

CEN352 Term Project Report

Employee Attrition Prediction Agent Using Rule-Based, Machine Learning, and Hybrid Sequential Pipeline

Worked by: Alesia Dardha, Suada Terolli

SWE 3C

Introduction

Employee attrition is a critical problem in human resource management, as unexpected employee turnover can result in significant financial and organizational costs. Predicting employee attrition is challenging due to class imbalance, uncertainty in human behavior, and the interaction of multiple personal and workplace-related factors.

The objective of this project is to design and implement an intelligent agent capable of predicting employee attrition using multiple artificial intelligence techniques. Rather than relying on a single method, this project integrates:

- a rule-based logical reasoning system,
- a machine learning classifier (Random Forest), and
- an extra hybrid sequential system that combines both approaches.

Problem Formulation (PEAS Framework)

Performance Measure (P).

The dataset used in this project is highly imbalanced, as the majority of employees do not leave the company. In such scenarios, using accuracy as the primary evaluation metric can be misleading, since a model could achieve high accuracy simply by predicting the majority class. For this reason, accuracy was not used as the main performance indicator. Instead, the model is evaluated using Area Under the ROC Curve (ROC-AUC), along with precision, recall, and F1-score. ROC-AUC is particularly suitable because it measures the model's ability to distinguish between employees who leave and those who stay across different decision thresholds. Recall is emphasized to ensure that most high-risk employees are correctly identified, while the F1-score provides a balanced view of precision and recall.

Environment (E).

The environment is a static and partially observable setting based on historical human resources data. Not all factors influencing employee decisions (such as personal motivation or external job opportunities) are available, introducing uncertainty into the task. Additionally, the environment is stochastic, as employee attrition depends on complex and unpredictable human behavior rather than deterministic rules.

Actuators (A).

The agent's actuator is a binary classification decision, where the system outputs either *Attrition Risk* or *No Attrition*. In the hybrid system, this decision may be produced either by the rule-based component or by the machine learning model, depending on whether the rules are confident enough to make a prediction.

Sensors (S).

The sensors correspond to employee-related attributes provided by the dataset, including demographic information (such as age), work-related variables (such as overtime status and years at the company), and satisfaction indicators (such as job satisfaction and work-life balance). These inputs are used by the intelligent agent to assess the likelihood of employee attrition.

Dataset and Preprocessing

The project uses the HR Employee Attrition dataset, a widely used benchmark dataset in human resource analytics and machine learning research. This dataset was chosen because it represents a realistic and practically relevant problem domain, where organizations aim to understand and predict why employees leave their jobs. Employee attrition is a well-known challenge in industry, making this dataset suitable for evaluating intelligent agents designed for decision support. Another important reason for selecting this dataset is its inherent class imbalance, as the majority of employees do not leave the company. This characteristic reflects real-world conditions and makes the problem more challenging, requiring careful model design and evaluation. As a result, the dataset provides a meaningful context for applying advanced performance metrics and hybrid AI approaches rather than relying on simple accuracy-based models. The dataset contains a diverse set of features, including demographic information, compensation details, job satisfaction indicators, and work-related attributes such as overtime and years at the company. This variety of features allows the implementation of both rule-based reasoning (using interpretable conditions) and machine learning models capable of capturing complex, non-linear relationships.

Preprocessing Steps

To ensure data quality, reproducibility, and fair model evaluation, several preprocessing steps were applied:

- Irrelevant or non-informative identifier attributes (such as employee number and constant-valued fields) were removed, as they do not contribute to predicting attrition and could introduce noise into the models.
- Categorical variables were transformed using one-hot encoding, enabling machine learning algorithms such as Random Forest to process non-numeric data effectively.

- The target variable (*Attrition*) was converted into a binary format, where 1 represents employees who left the company and 0 represents those who stayed.
- A stratified train-test split was applied to preserve the original class distribution of the dataset in both training and testing sets. This is particularly important for imbalanced datasets, as it ensures that evaluation results are representative and reliable.
- Feature names were stored after preprocessing to guarantee consistency between training, evaluation, and deployment, especially when integrating the machine learning model into the hybrid system.

Overall, this preprocessing pipeline ensures that all models are trained and evaluated under consistent conditions, enabling a fair comparison between the rule-based system, the Random Forest classifier, and the hybrid sequential agent.

Intelligent Agent Design

The employee attrition prediction system is designed as an intelligent agent that integrates multiple AI techniques. Instead of relying on a single approach, the system combines logical reasoning and statistical learning, allowing it to benefit from both interpretability and predictive power. This design choice reflects the complexity of human decision-making, where both clear patterns and uncertain relationships coexist.

The agent is composed of three main components: a rule-based attrition agent, a Random Forest learning agent, and a hybrid sequential agent that combines both approaches in a structured decision process.

