CNR-IRCrES Working Paper

HERA: Hierarchical Engine for Rating Assessment

A Sector-Specific Hierarchical Artificial Intelligence Framework for One-Year-Ahead Corporate Rating Forecasting

4/2026

Greta Falavigna

CNR IRCrES Istituto di Ricerca sulla Crescita Economica Sostenibile

ISSN (online) 2421-7158
ISSN (print) 2421-6798

Direttore	Elena Ragazzi
Direzione	CNR-IRCrES *Istituto di Ricerca sulla Crescita Economica Sostenibile* Strada delle Cacce 73, 10135 Torino, Italy Tel. +39 011 3977612 segreteria.ircres@cnr.it www.ircres.cnr.it
Sede di Roma	Via dei Taurini 19, 00185 Roma, Italy Tel. +39 06 49937809 / Fax +39 06 49937808
Sede di Milano	Via Corti 12, 20121 Milano, Italy Tel. +39 02 23699501 / Fax +39 02 23699530
Sede di Genova	Corso Ferdinando Maria Perrone 24, 16152 Genova, Italy Tel. +39 010 6598798

https://www.ircres.cnr.it/produzione-scientifica/

CNR-IRCrES Working Paper 4/2026

HERA: Hierarchical Engine for Rating Assessment

A Sector-Specific Hierarchical Artificial Intelligence Framework for One-Year-Ahead Corporate Rating Forecasting

Greta Falavigna

Affiliazione: CNR-IRCrES, Consiglio Nazionale delle Ricerche, Istituto di Ricerca sulla Crescita Economica Sostenibile, Strada delle Cacce 73, 00135 Torino, Italia

corresponding author: greta.falavigna@cnr.it

Abstract

This study develops HERA (Hierarchical Engine for Rating Assessment), a sector-specific artificial intelligence framework for one-year-ahead corporate rating forecasting. The proposed methodology decomposes the original multi-class rating problem into a hierarchical sequence of binary classification tasks estimated through Random Forest models. Separate HERA architectures are developed for the Production, Trade, and Services sectors, allowing the framework to capture sector-specific financial characteristics and heterogeneous risk profiles.

The predictive performance of the classifiers is evaluated through Out-of-Bag validation, while optimal classification thresholds are determined through F1-score maximization. Beyond forecasting accuracy, the framework provides information on the financial indicators associated with specific rating transitions through variable importance analysis.

The empirical results show that the determinants of corporate ratings differ across both economic sectors and rating categories. Debt-related indicators emerge as key predictors for rating transitions involving the highest and lowest rating classes, whereas measures of financial sustainability and capital structure play a more prominent role in intermediate rating transitions.

Overall, HERA provides not only one-year-ahead rating forecasts but also insights into the sector-specific financial drivers associated with future rating outcomes. These findings may be of interest to firms, financial institutions, and policymakers seeking to better understand the determinants of future corporate creditworthiness.

Keywords: Random Forest, Corporate Credit Risk, Sectoral Analysis.

DOI: 10.23760/2421-7158.2026.004

How to Cite

Falavigna, G. (2026). *HERA: Hierarchical Engine for Rating Assessment. A Sector-Specific Hierarchical Artificial Intelligence Framework for One-Year-Ahead Corporate Rating Forecasting* (CNR-IRCrES Working Paper 4/2026). Istituto di Ricerca sulla Crescita Economica Sostenibile. http://dx.doi.org/10.23760/2421-7158.2026.004

SUMMARY

1. INTRODUCTION

Corporate rating systems play a fundamental role in the assessment of firms' financial soundness and creditworthiness. Traditionally, rating assignment relies on financial statement analysis, accounting ratios, expert judgment, and statistical techniques aimed at evaluating a firm's ability to meet its financial obligations. In recent years, the increasing availability of firm-level financial information, together with the rapid diffusion of artificial intelligence and machine learning methodologies, has encouraged the development of data-driven approaches capable of supporting and enhancing traditional rating systems.

Despite the growing application of machine learning techniques to corporate risk assessment, many existing studies estimate predictive models on highly heterogeneous samples of firms belonging to different industries. Such an approach implicitly assumes that the relationship between financial indicators and credit quality remains stable across economic sectors. However, firms operating in production, trade, and service activities are characterized by different business models, operating structures, profitability dynamics, asset compositions, and financing patterns. Consequently, the factors that influence rating outcomes may vary substantially across sectors.

Sectoral heterogeneity raises an important issue for corporate rating analysis. If firms exhibit different financial structures and economic characteristics, it is reasonable to expect that accounting indicators do not contribute equally to rating determination across industries. Measures of profitability, leverage, indebtedness, financial sustainability, and capital structure may play distinct roles depending on the sector in which a firm operates. Therefore, understanding whether the determinants of rating assignments are sector-specific becomes as relevant as achieving high predictive accuracy.

This paper addresses these issues through the development of HERA (Hierarchical Engine for Rating Assessment), a sector-specific machine learning framework designed to forecast firms' future rating classes. The proposed architecture follows a hierarchical approach in which firms are first grouped into homogeneous macro-sectors according to their ATECO classification and are subsequently evaluated using dedicated sector-specific Random Forest models. Unlike traditional rating approaches based on a single predictive model, HERA explicitly incorporates sectoral heterogeneity within the rating process.

A further distinguishing feature of the framework is its predictive nature. Financial indicators are observed at time t, whereas the corresponding rating class is observed at time t+1. Consequently, HERA does not simply assign contemporaneous rating classes but performs a genuine one-year-ahead forecasting exercise, identifying the financial characteristics associated with firms' future creditworthiness.

The study is driven by the following research questions:

- **RQ1:** Can a sector-specific hierarchical architecture improve the forecasting of firms' future rating classes?
- **RQ2:** Do the financial determinants of rating transitions differ across economic sectors and rating categories?

To answer these questions, the paper develops a hierarchical network of binary Random Forest classifiers specifically designed to reproduce the ordinal structure of corporate ratings. The proposed architecture transforms a complex ten-class rating problem into a sequence of interconnected binary classification tasks, each calibrated through threshold optimization based on the F1-score. This strategy aims to improve classification stability while preserving the informational content of the original rating scale.

The contribution of this study is threefold. First, it introduces HERA, a novel hierarchical framework for one-year-ahead corporate rating forecasting. Second, it demonstrates how sectoral heterogeneity can be incorporated directly into the machine learning architecture through the estimation of dedicated sector-specific rating engines. Third, it investigates the relative

importance of financial indicators across sectors and rating transitions, providing evidence on the financial drivers that contribute most to the improvement or deterioration of firms' rating positions.
Beyond its predictive capabilities, HERA offers an interpretable framework that may support managers, analysts, and policymakers in identifying the financial dimensions that are most closely associated with future rating improvements. By combining sector specialization, hierarchical classification, and variable importance analysis, the framework provides insights not only into rating prediction but also into the economic mechanisms underlying firms' credit quality.
The remainder of the paper is organized as follows. Section 2 reviews the literature on corporate rating models and machine learning applications in credit risk assessment. Section 3 describes the dataset, the construction of financial indicators, and the sector classification procedure. Section 4 presents the HERA methodology and the hierarchical architecture of the framework. Section 5 discusses the empirical results and investigates the sector-specific determinants of rating assignments. Finally, Section 6 concludes and outlines directions for future research. The paper is complemented by two appendices presenting supplementary results that support and strengthen the empirical analysis.

