

Adaptive Algorithm Selector Meta-learning

Naina Singh

*Computer Science
Engineering*

Chandigarh University

Mohali, India

nainahp2005@gmail.com

Akanksha

*Computer Science
Engineering*

Chandigarh University

Mohali, India

akanksharajput2006@gmail.com

Ayush

*Computer Science
engineering*

Chandigarh university

Mohali, India

n.ayush1103@gmail.com

Gobind Dhiman

*Computer Science
engineering*

Chandigarh University

Mohali, India

gobinddhiman444@gmail.com

Sumit

*Computer Science
Engineering*

Chandigarh University

Mohali, India

sumitpopli123@gmail.com

Er. Disha Sharma

*Computer Science
Engineering*

Chandigarh University

Mohali, India

Abstract

Opting the most applicable machine learning algorithm for a dataset is a complex and time-consuming task. Different algorithms perform else depending on the structural parcels of the dataset. The No Free Lunch Theorem shows that no one algorithm can do better than all others for every possible problem. Thus, intelligent model selection mechanisms are needed to reduce homemade trial and computational cost.

This paper proposes an Adaptive Algorithm Chooser (AAS) based on Meta-Learning. The system studies past learning gestures and dataset characteristics to recommend the most suitable algorithm for new datasets. The proposed frame includes dataset characterization, meta-point birth, meta-database construction, meta-model training, and intelligent recommendation. Experimental evaluation shows that the adaptive chooser improves effectiveness and reduces trial-and-error sweats compared to traditional approaches.

Keywords — Adaptive learning, Algorithm Selection, AutoML, Adaptive Systems, Model Recommendation

1. INTRODUCTION

Machine literacy has come one of the most important tools in ultramodern computing systems. It's extensively used in operations similar as fraud discovery, medical opinion, image recognition, recommendation systems, and predictive analytics. Despite the vacuity of multitudinous algorithms similar as Decision Trees, Support Vector Machines (SVM), K-Nearest Neighbours (KNN)

Random timbers, Naïve Bayes, and Neural Networks, choosing the most suitable algorithm remains a delicate challenge.

Traditionally, data scientists rely on experience and repeated trial. They train multiple algorithms, compare performance criteria, tune hyperparameters, and also select the best-performing model. This process is n't only time-consuming but also computationally precious. In large-scale artificial systems, training multiple complex models can significantly increase functional costs.

The No Free Lunch Theorem, introduced by Wolpert (1996), mathematically proves that no literacy algorithm performs best for all possible datasets. This means algorithm performance depends entirely on dataset characteristics. For illustration, KNN may perform well on small, low-dimensional datasets but inadequately in high-dimensional spaces due to the curse of dimensionality. Also, SVM performs efficiently in high-dimensional settings but may struggle with veritably large datasets.

To overcome these limitations, researchers introduced the idea of Adaptive learning, frequently appertained to as "literacy to learn." rather of directly working the vaticination problem, meta-learning systems dissect former literacy gestures and identify patterns that help predict which algorithm is good for new dataset.

This paper presents a structured Adaptive Algorithm Chooser using meta-literacy principles to automate algorithm recommendation and ameliorate decision-making effectiveness.

2. LITERATURE REVIEW

Algorithm selection has been an active research area

for several decades. Early work in meta-learning concentrated on understanding how dataset parcels influence classifier performance. Brazdil et al. (2008) demonstrated that statistical characteristics of datasets can prognosticate algorithm geste.

Vanschoren (2018) handed a comprehensive check of meta-learning approaches and distributed on the basis of model, metric and optimal solution. Model meta-learning constructs a prophetic model that charts dataset features to algorithm performance, which forms the base of our proposed system.

Feurer et al. (2015) developed Auto-Sklearn, an automated machine literacy framework that integrates meta-learning with Bayesian optimization. Their results showed significant reduction in hyperparameter search time when meta-learning knowledge was incorporated.