1. Rule-Based Attrition Agent (Logical Reasoning)

The rule-based attrition agent is a knowledge-driven system that applies expert-inspired logical rules to identify high-risk or very stable employees. This approach was chosen to ensure interpretability and transparency, which are particularly important in human resources applications where decisions must be explainable. Rather than using fixed thresholds, the rules are calibrated using quantiles derived from the training data. This makes the system adaptive to the dataset and prevents arbitrary threshold selection. The agent follows a conservative decision strategy: a prediction is only made when a rule fires with high confidence. If no rule applies, the agent deliberately abstains from making a decision.

Key characteristics of the rule-based agent include: Conservative and cautious decision-making, Explicit abstention when confidence is low. High interpretability, as each decision can be traced back to a specific rule

Evaluation Results (Test Set):

- Coverage: 32.31%

- Precision: 0.4595
- Recall: 0.7727
- F1-score: 0.5763

The high recall demonstrates that when the agent predicts attrition, it correctly identifies most true attrition cases. However, the relatively low coverage indicates that many employees do not satisfy any rule conditions. This confirms that while rule-based systems are effective for high-confidence cases, they are not sufficient as a standalone solution in complex domains.

2. Random Forest Learning Agent (Statistical Learning)

To complement the limitations of the rule-based agent, a Random Forest classifier was implemented as a learning agent. Random Forest was selected because it is well-suited for structured tabular data and offers several advantages: robustness to noise, the ability to model non-linear relationships, and built-in feature importance estimation. The dataset's class imbalance motivated specific design decisions. Instead of using the default classification threshold, a custom probability threshold of 0.30 was applied. This choice prioritizes recall, ensuring that a higher proportion of employees at risk of attrition are detected, which is more appropriate for preventative HR decision-making. Additionally, class weighting was used during training to reduce bias toward the majority class.

Evaluation Results:

- ROC-AUC: 0.7831
- Precision: 0.3617
- Recall: 0.7234
- F1-score: 0.4823

The Random Forest achieves full coverage and strong ranking performance, as reflected by the ROC-AUC score. However, the lower precision highlights the trade-off introduced by class imbalance and threshold tuning. This behavior is expected in scenarios where detecting potential attrition is prioritized over avoiding false positives.

3. Hybrid Sequential Agent (Rules-First Strategy)

The hybrid sequential agent was designed to combine the strengths of both symbolic reasoning and machine learning. Instead of merging predictions arbitrarily, a rules-first sequential decision pipeline is used: 1)The rule-based agent evaluates each employee first 2)If the rule-based agent abstains, the Random Forest provides the prediction 3)Rule-based decisions are never overridden

This design ensures that highly interpretable rule-based decisions are preserved while achieving complete coverage through machine learning fallback. From an AI perspective, this agent can be classified as a learning agent with a model-based component, as it integrates learned knowledge with predefined reasoning.

Evaluation Results:

- ROC-AUC: 0.7813
- Precision: 0.3495
- Recall: 0.7660
- F1-score: 0.4800

The hybrid agent achieves high recall comparable to the rule-based system while maintaining full coverage like the Random Forest. This balance makes it particularly suitable for real-world HR screening scenarios, where both interpretability and completeness are required.

Feature Importance Analysis

Feature importance analysis from the Random Forest model reveals that employee attrition is influenced primarily by factors related to compensation, workload, experience, and stability. The most influential features include monthly income, age, overtime status, total working years, years at the company, manager stability, and commute distance. These results align with domain knowledge in human resources and validate the realism of the model. Importantly, the same features were also used to design and refine the rule-based agent. This ensures consistency between the symbolic and statistical components, strengthening the overall system design and interpretability.

Comparative Analysis

Model	Coverage	Precision	Recall	F1	ROC-AUC
Rule-Based	32%	0.46	0.77	0.58	N/A
Random Forest	100%	0.36	0.72	0.48	0.78
Hybrid Sequential	100%	0.35	0.77	0.48	0.78

The comparison highlights the complementary nature of the three approaches. The rule-based agent excels in interpretability and recall but lacks coverage. The Random Forest offers strong predictive performance and complete coverage but lower interpretability. The hybrid system successfully balances these trade-offs by combining interpretability, recall, and completeness in a single agent.

Ethical and Societal Considerations

Predicting employee attrition raises important ethical concerns, particularly regarding bias, transparency, and misuse of predictions. Features such as age or income may unintentionally introduce discriminatory effects if used irresponsibly. Additionally, opaque machine learning decisions can reduce trust among employees. To mitigate these risks, the system emphasizes transparency through rule-based reasoning and interpretable feature importance. Predictions are intended to be used strictly as decision-support tools, not as automated or punitive judgments. Human oversight remains essential when applying such systems in organizational settings.