2. Short literature review

Corporate rating models have long represented a fundamental tool for assessing firms' financial soundness and creditworthiness. Early contributions relied primarily on accounting-based indicators and statistical approaches aimed at predicting financial distress and default events. The pioneering work of Altman (1968) demonstrated the predictive value of financial ratios through discriminant analysis, laying the foundations for modern credit risk modeling. Subsequent studies expanded the range of methodologies employed in rating assessment, incorporating logistic regression, survival analysis, and various multivariate statistical techniques (Altman, 1968; Ohlson, 1980).
The growing availability of firm-level data and computational resources has stimulated increasing interest in machine learning approaches for credit risk assessment and rating prediction. Compared with traditional statistical techniques, machine learning models are capable of capturing complex nonlinear relationships and high-order interactions among explanatory variables. Among the most widely adopted machine learning techniques, Artificial Neural Networks (ANNs) have shown promising results in bankruptcy prediction, credit scoring, and corporate rating applications due to their ability to model complex and non-linear relationships among financial variables (Tam & Kiang, 1992; Atiya, 2001). In the context of small and medium-sized enterprises, Falavigna (2012) demonstrated that neural networks can effectively reproduce rating judgments even when only limited financial information is available. Using a multi-layer architecture composed of seven feed-forward neural networks, the proposed framework successfully simulated Bureau van Dijk rating assignments for Italian small firms, highlighting the capability of neural networks to learn and replicate complex rating mechanisms. More recently, Calabrese et al. (2024) employed neural network algorithms to predict firms' short-term financial constraints, showing that machine learning models optimized with sensitivity and specificity criteria can achieve satisfactory classification performance while also supporting the identification of financially vulnerable sectors and geographical areas.
Nevertheless, neural network models are often criticized for their limited interpretability and their sensitivity to parameter selection and data imbalance.
In last years, tree-based methods and ensemble learning algorithms have gained considerable attention in financial applications. Random Forests, introduced by Breiman (2001), combine bootstrap aggregation and random feature selection to generate robust classification models characterized by low variance and strong predictive performance. Several studies report that Random Forests frequently outperform traditional statistical approaches and perform competitively with neural networks in credit risk and default prediction tasks, while

simultaneously providing greater interpretability through variable importance measures (Lessmann et al., 2015; Zhu et al., 2019). Furthermore, the availability of Out-Of-Bag estimates allows an efficient assessment of predictive performance without requiring an independent validation sample.

Despite the growing body of evidence supporting the use of machine learning techniques in financial classification problems, relatively limited attention has been devoted to the role of sectoral heterogeneity in rating assessment. Firms operating in manufacturing, trade, and service sectors differ substantially in terms of asset structure, operating cycles, leverage patterns, and profitability dynamics. As a consequence, financial ratios that prove informative in one sector may exhibit limited explanatory power in another. Previous studies have emphasized the importance of industry-specific effects in financial distress prediction and corporate performance analysis (Platt & Platt, 1991; Grice & Ingram, 2001), suggesting that industry heterogeneity should be explicitly considered when developing predictive rating models.

This study contributes to the existing literature by proposing a hierarchical sector-specific framework for corporate rating assessment. In contrast to traditional one-size-fits-all approaches, the proposed HERA framework estimates dedicated machine learning models for homogeneous groups of firms. While previous studies have successfully employed neural networks to reproduce rating judgments (Falavigna, 2012) and to identify financially constrained firms (Calabrese et al., 2024), relatively little attention has been devoted to sector-specific machine learning architectures and to the possibility that financial indicators may exhibit different predictive relevance across economic sectors. The HERA framework seeks to address this gap by explicitly incorporating sectoral heterogeneity into the rating process.

In addition to evaluating predictive performance, the analysis investigates whether the relative importance of financial indicators differs across economic sectors, thereby providing further evidence on the sector-specific nature of rating determinants.

3. Dataset and Variable construction

3.1. Dataset

The empirical analysis is based on a dataset of Italian firms operating in the production, trade, and service sectors. Financial information was collected from firms' annual financial statements and was subsequently used to construct a set of accounting-based indicators commonly employed in corporate risk assessment and rating analysis. All data, financial accounts as well as rating scores, have been extracted from AIDA database provided by Bureau van Dijk.

The initial database comprises firms belonging to different sectors of economic activity, identified through the ATECO classification system.

Following a hierarchical sectoral approach, firms were grouped into three macro-sectors using the first two digits of the ATECO classification. The Production sector includes firms with ATECO codes from 01 to 43, covering primary activities, mining and quarrying, manufacturing industries, utilities, and construction activities. The Trade sector includes firms with ATECO codes from 45 to 47, corresponding to wholesale and retail trade and the repair of motor vehicles. The Services sector includes firms with ATECO codes from 49 to 99, covering transportation and storage, accommodation and food services, information and communication, professional activities, administrative services, education, health services, cultural activities, and other service activities. The resulting classification identifies three homogeneous groups of firms, namely Production, Trade, and Services, which represent the first decision level of the HERA architecture. Table 1 summerizes the classification of firms based on ATECO 2-digit codes.

Table 1: Classification of firms in Manufacturing, Trade and Services based on ATECO 2 digits code

Macro-sector	ATECO 2-digit codes	Description
Manufacturing	01-43	Primary activities, manufacturing, utilities and construction
Trade	45-47	Wholesale and retail trade; repair of motor vehicles
Services	49-99	Transportation, hospitality, ICT, professional activities, health, education and other services

This classification constitutes the first level of the HERA framework and reflects the assumption that firms operating in different economic environments may exhibit distinct financial structures, profitability patterns, and risk profiles. Table 2 presents the distribution of firms across sectors and years. The sample is predominantly composed of firms operating in the services sector, which consistently represents the largest share of observations. Production firms constitute the second largest group, while trade firms account for the smallest proportion of the sample over the entire period under analysis.

Table 2: Descriptive statistics of sample by sector and year (frequency and percentage values)

Year	Sector			
	Production	**Trade**	**Services**	**Total**
2016	210,939	115,505	288,052	614,496
	34.33%	18.8%	46.88%	100%
2017	210,939	115,505	288,052	614,496
	34.33%	18.8%	46.88%	100%
2018	210,939	115,505	288,052	614,496
	34.33%	18.8%	46.88%	100%
2019	210,939	115,505	288,052	614,496
	34.33%	18.8%	46.88%	100%
2020	210,939	115,505	288,052	614,496
	34.33%	18.8%	46.88%	100%
2021	210,939	115,505	288,052	614,496
	34.33%	18.8%	46.88%	100%
2022	210,939	115,505	288,052	614,496
	34.33%	18.8%	46.88%	100%
2023	210,939	115,505	288,052	614,496
	34.33%	18.8%	46.88%	100%
Total	1,687,512	924,040	2,304,416	4,915,968
	34.33%	18.8%	46.88%	100%

The observation period covers the years 2016-2023, yielding an initial sample of 614,676 firm-year observations. Prior to model estimation, several data-cleaning procedures were performed. Observations characterized by missing values, undefined accounting ratios, or non-finite values resulting from ratio calculations were excluded from the analysis. This procedure ensured the consistency and reliability of the financial indicators employed in the machine learning models.
The final dataset was subsequently divided according to sector and used to estimate sector-specific rating models. Unlike traditional approaches based on a single pooled sample, the

proposed methodology adopts separate models for each macro-sector in order to capture industry-specific financial dynamics.

3.2. Target Variable

The objective of the HERA framework is to predict firms' future rating classifications. To ensure a genuine forecasting exercise, financial indicators are computed using accounting information observed at year *t*, while the target variable corresponds to the rating class observed at year *t+1*. Therefore, the proposed models perform a one-year-ahead prediction of corporate credit quality. The target variable is derived from the official credit rating assigned to each firm and is transformed into a binary classification problem in order to estimate specialized rating models. Corporate rating information is obtained from the AIDA database (Bureau van Dijk), which assigns firms to a ten-level rating scale reflecting their financial soundness and credit risk. Table 3 presents the rating classes and their corresponding qualitative interpretations.

Table 3. AIDA Rating Classification and Qualitative Interpretation

Rating	**Description**
AAA	Highest credit quality and very low default risk
AA	Very high credit quality
A	High credit quality and strong repayment capacity
BBB	Adequate financial strength and moderate risk
BB	Speculative grade with increasing credit risk
B	Highly speculative grade
CCC	Substantial credit risk
CC	Very high probability of default
C	Extremely vulnerable financial condition
D	Default or severe financial distress situation

Source: Adapted from AIDA (Bureau van Dijk) corporate rating definitions.

The rating classes from AAA to BBB are generally considered investment-grade categories, while ratings from BB downward identify increasingly speculative and financially vulnerable firms.
In order to estimate the machine learning models, the original rating scale was transformed into a sequence of binary classification problems by grouping adjacent rating categories. This approach allows the identification of progressively deteriorating credit conditions while preserving the ordinal nature of the original rating scale.

