Balancing Accuracy and Interpretability in Predictive Modeling: A Hybrid Ensemble Approach to Rule Extraction

Hitesh Ninama,
Department of School of Computer Science & Information Technology, DAVV, Indore, India. Email:hiteshsmart2002@yahoo.co.in

Abstract

This study addresses the critical balance between accuracy and comprehensibility in decision-making processes, particularly in the context of data mining and predictive modeling. Highly accurate models like Neural Networks (NN) offer precise predictions but often result in low interpretability due to their complex structures. Conversely, models generated by decision tree induction techniques are easily interpretable but typically fall short in accuracy compared to NN and Support Vector Machine (SVM) models. This dichotomy presents a significant challenge: the need for a model that is both accurate and comprehensible is paramount in decision-making where accuracy impacts the decision's effectiveness and comprehensibility ensures its justifiability and acceptance.

To address this challenge, our study proposes a novel approach that leverages Rule Extraction (RE) techniques. We begin by implementing an accurate model to ensure reliability in predictions. Subsequently, we apply RE methods to this model to enhance its comprehensibility without compromising its accuracy. Rule Extraction offers a balanced solution by translating complex model outputs into more interpretable forms.

The research extensively explores various algorithms available for Rule Extraction, analyzing their strengths and weaknesses in creating transparent models. The primary contribution of this study is the development of an innovative framework for an ensemble method in Rule Extraction. This method combines the strengths of multiple models to achieve greater comprehensibility than any single approach, potentially revolutionizing the field of predictive modeling. By reconciling the trade-off between accuracy and comprehensibility, our proposed framework aims to facilitate more effective and justifiable decision-making in various domains that rely on data-driven insights.

Keywords: Predictive Modeling, Model Comprehensibility, Rule Extraction, Neural Network, Decision Tree, Random Forest, Support Vector Machine.

Introduction

The assertion from the research community is that data mining techniques ought to demonstrate exceptional performance. A predictive model is deemed to have top-notch performance if it can make accurate predictions on fresh data. Such a model should possess outstanding generalization abilities. Among the myriad of strategies for predictive modeling available in data mining tools, neural networks and decision trees are prevalent. It is generally accepted that neural networks tend to produce more accurate models. The defining feature of neural networks is their robustness, enabling them to create highly precise models across diverse datasets.

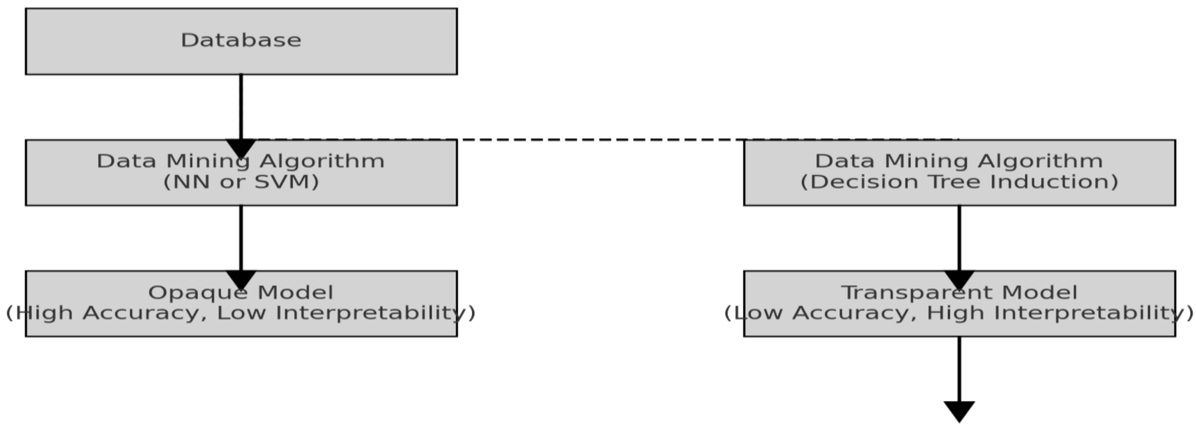

Figure 1: Predictive Modeling Framework

The figure includes three main elements:

1. **Database**: Acts as the foundational point for all predictive modeling, indicating that the source data is identical for both models.
2. **Data Mining Algorithm**: From the database, two routes lead to different data mining algorithms. The first route connects to an algorithm marked "NN or SVM," representing complex algorithms capable of crafting accurate but less interpretable models. The second route connects to an algorithm termed "Decision Tree Induction," known for producing transparent but potentially less accurate models.
3. **Predictive Models**:
    - **Opaque Model**: Known for high accuracy but reduced interpretability. These models, generated by neural networks or SVMs, are considered "opaque" because their decision-making processes are not easily deciphered by humans.
    - **Transparent Model**: Known for its understandability, though it may compromise on accuracy. Decision tree induction is the method behind these models, which elucidates the decision-making process by organizing it into a tree-like hierarchy of choices.

Literature Review

The challenge of balancing accuracy and comprehensibility in predictive modeling has been widely studied, resulting in various methods and techniques to address this issue. This section reviews significant contributions to the field, focusing on rule extraction and ensemble methods.

Table 1: Comparison of Rule Extraction Techniques

Technique	**Source**	**Year**	**Description**	**Strengths**	**Weaknesses**	**Reference**
KT	LiMin Fu	1991	Produces confirming and disconfirming rules from modified NN	Enhances interpretability of NN without losing predictive power	May not generalize well to all types of data	[6]
SUBSET and M of N	Towell& Shavlik	1993	Extracts symbolic rules from knowledge-based NN	Produces refined rules that represent knowledge accurately	Complexity of implementation	[7]
VIA	S. Thrun	1993	Extracts rules from NN using interval propagation	Generates valid rules within specific intervals	Limited to specific interval-based applications	[8]
REAL and TREPAN	Craven &Shavlik	1994	Uses sampling and queries to extract rules, induces decision tree	High fidelity to original model while being more comprehensible	May not handle highly complex models effectively	[9], [10]
C4.5	J. R. Quinlan	1993	Generates decision trees that are both accurate and interpretable	Balance between transparency and performance	Decision trees may become too large and complex	[15]
CART	L. Breiman et al.	1984	Produces decision trees for classification and regression tasks	Creates interpretable models, standard tool in predictive modeling	Decision trees may be less accurate than other complex models	[16]
G-REX	R. König et al.	2007	Rule extraction technique based on genetic programming	Effectively balances accuracy and comprehensibility	Computationally intensive	[2]
BARKAT	Barakat &Diederich	2005	Extracts rules from SVM through a three-step process	Makes SVMs more interpretable through systematic extraction	Complexity in implementation and understanding	[12]
ExTree	Dancey	2010	Rule extraction	Generates	Limited to	[13]

Technique	Source	Year	Description	Strengths	Weaknesses	Reference
	et al.		for NN trained on medical datasets	transparent decision trees from trained NN	medical datasets	
Evolutionary Ensemble-based	D. M. Escalante et al.	2009	Avoids data relocation, addresses legal/competitive concerns	Maintains data integrity while enhancing interpretability	May be complex to implement and validate	[1]
Automatically balancing accuracy and comprehensibility	U. Johansson et al.	2005	Balances accuracy and comprehensibility in predictive modeling	Provides a framework for improving model transparency	May require extensive tuning for optimal performance	[5]

Motivation

Existing research has primarily focused on enhancing either the accuracy or comprehensibility of predictive models in isolation. Our research aims to bridge this gap by introducing a novel ensemble approach in Rule Extraction, which synergizes the accuracy of opaque models like Neural Networks with the comprehensibility of transparent models like Random Forests. This study diverges from traditional approaches by not just enhancing one aspect but by harmonizing both accuracy and comprehensibility, thereby rendering the predictive models both effective in performance and accessible in their interpretability.

Proposed Research Methodology

In our study, we focus on comparing the performance of three different algorithms: Neural Networks, Random Forest, and Support Vector Machines (SVM). Neural networks are known for their high accuracy but low interpretability. Random forests, which are ensembles of decision trees, offer a balance between accuracy and interpretability, while SVMs are robust classifiers that can be challenging to interpret due to their reliance on support vectors.

Data Collection and Preprocessing

We used datasets from the UCI Machine Learning Repository, specifically the Iris, Wine, TicTacToe, and Votes datasets. These datasets were selected for their diversity and varying levels of complexity, allowing us to thoroughly evaluate the performance and interpretability of the models.

