A Comparative Study of Automated Machine Learning Systems Against Manual Approaches

S.B.S.S. Bharadwaj[1], P. Thrilochan Reddy[2], Dr. Diana moses[3]

[1]*Student, Dept of Artificial Intelligence and Data Science, Methodist college of engineering and technology, Telangana, India*
[2]*Student, Dept of Artificial Intelligence and Data Science, Methodist college of engineering and technology, Telangana, India*
[3]*Professor, Dept of Computer Science and Engineering, Methodist college of engineering and technology, Telangana, India*

Abstract- *Machine learning (ML) is a computers ability to help computer systems recognize patterns from data and draw conclusions from it. It allows machines to learn from data and make better decisions using various algorithms without being explicitly programmed. In the current era of artificial intelligence and machine learning, choosing the best model and adjusting its hyperparameters for the available variety of data is a very chaotic process. Here comes automated machine learning systems (Auto-ML), which provide a dynamic way to streamline this process and help overcome this challenge. By utilizing a variety of assessment measures including accuracy, RMSE, and MAE values, our study provides a clear comparison of the Auto-ML technique to the conventional manual approach. This research gives us a strong knowledge of the computational effectiveness and scalability of the models for a given data.*

Key Words- Hyperparameter Tuning, Artificial Intelligence, Auto-ML, Accuracy, computational efficiency.

1.INTRODUCTION

In recent years, the proliferation of machine learning (ML) applications in various fields has emphasized the importance of effectively selecting of appropriate models and fine-tuning hyperparameters for optimal performance.

The traditional approach to this task often involves manual research, where data scientists repeatedly try different algorithms and hyperparameter settings, a process that consumes both time and resources. However, the emergence of automatic machine learning (Auto-ML) systems offers a promising alternative by automating these labour-intensive tasks, which can democratize ML adoption and accelerate innovation.

The goal of this research is to compare an Auto-ML system that can automatically select the best-fit ML model for a given data set with hyperparameter optimization to a traditional machine learning approach. This system aims to streamline the ML pipeline from data processing to model evaluation using the latest technology in algorithm selection, hyperparameter optimization and performance evaluation. Through comprehensive evaluation, we aim to compare the performance of our Auto-ML system with manual methods and evaluate its performance in terms of predictive accuracy, computational efficiency, and ease of use.

The importance of this research lies in its potential to revolutionize the landscape of ML model development and deployment. By automating the labour-intensive and error-prone aspects of model selection and hyperparameter tuning,

AutoML can democratize ML knowledge, allowing industry professionals of diverse experience to harness the power of advanced predictive analytics. Additionally, by systematically comparing Auto-ML with manual approaches, we try to provide insight into the trade-offs involved and identify scenarios where each approach excels.

Here, we are going to use different machine learning models like RandomForest classifier, logistic regression, Support Vector Machines and K-Nearest Neighbour classifier for manual approaches and .For Auto-ML approaches, we will use the H2O Auto-ML model and the Auto-WEKA tool. We are going to use the most popular IRIS dataset for both the approaches. We will compare and evaluate using various parameters like model accuracy, RMSE and MAE values.

This research project contributes to the ongoing Auto-ML debate by introducing a new approach to automatic model selection and hyperparameter tuning. By empirically evaluating the performance of our Auto-ML system compared to manual approaches, we aim to inform about the potential benefits and limitations of applying automated techniques to the ML process.

2.LITERATURE REVIEW

There have been many research studies carried out on automated machine learning systems, all of which have used a significant number of Auto-ML algorithms. One of the research projects Conducted by thiloshan Nagaraj and guhanathan poravi focuses on automated machine learning (Auto-ML) systems, emphasizing the need to automate manual model creation processes. The article analyses existing Auto-ML approaches, with a specific focus on key components such as preprocessing, feature engineering, and model selection engines. The study highlights the importance of integrating functional end products, affordable data centers, Python-centric research, and neural networks for future Auto-ML advancements.

Another project ,conducted by Adithya Balaji and alexander allen compares the performance of different automatic machine learning frameworks on classification and regression datasets. While machine learning excels in classification, while TPOT performs well in Regression tasks the study emphasizes the importance of unbiased data analysis and suggests further research into specific techniques and other Auto-ML frameworks. The performance of various frameworks are presented in a research study carried out by Matthias Feurer, Katharina Eggensperger, Stefan Falkner, Marius Lindauer and Frank Hutter in the context of meta-learning.

In a research carried out by Jonas M. Kübler and VincentStimper,the mean deviation of the witness function is used as the test statistic with the aim of optimizing the power of the test by minimizing the squared loss. Experimental results show variable performance when testing different Auto-ML variants, with Auto-ML(class) performing similarly to MSE, while Auto-ML(bin) shows reduced performance. The study reveals that longer training times do not significantly affect the results in the original material. Auto-ML models like TPOT and H2O achieve around 81-82% accuracy, while manual models reach 100%. From the research conducted by Gururaj N Kulkarni, Sateesh Ambesange, Vijayalaxmi A and Anindita Sahoo, it can be concluded that Auto-ML saves time but may require additional tuning for improved accuracy

The Development of Auto-ML and potential benefits and Auto-ML architecture are discussed in a research paper by Chahal Vadalmiya, Nisha Gandhi, Gauri Khotele, Ansh Mishra and Prof. Rohini Jadhav. Future developments include the implementation of a graphical user interface to cater to non-technical users, further streamlining the feature engineering process and enhancing model accuracy as suggested by Akshat Chaturvediin his research project.