3.3. Financial Indicators

The explanatory variables employed in this study are based on a set of accounting indicators originally proposed by Falavigna (2012) for the simulation of corporate rating assessments using limited financial information. The selection of these variables is motivated by their ability to capture the main dimensions of firms' economic and financial conditions, including profitability, leverage, indebtedness, efficiency, and capital structure.
The adoption of the same set of indicators also ensures the continuity of the methodological framework developed in previous research, allowing a direct comparison between traditional Artificial Neural Network approaches and the machine learning techniques implemented in the HERA framework.
More importantly, the choice of these variables reflects the practical objective of extending rating assessment to firms characterized by limited financial disclosure. In the Italian context, a large number of small and medium-sized enterprises are allowed to file abbreviated financial

statements, which provide substantially less information than ordinary reports. The selected indicators were specifically designed to be computable even in the presence of simplified accounting information, making the proposed framework suitable for the evaluation of firms for which detailed financial data are unavailable (Falavigna, 2012).
This characteristic is particularly relevant for rating evaluation in SME-dominated economic systems, where information asymmetries frequently limit the application of more data-intensive risk assessment methodologies.
While the indicators remain identical across all models, their relative contribution to rating prediction is allowed to vary across sectors. This feature enables the investigation of whether the financial determinants of corporate ratings differ between production, trade, and service firms.
Specifically, the accounting variables required to compute the financial ratios include subscribed capital unpaid by shareholders, total net fixed assets, total current assets, shareholders’ equity, total provisions for risks and charges, employee severance indemnity provisions, total liabilities, value of production, cost of production, and financial expenses, together with the firm’s ATECO code, which is used to assign firms to the relevant macro-sector in the hierarchical architecture.
From these accounting variables, the model derives the following financial indicators:

1. DEBTA - Total Debt to Total Assets Ratio

$$DEBTA = \frac{TD}{Fixed\ Assets + Current\ Assets}$$

where TD represents total debt (the sum of total provisions for risks and charges, employee severance indemnity provisions, and total liabilities).
This index measures the overall level of indebtedness relative to total assets. Higher values indicate greater financial leverage.

2. DEBPROD - Debt to Production Ratio

$$DEBPROD = \frac{TD}{Value\ of\ Production}$$

The index measures the amount of debt supported by the firm’s productive activity. Higher values may indicate a heavier debt burden relative to operating capacity.

3. ROS - Return on Sales

$$ROS = \frac{Value\ of\ Production - Costs}{Value\ of\ Production}$$

The ROS measures operating profitability and the firm’s ability to generate earnings from sales.

4. ROA - Return on Assets

$$ROA = \frac{Value\ of\ Production - Costs}{Fixed\ Assets + Current\ Assets}$$

The ROA evaluates the profitability generated by the firm’s asset base.

5. DEBCOV - Debt Coverage Ratio

$$DEBCOV = \frac{Value\ of\ Production - Costs}{TD}$$

This indicator measures the firm's ability to generate operating earnings relative to total debt. Higher values indicate stronger debt-servicing capacity.

6. FAPROD - Fixed Assets to Production Ratio

$$FAPROD = \frac{Fixed\ Assets}{Value\ of\ Production}$$

The index captures the intensity of fixed capital employed in productive activities. Higher values generally characterize more capital-intensive firms.

7. FCCOV - Financial Charges Coverage Ratio

$$FCCOV = \frac{Financial\ Expenses}{Value\ of\ Production - Costs}$$

The indicator measures the weight of financial expenses relative to operating income. Higher values may signal greater financial pressure.

8. FAEQ - Fixed Assets to Net Equity

$$FAEQ = \frac{Fixed\ Assets}{Net\ Equity + Provisions - Unpaid\ Subscribed\ Capital}$$

This ratio evaluates the extent to which fixed investments are financed by long-term sources of funding. Lower values generally indicate a more balanced financial structure.

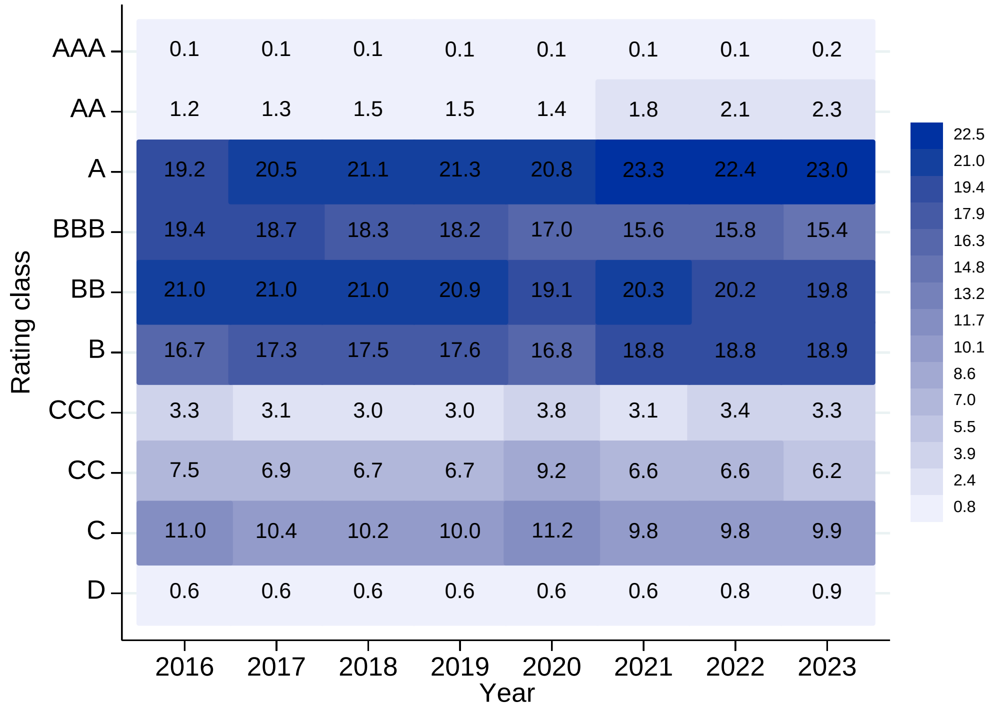
Table 4 shows descriptive statistics of indexes for the whole sample.

Table 4: Descriptive statistics of indicators for the whole sample (2016-2023).

Variable	Obs	Mean	Std. Dev.	Min	Max
DEBTA	4,902,893	4.769	88.677	0	53,395
DEBPROD	4,615,616	25.586	2591.562	-14,978	2,254,653
ROS	4,615,616	-0.925	30.702	-18,280	4,127
ROA	4,902,879	0.913	38.396	-21,606	10,888
DEBCOV	4,877,733	0.768	34.902	-5,120	18,619
FAPROD	4,615,616	33.237	6156.114	-131,714	6,135,180
FCCOV	4,767,208	1.975	267.338	-228,126	215,490
FAEQ	4,886,544	14.333	775.703	-93,027	1,587,577

Figure 1 reports the annual distribution of firms across the ten AIDA rating classes from 2016 to 2023. Values are expressed as percentages of the total number of firms observed in each year. Lighter shades indicate a lower share of firms belonging to a given rating category, whereas darker shades indicate larger frequencies. The figure provides a visual representation of the evolution of rating composition over time. Each cell of the heatmap represents the percentage of firms classified into a given rating category relative to the total number of firms observed in the corresponding year. For completeness, the underlying percentages are reported in Table 5.

Figure 1. Distribution of Firms Across Rating Classes by Year (%)



The distribution of ratings reveals that most Italian firms are concentrated in intermediate credit-quality classes, particularly BB and BBB. Nevertheless, the evidence suggests a gradual improvement in firms' financial conditions over time. Starting from 2018, a larger share of firms is observed in the A rating category, indicating an increasing concentration of firms in higher-quality rating classes. Although the proportion of firms assigned to the highest ratings (AA and AAA) remains relatively modest, the share of firms belonging to the most critical categories (C and D) is also limited, suggesting a generally satisfactory level of financial soundness within the sample.