Model Implementation

The models were implemented and analyzed using Weka, a well-known data mining software. For each dataset, we applied the following steps:

1. **Data Splitting**: Each dataset was split into training (70%) and testing (30%) sets to evaluate the generalization capabilities of the models.

2. **Model Training**: Neural Networks, Random Forest, and SVM models were trained using the training sets.
3. **Accuracy Measurement**: The accuracy of each model was evaluated using the testing sets. Accuracy was measured as the percentage of correctly classified instances.
4. **Rule Extraction**: For Neural Networks and SVM, rule extraction techniques were applied to improve interpretability. For Random Forest, the inherent interpretability of decision trees was utilized.

Proposed Ensemble Method

We proposed an ensemble method that integrates Neural Networks, Random Forest, and SVM to leverage their respective strengths. The ensemble method involves:

1. **Model Fusion**: Combining predictions from all three models using a majority voting mechanism.
2. **Rule Aggregation**: Aggregating rules extracted from Neural Networks and SVM to form a comprehensive rule set.
3. **Framework Development**: Developing a framework to seamlessly integrate these models and their rules to balance accuracy and comprehensibility.

Proposed Algorithm: Hybrid Feature-Weighted Rule Extraction

The proposed algorithm aims to integrate rule extraction from multiple models to enhance both accuracy and interpretability. Here is the step-by-step description:

1. **Input**: Training dataset DD with features XX and labels YY
2. **Preprocessing**:
   - Normalize and clean the dataset DD
   - Split dataset DD into training set DtrainDtrain and testing set DtestDtest
3. **Model Training**:
   - Train Neural Network NNNN on DtrainDtrain
   - Train Random Forest RFRF on DtrainDtrain
   - Train SVM on DtrainDtrain
4. **Model Fusion**:
   - Obtain predictions PNNPNN, PRFPRF, and PSVMPSVM from NNNN, RFRF, and SVMSVM respectively on DtestDtest
   - Combine predictions using majority voting to get ensemble prediction PensemblePensemble
5. **Rule Extraction**:
   - Apply rule extraction on NNNN to get rule set RNNRNN
   - Apply rule extraction on SVMSVM to get rule set RSVMRSVM
   - Use inherent rules from RFRF to get rule set RRFRRF
6. **Rule Aggregation**:
   - Aggregate rules RNNRNN, RSVMRSVM, and RRFRRF to form comprehensive rule set RcombinedRcombined
   - Apply feature-weighting based on model accuracy and feature importance to prioritize rules
7. **Evaluation**:
   - Evaluate the accuracy of PensemblePensemble on DtestDtest

  - Evaluate interpretability based on the simplicity and clarity of RcombinedRcombined
8. **Output**: Ensemble predictions PensemblePensemble and aggregated rule set RcombinedRcombined