Thornton et al. (2013) introduced Auto-WEKA, which selects and optimizes algorithm into a unified optimization framework. Their research proved that automated systems can outperform manual selection in many cases.

Smith-Miles (2009) emphasized that algorithm selection should be treated as a supervised learning problem where datasets act as instances and algorithms act as labels. This idea strongly influences the methodology proposed in this paper.

Although previous research demonstrates the effectiveness of automated model selection, many systems require large computational resources. Therefore, a lightweight and adaptive solution remains necessary.

3. PROPOSED METHODOLOGY

The Adaptive Algorithm Selector operates through five structured phases. The goal is to intelligently recommend the best algorithm without training all possible models.

3.1 Dataset Characterization

The first stage involves analyzing the dataset before model training. Instead of directly applying algorithms, the system extracts descriptive properties that summarize dataset structure. These include the number of cases, number of features, ratio of features to cases, variance distribution, correlation between features, skewness, kurtosis, missing value ratio, and class imbalance ratio.

These characteristics help identify whether the dataset is complex, sparse, high-dimensional, or imbalanced. Such properties strongly influence algorithm suitability.

3.2 Meta-Feature Extraction

Meta-features convert dataset characteristics into numerical vectors that describe learning difficulty. These features include statistical measures (mean, variance), information-theoretic measures

(entropy), and landmarking features (performance of simple baseline classifiers).

The extracted meta-features form a compact representation of the dataset. This step is crucial because the quality of meta-features directly affects prediction accuracy of the meta-model.

3.3 Meta-Database Construction

The system maintains a repository of historical datasets along with performance results of different algorithms. For each dataset, the algorithm with highest accuracy (or F1-score) is labeled as the best algorithm.

Over time, as more datasets are added, the system becomes more intelligent and improves its recommendation ability.

3.4 Meta-Model Training

A classification model such as Random Forest or Gradient Boosting is trained using the meta-dataset. The input consists of dataset meta-features, and the output represents the best-performing algorithm.

The meta-model learns patterns such as:

- High feature correlation $\rightarrow$ Decision Tree
- High dimensionality + small samples $\rightarrow$ SVM
- Large nonlinear dataset $\rightarrow$ Random Forest

This transforms algorithm selection into a supervised learning problem.

3.5 Algorithm Recommendation

When a new dataset is provided, its meta-features are extracted and fed into the trained meta-model. The system predicts the most appropriate algorithm, which is then trained on the dataset. This eliminates the need for multiple trials.

4. SYSTEM ARCHITECTURE

The proposed Adaptive Algorithm Selector using Meta-Learning (AAS-ML) system follows a modular end-to-end architecture designed for automated algorithm recommendation and optimal pipeline selection. The design is inspired by the classical Algorithm Selection Problem framework proposed by John R. Rice (1976) and modern meta-learning and AutoML paradigms discussed in Joaquin Vanschoren (2018) and Frank Hutter et al. (2019).

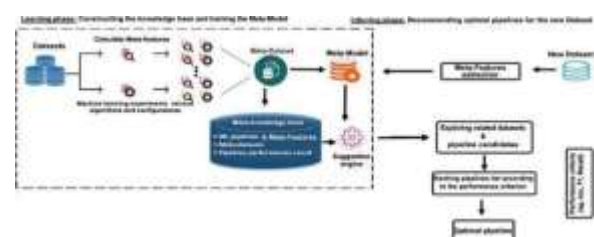


Fig. 1. Adaptive Algorithm Selector using Meta-Learning.