Table 5. Percentage Distribution of Firms Across AIDA Rating Classes by Year

Rating Class	**Year**								
	2016	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**Total**
AAA	0.07	0.08	0.10	0.10	0.08	0.11	0.13	0.16	0.10
AA	1.16	1.35	1.47	1.55	1.39	1.82	2.08	2.30	1.64
A	19.24	20.53	21.08	21.30	20.77	23.29	22.40	23.04	21.46
BBB	16.66	17.29	17.53	17.58	16.79	18.80	18.83	18.93	17.80
BB	21.04	21.01	21.01	20.89	19.06	20.28	20.19	19.83	20.41
B	19.42	18.73	18.32	18.22	17.05	15.60	15.80	15.42	17.32
CCC	11.05	10.39	10.22	10.04	11.23	9.76	9.83	9.87	10.30
CC	7.48	6.95	6.72	6.69	9.20	6.61	6.61	6.18	7.05
C	3.28	3.13	3.00	3.03	3.79	3.09	3.35	3.35	3.25
D	0.61	0.56	0.56	0.60	0.64	0.65	0.79	0.92	0.67
Total	100	100	100	100	100	100	100	100	100

Considering the firm distribution by sector, Figure 2 presents the heatmap of percentage values. Consistent with earlier evidence, the majority of firms are clustered in the intermediate rating categories, while the Services sector appears to be characterized by a more favorable credit quality profile.

Figure 2. Distribution of Firms Across Rating Classes by Sector (%)

Table 6 reports the descriptive statistics of the financial indicators employed in the analysis. The evidence highlights substantial differences across sectors, confirming the existence of heterogeneous financial structures among production, trade, and service firms. These differences provide initial support for the adoption of a sector-specific modelling strategy within the HERA framework.

Table 6. Mean Values and Standard Deviations of Financial Indicators Across Sectors

Financial indexes	**Production**		**Trade**		**Services**		**Total**	
	mean	**sd**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
DEBTA	6.383	110.163	3.348	67.493	4.154	77.798	4.769	88.677
DEBPROD	22.573	541.556	10.335	302.220	34.214	3,774.066	25.586	2,591.562
ROS	-0.862	30.715	-0.534	22.751	-1.135	33.472	-0.925	30.702
ROA	1.420	40.885	1.340	46.246	0.368	32.581	0.913	38.396
DEBCOV	0.891	33.908	0.921	42.182	0.615	32.282	0.768	34.902
FAPROD	10.356	443.809	3.667	139.728	62.559	9,036.470	33.237	6,156.114
FCCOV	2.822	299.974	2.870	293.989	0.985	227.419	1.975	267.338
FAEQ	12.332	255.680	6.255	119.983	19.059	1,110.268	14.333	775.703

4. Methodology: A Network of Random Forest

Among the machine learning techniques considered in this study, Random Forests were selected because of several characteristics that make them particularly suitable for corporate rating assessment. First, financial statement data are typically characterized by non-linear relationships, interaction effects, and heterogeneous distributions that are difficult to model through conventional statistical approaches. Random Forests are capable of capturing these complex patterns without requiring restrictive assumptions regarding the functional form of the relationships between explanatory variables and rating outcomes.

Second, the rating classes employed in this study are characterized by an uneven distribution of observations. Ensemble tree-based methods have been shown to be relatively robust in the presence of imbalanced classification problems and are generally less sensitive to distributional assumptions than many alternative techniques.

Third, Random Forests provide an effective mechanism for controlling overfitting through bootstrap aggregation and random feature selection. By combining a large number of decision trees estimated on different subsamples of the data, the algorithm reduces prediction variance while preserving classification accuracy.

A further advantage of Random Forests is their ability to provide measures of variable importance. This feature is particularly relevant for the objectives of the present study, as it allows the identification of the financial indicators that contribute most to rating classification within each sector. Consequently, the methodology not only supports accurate prediction but also provides useful insights into the sector-specific determinants of corporate ratings.

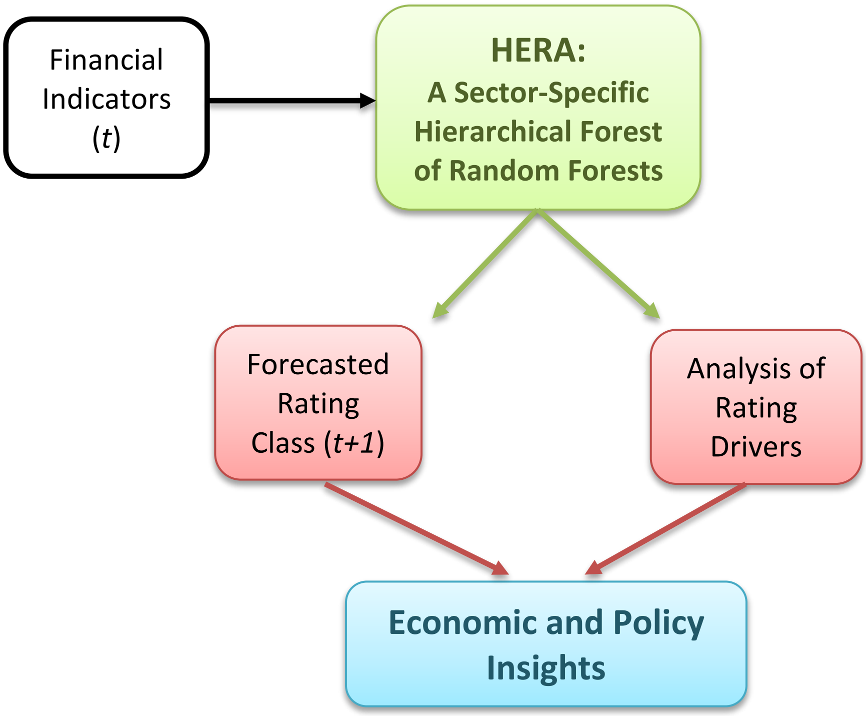
Finally, Random Forests generate probabilistic classification scores that can be subsequently calibrated through alternative threshold-selection criteria, including the F1-score and the Youden Index. This characteristic is particularly valuable in the context of corporate rating systems, where the costs associated with classification errors may differ substantially across rating classes.

For these reasons, Random Forests represent a suitable methodology for the development of the HERA framework, combining predictive performance, robustness, interpretability, and flexibility in the presence of heterogeneous sectoral characteristics.

Preliminary experiments were also conducted using Artificial Neural Networks (ANNs). Although these models provided satisfactory classification results, the final HERA architecture relies on Random Forest classifiers because of their greater interpretability, robustness, and ability to provide variable importance measures. Summary ANN results are reported in Appendix A.

Figure 3 summarizes the conceptual architecture of the HERA framework. Financial indicators observed at time *t* constitute the input of a sector-specific hierarchical ensemble of Random Forest classifiers. The framework simultaneously produces two outputs: a one-year-ahead forecast of firms' rating classes and an analysis of the financial drivers underlying rating transitions. The integration of these complementary perspectives allows HERA not only to predict future creditworthiness outcomes, but also to provide economic and policy insights regarding the financial characteristics associated with subsequent rating improvements or deteriorations.

Figure 3. HERA: From Financial Indicators to Rating Forecasts and Economic Insights



4.1. Random Forest Estimation

A Random Forest is an ensemble learning technique that combines a large number of decision trees to perform classification tasks. Each tree is estimated on a different bootstrap sample of the original dataset, while a random subset of predictors is considered at each split. The final prediction is obtained by aggregating the classifications generated by all trees through a majority-voting mechanism. The Random Forest classifiers were estimated using the original financial indicators. Unlike distance-based or gradient-based machine learning techniques, Random Forests do not require variable standardization, as tree-based models are generally insensitive to differences in measurement scales.
The mathematical formalization of Random Forest model, is the following:

$$RF(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

where B denotes the number of trees in the forest and $T_b(x)$ is the prediction generated by the b-th decision tree. In classification problems, the final class is assigned according to the majority vote of the trees.
This approach reduces overfitting, improves predictive stability, and allows complex non-linear relationships among variables to be captured without requiring strong distributional assumptions (Breiman, 2001).
A distinctive feature of the Random Forest algorithm is the availability of an internal validation procedure known as the Out-of-Bag (OOB) evaluation. Each decision tree is estimated using a bootstrap sample randomly drawn from the original dataset. Since observations are sampled with replacement, a fraction of the original data is not included in the estimation of a given tree. On average, approximately one-third of the observations remains excluded from the bootstrap sample and can therefore be used as an independent validation set.
For each observation, the OOB prediction is obtained by aggregating the votes of all trees that did not use that observation during the estimation phase. Consequently, OOB predictions provide an

internal assessment of predictive performance without requiring a separate hold-out sample. The resulting OOB measures are commonly considered reliable estimates of the model's out-of-sample classification ability.