Finally the research conducted by Nalawade Shruti, , Konde Priyadarshani, , Aashlesha Modhe, Bhise Sneha, Pokharkar Gayatri, and Bhalerao D.N emphasizes the significance of automated systems in maintaining data, ensuring rapid operations, and enhancing security through automated machine learning systems.

3.METHODOLOGY

In the proposed methodology, this study explores mixed method approach. We have explored various methods of implementing Auto-ML and have opted for two techniques for this research are H2O Auto-ML model and Auto Weka tool for Auto-ML approaches, ML algorithms like Random-Forest classifier, K-Nearest Neighbour classifier, Support Vector Machines and logistic regression are taken up for the conventional manual approach. Hyper parameter tuning is carried out by THE GridsearchCV method

3.1.H2O Auto-ML

Developed by H2O.ai, H2O Auto-ML revolutionizes the machine learning model development process by automating critical steps in the pipeline. With a focus on simplicity and efficiency, it selects optimal algorithms adapted to dataset characteristics and prediction tasks, from traditional methods such as random forests to state-of-the-art deep learning architectures. The strength of the platform lies in its automated hyperparameter optimization, which uses advanced techniques such as Bayesian optimization and web search to efficiently tune model settings for maximum performance.

In addition, H2O Auto-ML incorporates aggregate learning, combining knowledge from different models to improve prediction accuracy and reliability.

Scalability is a key feature of H2O Auto-ML, allowing it to smoothly handle large datasets. By leveraging parallel processing capabilities, the framework accelerates model training and evaluation processes, making it ideal for high-data real-world applications.

H2O Auto-ML provides a comprehensive solution for automating machine learning tasks, significantly reducing manual intervention and accelerating the journey from data to insight.

3.2.Auto-WEKA

Auto-WEKA is an automated machine learning tool in the WEKA series designed to streamline the model building process by automating model selection and hyperparameter optimization. It evaluates various algorithms in the WEKA library, to identify the best fit for a given dataset and prediction task.

Using techniques such as Bayesian optimization and genetic algorithms. Auto-WEKA effectively tunes model hyperparameters to improve performance. Users can customize the tool by setting evaluation metrics and calculation time limits.

Auto-WEKA integrates cross-validation methods to accurately evaluate model performance and provides metrics such as accuracy, RMSE and MAE score for evaluation.

3.3.Comparision

As we know Auto-ML is completely responsible for model selection and hyperparameter tuning. We evaluate and compare this model with traditional manual approaches for which we have taken several algorithms.

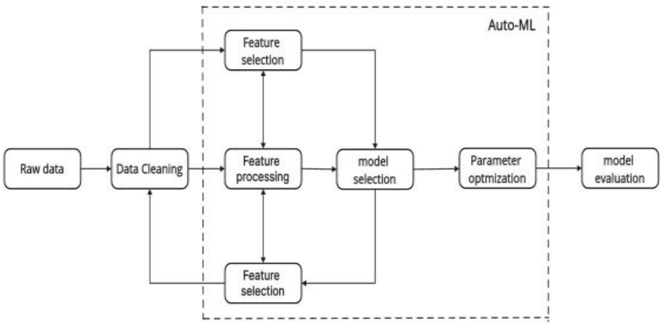

Fig -1: Architecture of Auto-ML

For manual approaches we have taken Random-Forest classifier, SVM, Logistic regression, Gradient boosting and KNN classifiers with the GridSearchCV hyperparameter tuning technique with a 5 fold cross-validation measure for both approaches, we iterate through each algorithm

H2O and Auto-WEKA are used for the implementation of the Auto-ML approach where the complete model selection to evaluation is carried out dynamically

Both approaches are compared using various measures like model accuracy, computational efficiency and scalability.

4.RESULTS

Using the conventional manual approach of machine learning models, we obtain the highest accuracy of the model by optimizing the parameters to be 90.9% this accuracy was further improved by hyperparameter tuning resulted in 100.0% accuracy achieved by several linear models for this dataset.