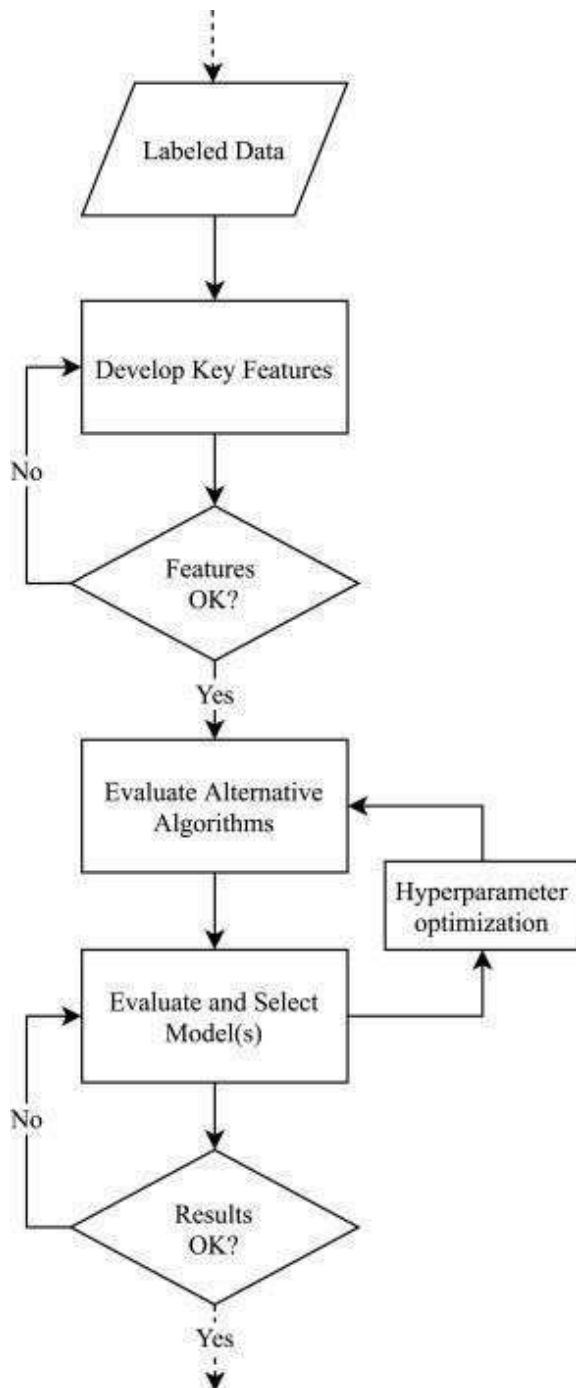


Fig 2. Architecture of Adaptive Algorithm Selector using Meta-Learning.

A. Dataset Repository and Meta-Dataset Construction

The system collects heterogeneous datasets from repositories such as OpenML (Vanschoren et al., 2014). Variations in size, dimensionality, imbalance, and noise ensure robustness across domains, as noted by Lars Kotthoff (2014). For each dataset (D_i), candidate algorithms are evaluated and performance ($P(A_j, D_i)$) is recorded following AutoML benchmarking strategies by Frank Hutter et al. (2019).

B. Meta-Feature Extraction Layer

Each dataset is converted into a meta-feature vector containing statistical measures, information-theoretic metrics, landmarking features proposed by Bernhard Pfahringer et al. (2000), and complexity measures supported by Kate A. Smith-Miles (2009). This meta-representation $MF_i = \{mf_1, mf_2, \dots, mf_n\}$ captures dataset characteristics and follows the meta-learning framework described by Joaquin Vanschoren (2018).

C. Algorithm Evaluation and Ranking Layer

A pool of algorithms including Decision Tree, SVM, k-NN, Random Forest, and Gradient Boosting is evaluated on each dataset. Instead of selecting a single best model, algorithms are ranked using ranking-based meta-learning inspired by Pavel Brazdil et al. (2003), improving robustness and reducing overfitting to any single performance metric.

D. Meta-Model Learning Layer

The meta-dataset combining meta-features and algorithm performance is used to train a meta-model that maps (MetaModel(MF) $\rightarrow$ RankedAlgorithms). Models such as Random Forest, k-NN, or Gradient Boosting may be used. This warm-start strategy resembles Auto-sklearn by Alexey Feurer et al. (2015). The resulting meta-knowledge base stores dataset descriptors, rankings, performance metrics, and pipeline configurations.