In this study, the performance of each Random Forest classifier composing the HERA framework is evaluated through OOB predictions. Specifically, Accuracy, Error Rate, and Area Under the Receiver Operating Characteristic Curve (AUC) are computed on the basis of OOB classifications, thus providing an estimate of the model's generalization capability.

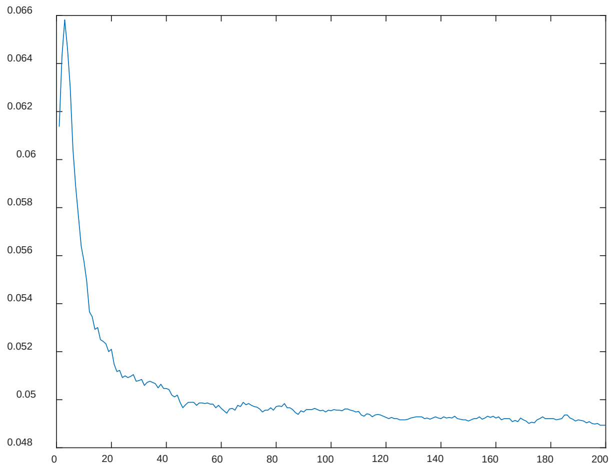
In addition to providing an internal validation procedure, the OOB framework can be used to evaluate the stability of the Random Forest as the number of trees increases. As additional trees are incorporated into the ensemble, the OOB error typically decreases and progressively converges towards a stable value. The convergence of the OOB error indicates that the classifier has reached a satisfactory level of predictive stability and that further increases in the number of trees are unlikely to produce substantial improvements in performance.

To illustrate this process, Figure 4 reports the evolution of the OOB error for the Random Forest that in the HERA framework is responsible for distinguishing between the AAA and AA rating classes. Estimations are provided separately for the Production, Trade, and Services sectors (see Figure 5). Figure 4 presents the OOB error convergence for RF8 (AAA versus AA) across the three sector-specific HERA engines: (a) Production, (b) Trade, and (c) Services. The x-axis reports the number of trees included in the ensemble, while the y-axis shows the corresponding OOB classification error. Each Random Forest classifier was estimated using 200 trees. Figure 3 shows that the OOB error decreases rapidly during the initial stages of model estimation and subsequently stabilizes as the number of trees increases. This behaviour is observed across all three sectors, confirming the stability of the Random Forest classifiers adopted within the HERA framework. The minimum OOB error identifies the optimal forest size retained for the corresponding classifier. Beyond this point, additional trees provide only marginal improvements in predictive performance while increasing computational complexity. Overall, the convergence pattern provides evidence of the robustness of the sector-specific Random Forest models and supports the suitability of the selected forest size for the HERA architecture.

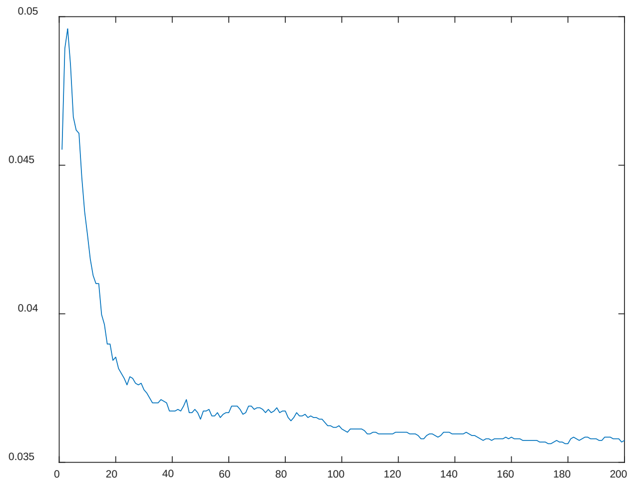
Consequently, the adoption of Random Forests serves a dual purpose within HERA: improving rating forecasting performance and providing interpretable evidence on the sector-specific financial drivers underlying rating transitions. This dual perspective represents one of the main strengths of the proposed framework.

Figure 4. OOB Error Convergence of RF8 (AAA versus AA) Across the Three Sector-Specific HERA Engines

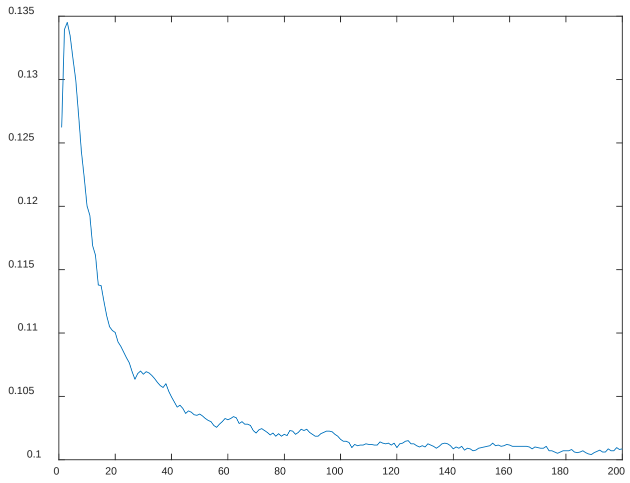
(a): Production: Number of trees: 198: Minimum Error OOB = 4.89%



(b) Trade: Number of trees: 173: Minimum Error OOB = 3.57%



(c) Services: Number of trees: 189: Minimum Error OOB = 10.08%



4.2. Threshold Optimization and Performance Evaluation

The performance of each binary Random Forest classifier was evaluated through three complementary measures: Accuracy, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC, Fawcett 2006).

Accuracy measures the proportion of correctly classified observations and is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP, and FN denote true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), respectively.

Although Accuracy is one of the most widely used classification metrics, it may provide misleading results in the presence of highly unbalanced classes. In such cases, a classifier may achieve a high level of accuracy simply by assigning most observations to the dominant class, while exhibiting poor predictive ability for the minority class.

Threshold optimization plays a crucial role in this context because the ultimate goal is not solely the minimization of an objective function, but the identification of decision boundaries that support robust and balanced classification outcomes. In imbalanced rating datasets, relying exclusively on aggregate performance metrics may lead to thresholds that favor the most frequent classes. The optimization procedure therefore seeks a threshold that maximizes predictive effectiveness while preserving the model's ability to accurately distinguish firms across different rating categories.

For this reason, the threshold optimization procedure was based on the F1-score (Sokolova and Lapalme, 2009), which combines precision and recall into a single measure:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

The F1-score corresponds to the harmonic mean of precision and recall and reaches its maximum value only when both measures are simultaneously high. Consequently, it provides a more balanced assessment of classification performance than overall accuracy, particularly when the distribution of observations across classes is uneven.

Following the threshold calibration approach proposed by Falavigna (2012), a grid-search procedure was implemented to identify the optimal classification threshold. However, unlike the original framework, which selected the threshold by minimizing the overall classification error, the present study adopts the threshold that maximizes the F1-score. This modification is particularly relevant in rating applications, where the cost of misclassifying financially fragile firms may be substantially higher than the cost of other classification errors. Therefore, maximizing the F1-score allows the model to achieve a more balanced trade-off between false positives and false negatives.

To assess the discriminant ability of the classifiers independently of the selected threshold, the analysis also considers the Area Under the Receiver Operating Characteristic Curve (AUC). The ROC curve plots the True Positive Rate (Sensitivity) against the False Positive Rate for all possible threshold values. The AUC can be interpreted as the probability that the model assigns a higher score to a randomly selected positive observation than to a randomly selected negative observation.

An AUC value equal to 0.5 indicates performance equivalent to random classification, whereas an AUC equal to 1 denotes perfect discrimination. Since AUC does not depend on any specific threshold, it provides a robust measure of the overall ranking capability of the model and represents a useful complement to Accuracy and F1-score in the evaluation of rating prediction systems.