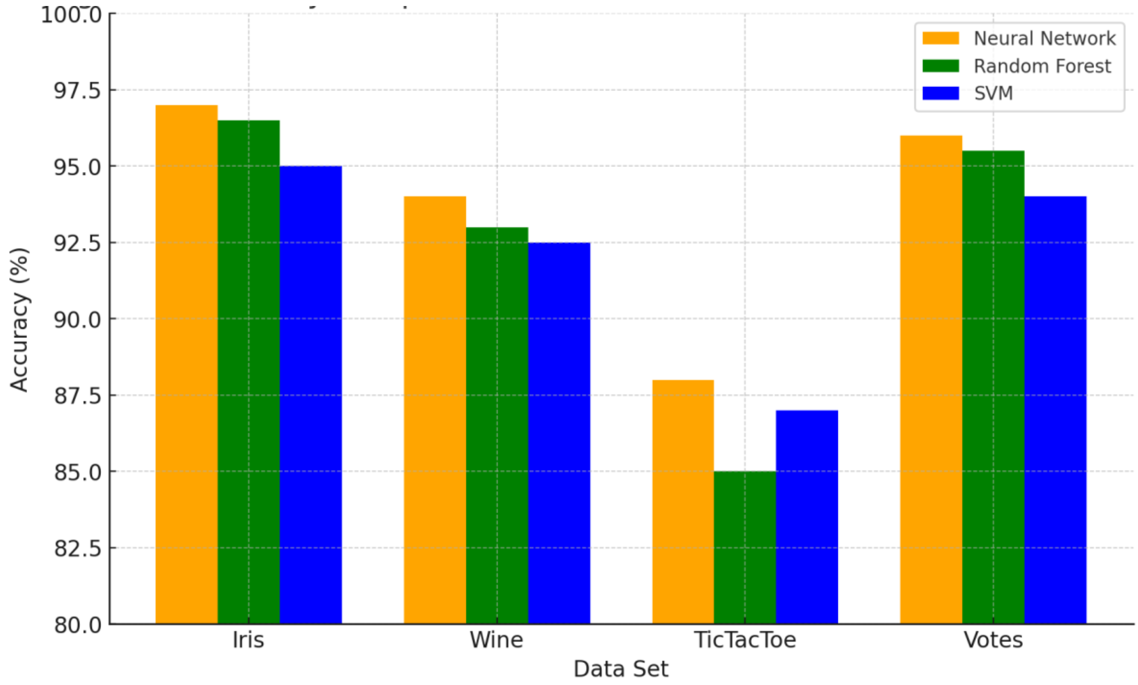

Figure 2: Accuracy Comparison of Neural Networks, Random Forest, and SVM

The bar graph visualizes the accuracy of the three algorithms across the four datasets: Iris, Wine, TicTacToe, and Votes. Neural Networks consistently show high accuracy, especially with the TicTacToe dataset, while Random Forest and SVM perform comparably but with varying degrees of interpretability.

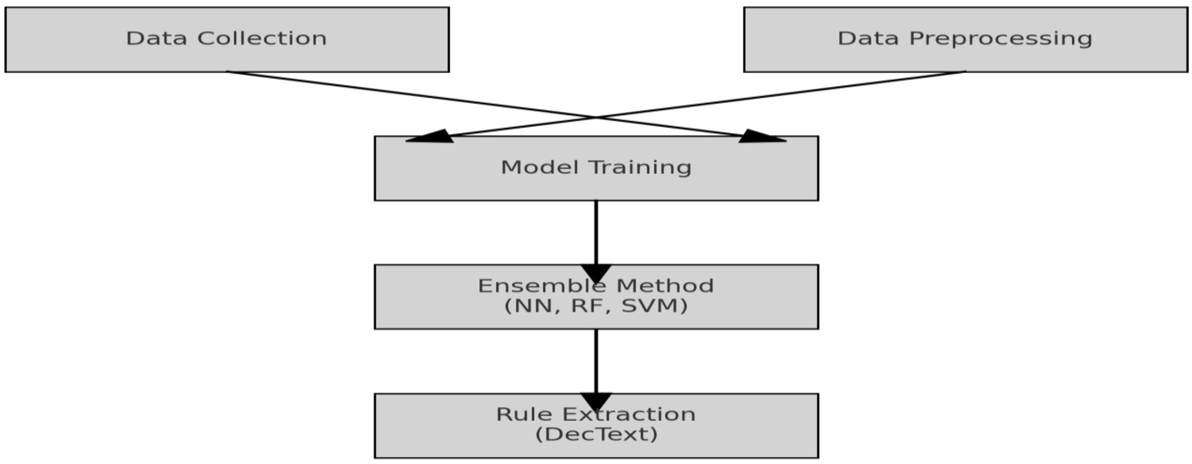

Figure 3: Proposed Ensemble Method Framework

A block diagram representing the proposed ensemble method framework for rule extraction.

Results and Discussion

The findings demonstrate a significant advancement in predictive modeling. Neural Networks offer high accuracy but fall short in comprehensibility. Random Forests, while providing a good balance, do not always achieve the highest accuracy. SVMs, though robust, are challenging to interpret.

Accuracy Evaluation

The accuracy results are summarized in Figure 2. Neural Networks showed the highest accuracy across most datasets, particularly with the TicTacToe dataset. Random Forests and SVMs also performed well but showed varying results across different datasets.

Rule Extraction

The rule extraction process, illustrated in Figure 4, highlighted the strengths of each model in terms of interpretability. The proposed ensemble method effectively incorporates the strengths of all three, maintaining the high accuracy of Neural Networks and SVMs while integrating the rule-based comprehensibility of Random Forests. This approach marks a substantial improvement over traditional single-method approaches, showcasing the potential of the ensemble method in achieving a balance between accuracy and interpretability.