Ethical and Societal Considerations

Using artificial intelligence to predict employee attrition comes with several ethical concerns that must be carefully considered. Since the model relies on personal and work-related data, there is a risk that certain features, such as age, income level, or overtime status, could unintentionally lead to biased or unfair outcomes if the system is misused. For example, predictions could reinforce existing workplace inequalities if they are interpreted without proper context. Another important concern is transparency. Machine learning models, especially ensemble methods like Random Forests, can be difficult to interpret, which may reduce trust among employees and decision-makers. If individuals are affected by predictions they do not understand, this can create discomfort or resistance toward the system. To address these issues, this project places emphasis on interpretability and responsible use. The inclusion of a rule-based component allows decisions to be explained in simple and understandable terms, while feature importance analysis helps clarify which factors influence predictions. Most importantly, the system is intended to support human decision-making rather than replace it. Predictions should be used as early warning signals, not as automatic or punitive decisions, and human oversight should always remain a key part of the process.

Conceptual Evaluation Focus

1. Agent Design and Task Environment Analysis

The implemented system can be classified as an intelligent agent operating in a decision-support setting. More specifically, the overall system represents a learning agent with a model-based component, as it integrates both predefined logical rules and a learned statistical model.

- The rule-based agent corresponds to a simple reflex agent. It selects actions (attrition or no attrition) based on condition–action rules derived from observable employee attributes. These rules do not maintain an internal state beyond the current input, but they are calibrated using training data statistics, making them adaptive rather than purely static.
- The Random Forest classifier represents a learning agent, as it improves its decision-making behavior through experience (training data). It builds an internal model that maps

employee features to attrition risk probabilities and generalizes beyond explicitly defined rules.

- The hybrid sequential agent combines both approaches and can be viewed as a model-based learning agent, since it incorporates structured domain knowledge (rules) together with a learned predictive model. The agent follows a rules-first decision policy, ensuring that high-confidence symbolic knowledge is preserved while falling back on statistical learning when necessary.

Task Environment Characteristics

The agent operates in a partially observable environment, as not all factors influencing employee attrition (e.g., personal motivations, external job opportunities) are available in the dataset. The environment is static, because predictions are made based on historical snapshots of employee data rather than continuously changing states. It is also stochastic, since employee attrition is influenced by uncertain and probabilistic human behavior rather than deterministic rules.

The environment is discrete in terms of actions (attrition vs no attrition) and mostly discrete/continuous in perception, as the input features include both categorical variables (e.g., overtime status) and continuous variables (e.g., income, years at company).

2. Model Selection and Justification

Random Forest is particularly suitable for this task for several reasons:

- The dataset is tabular and heterogeneous, containing both numerical and categorical features. Random Forest handles such data naturally without requiring complex feature transformations.
- The problem involves non-linear relationships between features such as income, overtime, job satisfaction, and years at the company. Unlike linear models, Random Forest can capture these interactions effectively.
- The dataset is highly imbalanced, and Random Forest supports class weighting and probability threshold tuning, allowing recall-oriented decision-making without modifying the core algorithm.

In comparison to alternative models:

- Decision Trees alone are prone to overfitting and lack robustness.
- Support Vector Machines (SVMs) are sensitive to feature scaling and are less interpretable in high-dimensional categorical settings.
- Gradient boosting methods (e.g., XGBoost) offer strong performance but introduce higher complexity and reduced interpretability, which is undesirable in HR decision-support systems.

Random Forest provides a strong balance between predictive performance, robustness, interpretability, and computational efficiency, making it an appropriate choice for this application.

3. Correctness and Efficiency Considerations

From a correctness perspective:

- The rule-based agent is logically correct within its defined knowledge base, as each prediction follows explicitly defined and calibrated conditions.
- The learning agent's correctness is evaluated empirically using ROC-AUC, recall, precision, and F1-score, which are more appropriate than accuracy for imbalanced classification problems.
- The hybrid sequential agent improves correctness by ensuring that high-confidence rule-based decisions are never overridden while guaranteeing full coverage through the machine learning fallback.

From an efficiency perspective:

- The rule-based agent operates in constant time $O(1)$ per instance, making it highly efficient.
- The Random Forest classifier has prediction complexity proportional to the number of trees and their depth, but remains efficient enough for real-time HR screening scenarios.
- The hybrid agent introduces minimal overhead, as it first applies inexpensive rule checks and only invokes the learning model when necessary.

Overall, the system achieves a practical balance between decision correctness, computational efficiency, and real-world usability, which is essential for deployment in organizational decision-support environments.

Conclusion and Future Work

In this project, an intelligent agent for employee attrition prediction was designed and implemented using multiple artificial intelligence techniques. By combining a rule-based system with a machine learning model in a hybrid sequential approach, the system is able to balance interpretability, predictive performance, and full decision coverage. This demonstrates how symbolic reasoning and statistical learning can complement each other when applied to a real-world problem. The evaluation results show that while the rule-based system performs well in high-confidence cases, it is limited in coverage. The Random Forest model provides strong overall predictive ability, and the hybrid system successfully integrates the strengths of both approaches. This confirms that using multiple AI techniques together can lead to more robust and practical solutions.