In addition to Accuracy and F1-score, the analysis also considers the Youden Index (Youden, 1950; Habibzadeh, 2025), a widely adopted measure for evaluating binary classifiers. The Youden Index is defined as:

$$Youden = Sensitivity + Specificity - 1$$

where:

$$Sensitivity = \frac{TP}{TP + FN}$$

and

$$Specificity = \frac{TN}{TN + FP}$$

The Youden Index ranges from 0 to 1, with higher values indicating a better balance between the ability to correctly identify positive observations and the ability to correctly identify negative ones. A value of zero corresponds to a classifier performing no better than random classification, whereas a value of one denotes perfect discrimination. Similar algorithm is provided by Falavigna et al. (2019).
Although the Youden Index is computed for all threshold values and provides useful information on the trade-off between sensitivity and specificity, the present study adopts the threshold that maximizes the F1-score. This choice is motivated by the characteristics of the rating classification problem, where class frequencies are often unbalanced and the correct identification of the minority class is particularly relevant. While the Youden Index assigns equal importance to sensitivity and specificity, the F1-score explicitly balances precision and recall, providing a more informative measure of classification quality in the presence of class imbalance. Consequently, F1-based threshold calibration is deemed more appropriate for the development of the HERA framework.
Therefore, AUC is used to evaluate the overall discriminating ability of the model, whereas F1-score is used to identify the optimal decision threshold for operational rating assignment.

4.3. The Hierarchical Architecture of HERA: A Forest of Random Forests

Furthermore, although the original rating variable is expressed on a ten-class ordinal scale ranging from AAA to D, the estimation strategy does not rely on a single multiclass Random Forest model. Instead, HERA adopts a hierarchical network of binary Random Forest classifiers, each designed to distinguish between two adjacent groups of rating categories.
This choice is motivated by both methodological and practical considerations. Binary classification models are generally more stable, easier to calibrate, and less affected by class imbalance than highly fragmented multiclass classifiers. Moreover, the decomposition of the rating assignment process into a sequence of binary decisions preserves the ordinal nature of the rating scale and reduces the risk of misclassification across distant rating categories.
In addition, the hierarchical architecture was adopted for both methodological and economic reasons. First, corporate ratings are naturally ordered according to different levels of credit risk, suggesting a hierarchical rather than a purely multiclass representation of the classification problem. Second, the decomposition of the original rating scale into a sequence of binary classification tasks reduces the complexity of the learning problem, allowing each Random Forest classifier to focus on a more homogeneous subset of firms. Finally, the hierarchical structure provides a meaningful economic interpretation of the classification process, as each node corresponds to a specific rating boundary and captures different stages of financial deterioration or improvement. The resulting network of Random Forest classifiers therefore combines predictive flexibility with an interpretable representation of corporate credit risk dynamics.
The classification threshold associated with each Random Forest is not fixed a priori. Following the threshold optimization procedure described above, an optimal threshold is estimated for every binary classifier through F1-score maximization. As a result, each node of the HERA architecture is characterized by a dedicated Random Forest model and a corresponding optimized decision threshold.
Figure 5 illustrates the internal architecture of HERA for a given sector. Starting from the complete set of firms belonging to a specific macro-sector, observations are progressively classified through a sequence of binary Random Forest models. At each stage, a classifier separates firms into two increasingly homogeneous rating groups, which are subsequently analysed by additional classifiers until a unique rating category is assigned.
It is important to emphasize that the architecture shown in Figure 5 is independently estimated for each of the three macro-sectors identified through the ATECO classification, namely

Production, Trade, and Services. Consequently, HERA consists of three sector-specific rating engines, each characterized by the same hierarchical structure composed of nine interconnected Random Forest classifiers and nine optimized F1-based classification thresholds.
Because the explanatory variables are observed at time *t* while the target rating is measured at time *t+1*, HERA performs a genuine one-year-ahead forecasting exercise. Therefore, the framework should not be interpreted merely as a rating classification system, but rather as a predictive architecture designed to forecast firms' future creditworthiness.
Overall, the hierarchical decomposition of the rating process transforms a complex ten-class prediction problem into a sequence of simpler binary decisions, thereby enhancing classification stability, facilitating threshold calibration, and improving the reliability of rating forecasts in the presence of heterogeneous and unbalanced data.

5. RESULTS

The empirical results reported in this section are based on the Out-of-Bag (OOB) evaluation procedure provided by the Random Forest algorithm. For each tree, approximately one-third of the observations are excluded from the bootstrap sample used for estimation and subsequently employed for model validation. Consequently, performance indicators are computed using observations that are not involved in the construction of the corresponding decision trees, reducing the risk of overfitting and providing a reliable assessment of predictive performance.
All Random Forest models composing the HERA framework were implemented in MATLAB R2024a. The reported results therefore refer to the predictive performance obtained through OOB validation rather than to in-sample classifications.

5.1. Performance of the HERA classifiers

Since each sector-specific HERA engine consists of nine binary Random Forest classifiers, performance measures are evaluated for every classifier separately, together with the corresponding optimal classification threshold obtained through F1-score maximization.
Table 7 reports the performance of the Random Forest classifiers composing the HERA framework. For each binary classifier, the table presents the optimal classification threshold obtained through F1-score maximization, the corresponding maximum F1-score, and a set of performance measures. Specifically, AUC and Accuracy are computed using Out-of-Bag (OOB) predictions and therefore provide an assessment of the classifier's out-of-sample predictive ability. In contrast, the reported Error Rate refers to the final classification error obtained after applying the F1-optimized threshold and is calculated as the proportion of misclassified observations over the total number of observations. Consequently, the Error Rate reflects the operational performance of the HERA classification system once the decision threshold has been calibrated. Since each of the three sector-specific HERA engines consists of nine binary Random Forest classifiers, model performance is assessed separately for every classifier in order to identify potential differences in predictive power across rating segments and economic sectors.
Table 8 reports the Out-of-Bag (OOB) performance of the terminal classifiers included in the HERA framework. For each classifier, the table compares the F1-score obtained using the conventional classification threshold of 0.5 with the F1-score associated with the threshold selected through the F1-maximization procedure adopted by HERA. The table also reports the corresponding OOB accuracy and classification error.
The results show that, in most cases, the HERA calibration procedure produces higher OOB F1-scores than the conventional threshold, although the improvement is generally modest. This finding suggests that the probability estimates generated by the Random Forest classifiers already provide a satisfactory degree of class separation and that threshold optimization mainly acts as a refinement of the classification rule rather than as a primary source of predictive performance.

At the same time, considerable heterogeneity can be observed across rating transitions. Some classifiers exhibit relatively low F1-scores despite high levels of OOB accuracy and limited classification errors. This apparent discrepancy is largely attributable to the highly unbalanced nature of several binary classification tasks. In such cases, correctly identifying the minority rating category remains challenging, and F1-scores provide a more demanding measure of predictive performance than overall accuracy. The comparison between in-sample and OOB performance suggests that exceptionally high training F1-scores do not necessarily translate into equally strong out-of-sample performance. This result reinforces the importance of OOB validation, particularly for highly unbalanced rating transitions. From a methodological perspective, these results confirm the usefulness of threshold calibration while also suggesting the need for further investigation of alternative calibration criteria. In particular, future research could assess whether threshold selection based on the Youden Index, or other measures balancing sensitivity and specificity, may provide more stable performance across different levels of class imbalance.

Before discussing the results, it is important to distinguish between in-sample and Out-of-Bag (OOB) performance measures. In-sample performance is computed using the same observations employed to estimate the Random Forest classifiers and therefore reflects the ability of the model to fit the training data. Although useful for assessing the internal consistency of the classification framework, in-sample measures may provide optimistic estimates of predictive performance.

In contrast, OOB performance is evaluated using observations that are not included in the bootstrap sample used to grow each tree. Since these observations are not directly involved in the estimation of the corresponding trees, OOB measures provide an internal approximation of out-of-sample predictive performance and are generally considered a more reliable indicator of the model's generalization ability.

Consequently, the performance measures reported in Tables 7 and 8 serve different purposes. Table 7 summarizes the performance of the calibrated classifiers on the estimation sample, whereas Table 8 evaluates the same classifiers using OOB predictions. The comparison between the two sets of results provides useful information on the robustness of the HERA framework and on the extent to which the observed predictive performance generalizes beyond the training data.

It should be noted that the performance measures reported in Tables 7 and 8 are not directly and fully comparable. The results reported in Table 7 are based on the estimation sample and therefore reflect in-sample performance, whereas Table 8 relies on Out-of-Bag predictions and provides an assessment of the generalization ability of the classifiers. The comparison between the two tables is nevertheless informative because it allows the identification of potential differences between training and out-of-sample performance, particularly for highly unbalanced rating transitions.

Although threshold selection in the HERA framework is based on F1-score maximization, maximum Youden Index values and the associated optimal thresholds were also computed as a robustness analysis. These results are reported in Appendix B.