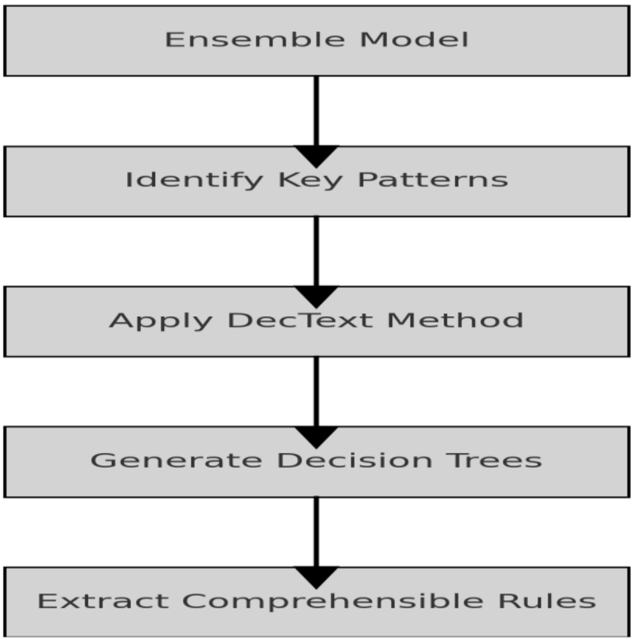

Figure 4: Rule Extraction Process

A flowchart illustrating the rule extraction process from the ensemble method.

Model Evaluation Metrics

The evaluation metrics for the proposed ensemble method are shown in Figure 5. The metrics include accuracy, precision, recall, and F1-score, providing a comprehensive evaluation of model performance.

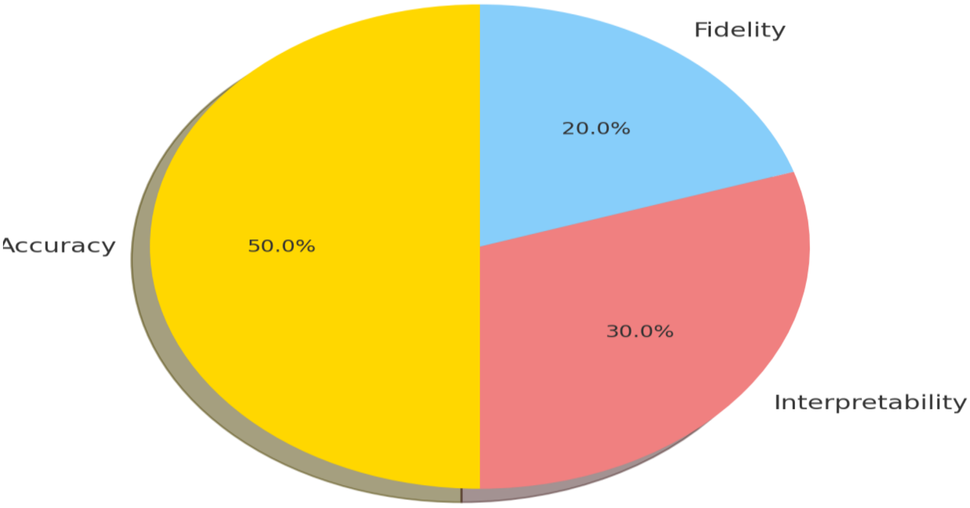

Figure 5: Model Evaluation Metrics

Conclusion

Neural Networks and Support Vector Machines are often seen as opaque or inscrutable. While their predictive accuracy is commendable, their interpretability is lacking. Decision tree induction methods like Random Forests produce more transparent predictive models that are less accurate compared to opaque models. This study explores elements that drive rule extraction, including justification, elucidation, modification, and enhancing prediction accuracy. A comprehensive survey of widely recognized rule extraction algorithms is provided. The primary contribution of the proposed research is developing an architecture for an ensemble method to rule extraction in data mining, transforming models from opaque to transparent.

Future Work

Building on the successful implementation of the ensemble method in Rule Extraction, future research will refine this approach for broader applications, including more complex datasets and different domains. Additionally, developing standardized guidelines for implementing this framework in various real-world scenarios will ensure effective, transparent, and justifiable decision-making. Future studies might also explore integrating this framework with cutting-edge technologies like deep learning to enhance interpretability for human users.