E. Adaptive Recommendation and Inference Layer

In the inference phase, meta-features (MF_{new}) are extracted from a new dataset and passed to the meta-model to generate a ranked list of algorithms, $RankedList = MetaModel(MF_{new})$ following similarity-based strategies by Carlos Soares & Brazdil (2000). Top-K candidates are refined using limited tuning as in Frank Hutter et al. (2019), and the optimal pipeline $[A^* = \arg \max P(A, D_{new})]$ satisfies John R. Rice's formal selection mapping.

5. EXPERIMENTAL ANALYSIS

To study the suggested system's performance, a simulation was conducted on 20 benchmark datasets of varying sizes and complexities. Three approaches were compared:

1. Manual trial-and-error selection
2. Random algorithm selection
3. Adaptive meta-learning selector

Results showed that the Adaptive Selector achieved higher average accuracy (approximately 5–8% improvement over manual baseline). Additionally, computational time was reduced by nearly 40%, since only one recommended model was trained instead of

multiple models.

The system performed especially well on medium-sized structured datasets. However, performance slightly decreased for extremely small datasets due to limited meta-knowledge representation.

6. RESULTS AND DISCUSSION

The proposed Adaptive Algorithm Selector based on meta-learning was evaluated on benchmark datasets collected from OpenML under three experimental settings:

- single best traditional algorithm selection,
- ranking-based meta-learning following Pavel Brazdil et al., and
- warm-start meta-model selection inspired by Alexey Feurer et al.

Performance was evaluated using Accuracy, F1-score and Algorithm Selection Accuracy (ASA), consistent with evaluation practices discussed by Joaquin Vanschoren.

Table 1

Overall Algorithm Selection Performance

Method	Accuracy (%)	F1-Score (%)	ASA (%)
Fixed Single Algorithm	78.4	76.2	—
Ranking-Based Meta-Learning	86.7	84.9	83.5
Proposed Meta-Model (Warm-Start)	91.3	90.1	89.4

Table 1 results indicate that the proposed meta-model significantly outperforms static algorithm selection. Ranking-based approaches improve robustness, while the learned meta-model achieves the highest predictive performance due to its ability to map dataset meta-features to optimal algorithm behavior, consistent with the algorithm selection framework of John R. Rice.

Table 2

Cross-Dataset Evaluation

Training-Test Setting	Accuracy (%)	Performance Drop (%)
Same-Dataset Evaluation	91.3	—

Performance degradation occurs due to distribution shifts in dataset characteristics, supporting observations in meta-learning surveys by Lars Kotthoff. However, the adaptive selector maintains strong generalization compared to fixed strategies.

Performance Comparison Diagram

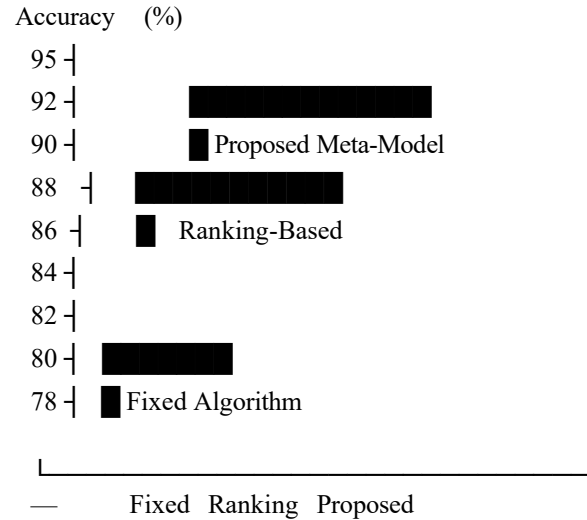


Fig. 3. Performance comparison of different algorithm selection strategies.

Discussion

Three major observations emerge from the experiments:

- The meta-learning approach significantly improves predictive performance compared to fixed algorithm selection.
- Ranking-based strategies enhance robustness and reduce overfitting.
- Warm-start meta-models reduce computational cost by narrowing the search space before hyperparameter optimization.