Figure 5. HERA Architecture: Sector-Specific Hierarchical Network of Binary Random Forest Classifiers for One-Year-Ahead Rating Forecasting

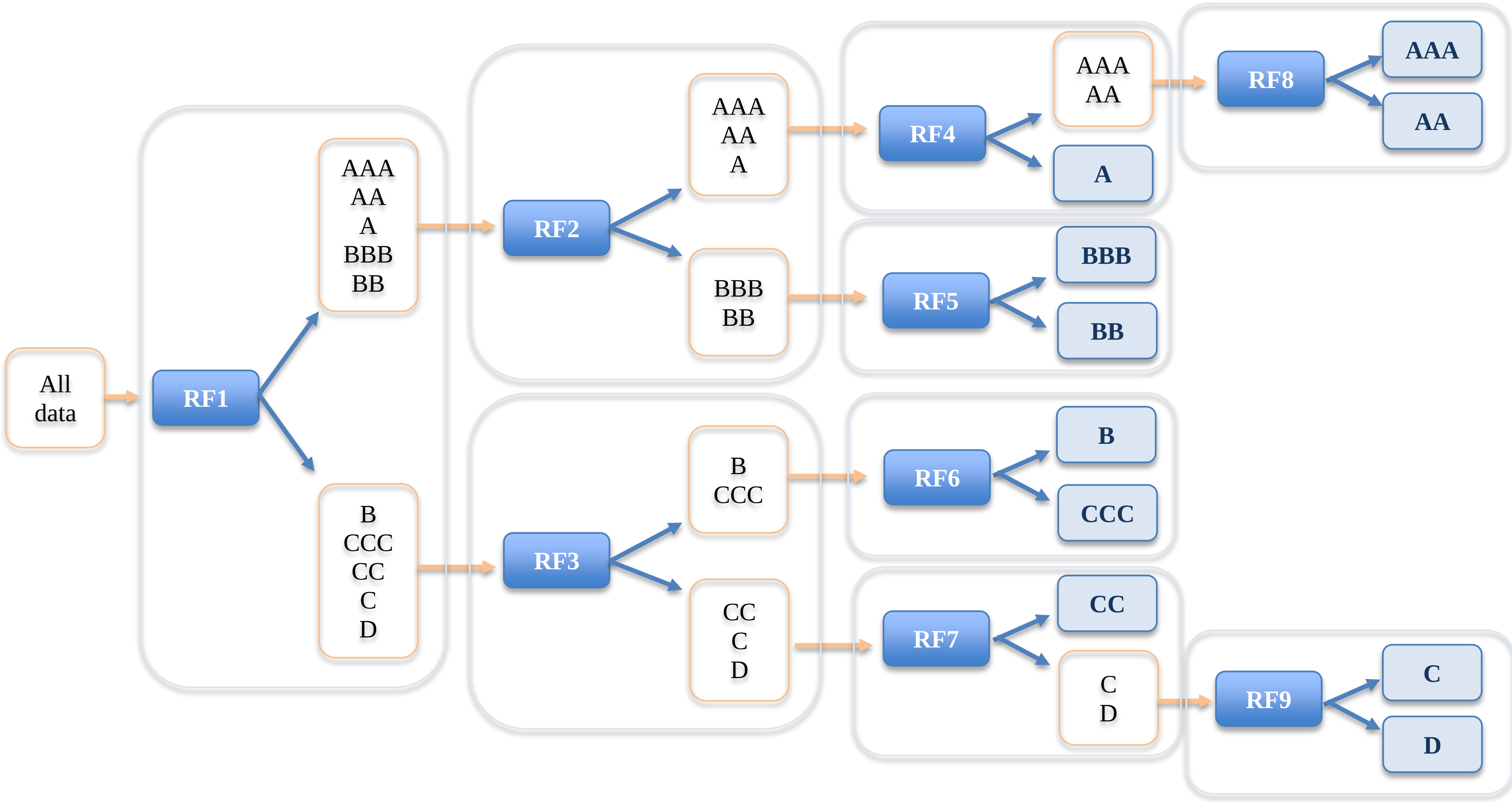


Table 7. Thresholds and Out-of-Bag Performance Measures of the HERA Random Forest Classifiers

	Production					Trade					Services				
RF	Optimal Threshold	F1 (Optimal Threshold)	AUC (OOB, 0.5)	Accuracy (OOB, 0.5)	Final Classification error (Optimal Threshold)	Optimal Threshold	F1 (Optimal Threshold)	AUC (OOB, 0.5)	Accuracy (OOB, 0.5)	Final Classification error (Optimal Threshold)	Optimal Threshold	F1 (Optimal Threshold)	AUC (OOB, 0.5)	Accuracy (OOB, 0.5)	Final Classification error (Optimal Threshold)
RF 1	0.45	0.9996	0.8704	0.8091	0.0474%	0.48	0.9998	0.8636	0.8014	0.0269%	0.47	0.9992	0.8411	0.7854	0.0992%
RF 2	0.49	0.9998	0.8769	0.8130	0.0159%	0.49	0.9998	0.8937	0.8311	0.0116%	0.48	0.9994	0.8769	0.8130	0.0159%
RF 3	0.45	0.9993	0.8143	0.8090	0.1122%	0.51	0.9997	0.7674	0.8039	0.0441%	0.47	0.9992	0.7883	0.7705	0.1140%
RF 4	0.41	1.0000	0.8251	0.8869	0.0000%	0.42	1.0000	0.8189	0.9051	0.0000%	0.41	1.0000	0.8321	0.9636	0.0000%
RF 5	0.49	0.9999	0.7614	0.7007	0.0069%	0.46	0.9999	0.7794	0.7170	0.0570%	0.46	0.9998	0.7251	0.6705	0.0213%
RF 6	0.47	0.9996	0.7593	0.7323	0.0476%	0.52	0.9998	0.7507	0.7430	0.0252%	0.47	0.9993	0.7527	0.7014	0.0793%
RF 7	0.57	0.9994	0.8109	0.9666	0.1159%	0.51	0.9996	0.8342	0.9544	0.0755%	0.58	0.9994	0.8096	0.9477	0.1164%
RF 8	0.31	1.0000	0.7088	0.9511	0.0000%	0.32	1.0000	0.7847	0.9643	0..0000%	0.31	1.0000	0.6987	0.8991	0.0000%
RF 9	0.51	0.9993	0.7190	0.7212	0.0953%	0.48	0.9994	0.6974	0.7051	0.0791%	0.54	0.9991	0.7196	0.7220	0.1208%

Notes: Optimal thresholds were selected through F1-score maximization. AUC and Accuracy are computed using Out-of-Bag predictions. Classification Error corresponds to the percentage of misclassified observations obtained after applying the optimal threshold.

Table 8. Out-of-Bag Performance Measures with standard threshold (i.e., 0.5) and the optimal HERA threshold

	Production				Trade				Services			
RF	**F1 (OOB, 0.5)**	**F1 (OOB, optimal threshold)**	**Accuracy (OOB, optimal threshold)**	**Final Classification error (OOB, optimal threshold)**	**F1 (OOB, 0.5)**	**F1 (OOB, optimal threshold)**	**Accuracy (OOB, optimal threshold)**	**Final Classification error (OOB, optimal threshold)**	**F1 (OOB, 0.5)**	**F1 (OOB, optimal threshold)**	**Accuracy (OOB, optimal threshold)**	**Final Classification error (OOB, optimal threshold)**
RF1	0.8530	0.8557	0.8081	19.19%	0.8516	0.8531	0.8044	19.56%	0.8347	0.8372	0.7851	21.49%
RF2	0.7249	0.7276	0.8130	18.70%	0.7438	0.7459	0.8309	16.91%	0.7345	0.7432	0.7787	22.13%
RF3	0.8776	0.8795	0.8073	19.27%	0.8805	0.8797	0.8033	19.67%	0.8491	0.8517	0.7709	22.91%
RF4	0.3756	0.4400	0.8824	11.76%	0.3375	0.3900	0.9018	9.82%	0.2734	0.3290	0.9628	37.20%
RF5	0.6790	0.6830	0.7002	29.98%	0.6835	0.6964	0.7157	28.43%	0.6458	0.6630	0.6670	33.30%
RF6	0.8091	0.8134	0.7328	26.72%	0.8251	0.8221	0.7414	25.86%	0.7660	0.7734	0.7022	29.78%
RF7	0.9711	0.9700	0.9423	5.77%	0.9765	0.9765	0.9543	4.57%	0.9730	0.9719	0.9458	5.42%
RF8	0.0777	0.2469	0.9432	5.68%	0.0634	0.1762	0.9589	4.11%	0.0861	0.2299	0.8769	12.31%
RF9	0.8120	0.8106	0.7205	27.95%	0.7951	0.7985	0.7058	29.42%	0.8141	0.8060	0.7179	28.21%

5.2. Sector-specific determinants of rating transitions

Variable importance is evaluated through the Out-of-Bag permutation importance provided by the Random Forest algorithm. For each predictor, importance is measured as the increase in classification error generated by randomly permuting the values of that variable while keeping all other predictors unchanged. Consequently, larger values indicate a stronger contribution of the corresponding financial indicator to the rating classification process. Since permutation importance measures variations in predictive performance rather than proportions of explained variance, the resulting values should be interpreted as relative importance scores and do not necessarily sum to 100.