Through efficient exploratory data analysis and parameter optimization ,the manual approach was able to achieve this level of accuracy.

```
Model: RandomForest
Best Parameters: {'max_depth': None, 'n_estimators': 10}
Cross-validation Accuracy: 1.0

Model: SVM
Best Parameters: {'C': 0.1, 'kernel': 'linear'}
Cross-validation Accuracy: 1.0

Model: LogisticRegression
Best Parameters: {'C': 0.1, 'penalty': 'l2'}
Cross-validation Accuracy: 1.0

Model: GradientBoosting
Best Parameters: {'learning_rate': 0.01, 'n_estimators': 50}
Cross-validation Accuracy: 0.9916666666666668

Model: KNN
Best Parameters: {'n_neighbors': 3, 'weights': 'uniform'}
Cross-validation Accuracy: 1.0

Test Accuracy of the Best Model: 1.0
```

Fig -2: H2O Auto-ML Results

For the Auto-ML approach, we used the H2O Auto-ML library in python. After hyperparameter tuning, we can observe that the stacked ensemble model was the best fit with an accuracy of 79.3% for this dataset.

```
Model Details
=============
H2OStackedEnsembleEstimator : Stacked Ensemble
Model Key: StackedEnsemble_BestOfFamily_1_AutoML_1_20240415_165908

Model Summary for Stacked Ensemble:
key                                    value
-------------------------------------  ---------------
Stacking strategy                      cross_validation
Number of base models (used / total)   3/5
# GBM base models (used / total)       1/1
# DeepLearning base models (used / total)  1/1
# DRF base models (used / total)       1/2
# GLM base models (used / total)       0/1
Metalearner algorithm                  GLM
Metalearner fold assignment scheme     Random
Metalearner nfolds                     5
Metalearner fold_column
Custom metalearner hyperparameters     None

stackedensemble prediction progress: |██████████████████████████████████
█████| (done) 100%
Test Accuracy: 0.7931034482758621
```

Fig -3: H2O Auto-ML Results

Another tool used for the Auto-ML approach is Auto-WEKA, which resulted in an accuracy of 100.0% for the k-Nearest Neighbour classifier making it the best fit for this dataset. It has took 935 configurations for different classifiers with different parameters to obtain the best output for this data.

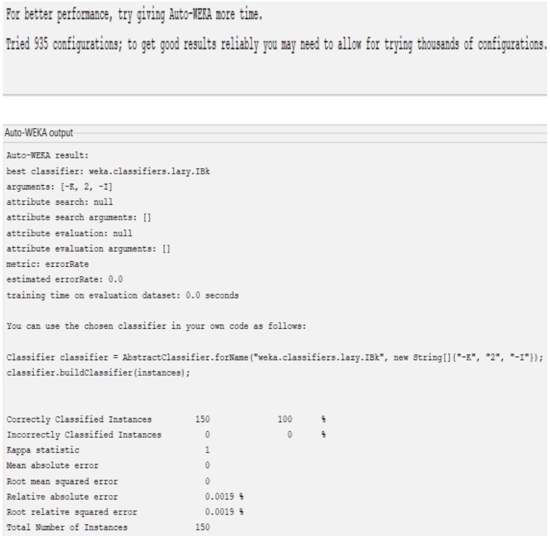

Fig -4: Auto-Weka Results

5.CONCLUSION

Together, our research highlights the nuanced dynamics of Auto-ML and manual approaches in the context of machine learning model development. Through rigorous testing and analysis, we found that while Auto-ML systems demonstrate impressive performance and scalability, manual approaches are often superior for scenarios that require fine-tuning and domain-specific expertise.

Our results emphasize the importance of considering task complexity, resource constraints, and operational expertise when choosing between Auto-ML and manual methods. Auto-ML platforms systematically iterate through various algorithms, hyperparameters, and preprocessing techniques to find the best performing model, making machine learning accessible to users with little domain knowledge, On the other hand manual approach requires not only domain and technical knowledge but also an enormous amount of time and computational efficiency. Finally, each having its own pros and cons that are completely task and data specific.

Through our research we conclude that improvising models can also be achieved by combining both approaches in a hybrid framework promising to maximize the benefits of automation and harness human intelligence where it is most effective.

You can access the project files using this link :

https://drive.google.com/file/d/1F_NY9MZqGJx7kPuXxC7frNJTUJvvwzYp/view?usp=sharing

REFERENCES

[1] A Review on Automated Machine Learning (Auto-ML) Systems, Thiloshon Nagarajah, Guhanathan Poravi,I2CT 2019.

[2] Benchmarking Automatic Machine Learning Frameworks, Adithya Balaji, Alexander Allen,2018.

[3] Auto-Sklearn 2.0: Hands-free Auto-ML via Meta-Learning, Matthias Feurer, Katharina Eggensperger, Stefan Falkner, Marius Lindauer, Frank Hutter, JMLR,2022.

[4] AutoML Two-Sample Test, Jonas M. Kübler, Vincent Stimper,2023.

[5] Efficient and Robust Automated Machine Learning, Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Tobias Springenberg, Manuel Blum, Frank Hutter, University of Freiburg, Germany.

[6] Comparision of diabetic prediction AutoML model with customized model, Gururaj N Kulkarni, Sateesh Ambesange, Vijayalaxmi A, Anindita Sahoo, ICAIS-2021.

[7] Development of auto-ml tool, Chahal Vadalmiya, Nisha Gandhi, Gauri Khotele, Ansh Mishra, Prof. Rohini Jadhav, IRJET,2023.

[8] Automatic Feature Engineering tool for Machine learning, Akshat Chaturvedi, IRJET, 2021.

[9] Automatic question paper formation system, Nalawade Shruti, , Konde Priyadarshani, , Aashlesha Modhe, Bhise Sneha, Pokharkar Gayatri, Bhalerao D.N, IRJET, 2024.