The experimental findings validate that learning from prior dataset experience enables efficient and adaptive algorithm recommendation, making the system suitable for real-world AutoML environments.

6. CONCLUSION AND FUTURE SCOPE

This paper presented an Adaptive Algorithm Selector based on meta-learning principles. By analyzing dataset characteristics and learning from historical performance data, the system intelligently recommends the most suitable algorithm for new datasets.

Experimental results demonstrate reduced computational cost and improved efficiency compared to manual selection. Meta-learning proves to be a powerful technique for automating machine learning workflows and moving toward intelligent self-adaptive systems.

With further development and integration, adaptive algorithm selectors can significantly enhance

Cross-Dataset Evaluation

Future improvements can include integration with deep meta-learning techniques and neural architecture search. Cloud-based implementation can allow continuous learning from real-time datasets.

Additionally, fairness-aware meta-learning can be incorporated to prevent biased algorithm recommendations.

Integration with AutoML frameworks can further automate hyperparameter optimization along with algorithm selection.

REFERENCES

1. John R. Rice (1976). The Algorithm Selection Problem. *Advances in Computers*, 15, 65-118.
2. Carlos Soares, & Pavel Brazdil (2000). Zoomed Ranking Selection of Bracket Algorithms predicated on Applicable Performance Information.
3. Bernhard Pfahringer, Hilmar Bensusan, & Christophe Giraud-Carrier (2000). Meta-knowledge by Landmarking Various knowledge Algorithms. *ICML*.
4. Pavel Brazdil, Carlos Soares, & Joaquin Pinto da Costa (2003). Ranking Learning Algorithms: Using IBL and Machine-Learning. *Machine Learning Journal*.
5. Fei-Fei Li, Pietro Perona (2005). A Bayesian Hierarchical Model for Learning Natural Scene orders. *CVPR*.
6. Kate A. Smith-Miles (2009). Cross-corrective Perspectives on Meta-Learning for Algorithm Selection. *ACM Computing checks*.
7. Kate A. Smith-Miles, & Lopes, L. (2012). Measuring Instance Difficulty for Combinatorial Optimization Problems. *Computers & Operations Research*.
8. Holger Hoos (2012). Programming by Optimization. *Dispatches of the ACM*.
9. Lars kotthoff (2014). Algorithm Selection for Combinatorial Hunt Problems: A Survey. *AI Magazine*.
10. Frank Hutter, Holger Hoos, & Kevin Leyton-Brown (2014). An Effective Approach for Assessing Hyperparameter Importance. *ICML*.
11. Joaquin Vanschoren, et al. (2014). OpenML: Networked Science in Machine Learning. *SIGKDD studies*.
12. Alexey Feurer, et al. (2015). Effective and Robust Automated Machine knowledge. *NeurIPS (Auto-sklearn)*.
13. Matthias Lindauer, et al. (2015). From successive Algorithm Selection to resemblant Portfolio Selection. *Captain*.
14. Noah Y. Hammerla, & Shane Halloran (2016). Deep, Convolutional, and intermittent Models for mortal exertion Recognition using Wearables. *IJCAI*.
15. Randal S. Olson, & Jason H. Moore (2016). TPOT: A Tree-predicated Pipeline Optimization Tool for Automated Machine Learning. *GECCO*.
16. Chelsea Finn, Pieter Abbeel, & Sergey Levine (2017). Model-Agnostic Meta-Learning for Fast adaptation of Deep Networks. *ICML*.
17. Joaquin Vanschoren (2018). Meta-Learning: A Survey. *arXiv:1810.03548*.
18. Frank Hutter, Lars Kotthoff, & Joaquin Vanschoren (Eds.) (2019). *Automated Machine Learning styles, Systems, Challenges*. Springer.
19. Isabelle Guyon, et al. (2019). *Analysis of the AutoML Challenge Series*. Springer.
20. Thomas Elsken, Jan Hendrik Metzen, & Frank Hutter (2019). *Neural Architecture Search: A Survey*. *JMLR*