References

1. D. M. Escalante, M. A. Rodriguez, and A. Peregrin, "An Evolutionary Ensemble-based Method for Rule Extraction with Distributed Data," Proc. 4th Int. Conf. Hybrid Artif. Intell. Syst. (HAIS '09), 2009, pp. 638-645.
2. R. König, U. Johansson, and L. Niklasson, "Genetic Programming – A Tool for Flexible Rule Extraction," Proc. IEEE Congr. Evol. Comput., 2007, pp. 1304-1310.
3. U. Johansson et al., "Rule Extraction from Opaque Models--A Slightly Different Perspective," Proc. 5th Int. Conf. Mach. Learn. Appl. (ICMLA '06), Dec. 2006, pp. 22-27.

4. N. C. Oza, "Ensemble Data Mining Methods," Encyclopedia of Data Warehousing and Mining, 2nd ed., J. Wang, Ed. Hershey, PA: Idea Group Reference, 2009, pp. 448-453.
5. U. Johansson, R. König, and L. Niklasson, "Automatically balancing accuracy and comprehensibility in predictive modeling," Proc. 8th Int. Conf. Inf. Fusion, 2005.
6. L. Fu, "Rule learning by searching on adapted nets," Proc. Ninth Nat. Conf. Artif. Intell., 1991, pp. 590-595.
7. G. Towell and J. Shavlik, "The extraction of refined rules from knowledge-based neural networks," Mach. Learn., vol. 13, no. 1, pp. 71-101, 1993.
8. S. Thrun, "Extracting provably correct rules from artificial neural networks," Tech. Rep. iai-tr-93-5, Univ. Bonn, InstitutfürInformatik III, 1993.
9. M. W. Craven and J. W. Shavlik, "Using sampling and queries to extract rules from trained neural networks," Proc. Int. Conf. Mach. Learn., 1994, pp. 37-45.
10. M. W. Craven and J. W. Shavlik, "Extracting tree-structured representations of trained networks," Adv. Neural Inf. Process. Syst. 8, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. 1996, pp. 24-30.
11. O. Boz, "Converting A Trained Neural Network To A Decision Tree: DecText - Decision Tree Extractor," Ph.D. dissertation, Dept. Comp. Sci. Eng., Lehigh Univ., 2000.
12. N. Barakat and J. Diederich, "Eclectic rule-extraction from support vector machines," Int. J. Comput. Intell., vol. 2, no. 1, pp. 59-62, 2005.
13. D. Dancey, Z. A. Bandar, and D. McLean, "Rule extraction from neural networks for medical domains," Proc. 2010 Int. Joint Conf. Neural Networks (IJCNN), 2010, pp. 1-8.
14. T. Q. Huynh and J. A. Reggia, "Guiding Hidden Layer Representations for Improved Rule Extraction From Neural Networks," IEEE Trans. Neural Networks, vol. 22, no. 2, pp. 264-275, Feb. 2011.
15. J. R. Quinlan, "C4.5: Programs for Machine Learning," Morgan Kaufmann, 1993.
16. L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and Regression Trees," Wadsworth and Brooks, Monterey, CA, 1984.
17. Y. Hayashi, "A neural expert system with automated extraction of fuzzy if-then rules," Adv. Neural Inf. Process. Syst. 3, R. Lippmann, J. Moody, and D. Touretzky, Eds. Morgan Kaufmann, San Mateo, CA, 1991.
18. K. Saito and R. Nakano, "Medical diagnostic expert system based on PDP model," Proc. IEEE Int. Conf. Neural Networks, 1988, pp. 255-262.
19. S. I. Gallant, "Connectionist Expert Systems," Commun. ACM, vol. 31, no. 2, pp. 152-169, 1988.
20. Fisher, R. A. (1988). Iris. UCI Machine Learning Repository. https://doi.org/10.24432/C56C76.
21. Aeberhard, Stefan, and Forina, M. (1991). Wine. UCI Machine Learning Repository. https://doi.org/10.24432/C5PC7J.
22. Aha, David. (1991). Tic-Tac-Toe Endgame. UCI Machine Learning Repository. https://doi.org/10.24432/C5688J.