nishant Gaurav, et al. - Proceedings of the 10th International Conference on Machine Learning and Computing, 2018. DOI: 10.1145/3183713.3183756
3. "AutoML for Sentiment Analysis: A Comparative Study" by Akshay Sanjay Potdar, et al. - Proceedings of the 2019 International Conference on Communication and Signal Processing, 2019. DOI: 10.1145/3323709.3323734
4. "Comparative Analysis of AutoML Frameworks for Tabular Data" by Rajat Sen, et al. - Proceedings of the 2021 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, 2021. DOI: 10.1145/3464013.3464016
5. "AutoML: A Survey of the State-of-the-Art" by Quoc V. Le, et al. - arXiv:1908.00709
6. "Efficient and Robust Automated Machine Learning" by Roberto Rossi, et al. - arXiv:2012.01077
7. "Auto-sklearn: Efficient and Robust Automated Machine Learning" by Matthias Feurer, et al. - arXiv:1907.00903
8. "Why AutoML is the future of AI in India" by DataQuest India - <https://www.dqindia.com/why-automl-is-the-future-of-ai-in-india/>
9. "An Empirical Study of AutoML Tools for Classification" by Florian Pfisterer, et al. - arXiv:2101.05578
10. "AutoML with Reinforcement Learning for Text Classification" by Rahul Kumar, et al. - Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics, 2018. DOI: 10.1145/3243107.3243228



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details