From a managerial perspective, the analysis is not focused on identifying the most important financial indicator in general, but rather on understanding which variables are most relevant for a specific rating transition. Consequently, the proposed approach allows the investigation of the financial factors associated with improvements or deteriorations in firms' rating positions within each sector.

For example, a profitability indicator may play a crucial role in distinguishing firms belonging to the highest rating classes, while leverage-related variables may become more relevant when differentiating between financially distressed firms. Therefore, the importance scores provide insights not only into the determinants of rating assignments, but also into the financial dimensions on which firms should focus in order to improve their future rating position.

The interpretation of variable importance thus shifts from a purely predictive perspective to a managerial one, allowing the identification of the financial factors that are most closely associated with specific rating improvements within each economic sector.

Tables 9, 10, and 11 report the variable importance scores obtained from the Random Forest classifiers composing the HERA architecture for the Production, Trade, and Services sectors, respectively. As discussed in Section 4.4, the original ten-class rating scale is decomposed into a sequence of binary classification tasks. Consequently, each Random Forest is associated with a specific rating transition within the hierarchical structure, and its importance scores measure the contribution of each financial indicator to that particular classification decision.

More specifically, RF4 distinguishes firms classified in the highest rating segment (AAA-AA) from firms belonging to the A rating class. RF5 further separates firms rated BBB from those rated BB. RF6 distinguishes firms classified as B from those belonging to the CCC category, while RF7 separates firms in the CC segment from those assigned to the most critical rating classes (C-D). Finally, RF8 identifies firms belonging to the AAA class versus AA, whereas RF9 distinguishes between the C and D classes.

Therefore, the reported importance scores should not be interpreted as global measures of variable relevance. Instead, they indicate the contribution of each financial indicator to a specific rating transition. This perspective allows the identification of the financial factors that are most strongly associated with improvements or deteriorations in firms' rating positions and makes it possible to investigate whether such determinants differ across sectors and rating categories. In other words, from a practical perspective, these results provide insights into the financial dimensions that firms may need to strengthen in order to improve their future rating position within a given sector. In addition, variable importance identifies the predictors that contribute most to a given rating transition but does not provide information regarding the direction of their effect.

In order to obtain a preliminary indication of the direction associated with the most influential predictors, Tables 12-14 provide a preliminary indication of the direction associated with the most influential predictors identified by the Random Forest classifiers. For each terminal classifier, the tables report the mean and median values of the most important variable for the two rating categories involved in the corresponding classification task. Given the presence of skewed distributions and extreme observations, median values are considered more informative than arithmetic means.

Although these statistics do not allow causal interpretations, they offer useful insights into the financial characteristics associated with different rating outcomes. The results suggest that the same financial indicator may play different roles across rating transitions, reinforcing the hypothesis that rating determinants are both sector-specific and transition-specific. Consequently, HERA not only forecasts firms' future rating classes but also provides preliminary evidence on the financial dimensions associated with potential rating improvements or deteriorations.

Table 9: Variable Importance Scores of the HERA Random Forest Classifiers: Production Sector

Variable	RF4	RF5	RF6	RF7	RF8	RF9
DEBTA	7.8594	14.737	12.321	4.9718	2.6002	9.6682
DEBPROD	7.2092	12.747	7.3411	3.0343	2.811	5.3315
ROS	5.508	10.387	5.7197	2.5614	2.5375	4.1815
ROA	5.5538	6.6009	4.0892	4.0114	2.2369	4.2198
DEBCOV	5.373	8.008	5.0718	3.7872	2.3017	4.9007
FAPROD	7.9251	11.608	8.1384	2.8011	2.3261	4.9596
FCCOV	6.437	23.263	5.6616	1.3036	1.1589	3.6521
FAEQ	7.165	12.421	15.256	1.9698	2.1507	8.0336

Table 10: Variable Importance Scores of the HERA Random Forest Classifiers: Trade Sector

Variable	RF4	RF5	RF6	RF7	RF8	RF9
DEBTA	7.3293	15.004	8.7549	2.8192	1.7446	5.1761
DEBPROD	5.9585	10.532	8.6478	2.5815	2.2137	5.1508
ROS	5.5722	7.9256	5.6296	1.9795	1.9021	3.154
ROA	3.7605	6.1429	4.2761	2.283	1.5456	3.6351
DEBCOV	4.4319	7.5644	4.3797	1.976	1.9135	3.739
FAPROD	5.3482	9.4791	8.0334	2.4047	1.7514	3.9853
FCCOV	6.1589	20.763	8.2691	0.64018	0.89305	1.8211
FAEQ	4.8285	8.7095	12.603	1.1879	1.3323	4.6184

Table 11: Variable Importance Scores of the HERA Random Forest Classifiers: Services Sector

Variable	RF4	RF5	RF6	RF7	RF8	RF9
DEBTA	5.3485	14.473	12.804	5.0968	2.8422	9.5339
DEBPROD	4.5408	12.288	8.7167	3.2274	3.4271	6.705
ROS	4.9144	8.9264	7.5736	3.0075	2.5856	5.1314
ROA	3.8206	8.38	5.4044	4.5469	2.3796	5.499
DEBCOV	4.2573	8.353	5.0557	5.0034	2.6401	6.714
FAPROD	4.7845	12.361	8.2782	3.6299	2.6777	6.8909
FCCOV	3.1182	18.083	7.631	1.1655	0.99617	4.7212
FAEQ	3.919	18.885	25.573	2.2667	2.1245	9.423

Table 12 provides a preliminary indication of the direction associated with the most influential predictors identified by the HERA classifiers in the Production sector. Although variable importance scores do not provide information on the sign of the relationship, the comparison of mean and median values across rating categories offers initial insights into the financial characteristics associated with future rating outcomes.

The results highlight the recurrent role of debt-related indicators across several rating transitions. DEBTA emerges as the most important predictor in RF4, RF7, and RF9, while DEBPROD is

identified as the dominant variable in RF8. In particular, lower median values of DEBTA are generally associated with the higher rating categories in the lower segments of the rating scale (RF7 and RF9), suggesting the relevance of indebtedness in distinguishing financially distressed firms from firms characterized by a lower level of credit risk.
For intermediate rating transitions, FCCOV (RF5) and FAEQ (RF6) emerge as the most relevant predictors. Firms forecasted to belong to the higher rating categories (BBB and B, respectively) exhibit lower median values of FCCOV and higher median values of FAEQ than firms assigned to the lower rating categories. Overall, these findings suggest that financial sustainability and capital structure play a particularly important role in shaping future rating outcomes in the Production sector.

Table 12. Mean and Median Values of the Most Important Predictor Across Rating Categories: Production Sector

RF	Most Important Variable	Mean (Class 0)	Mean (Class 1)	Median (Class 0)	Median (Class 1)	Rating Transition
RF8	DEBPROD	11.012	50.159	0.325	0.270	AA vs AAA
RF4	FAPROD	5.559	7.002	0.121	0.160	A vs AAA-AA
RF5	FCCOV	4.464	2.475	0.148	0.053	BB vs BBB
RF6	FAEQ	15.816	17.842	0.659	1.101	CCC vs B
RF7	DEBTA	26.648	20.483	1.177	0.971	D-C vs CC
RF9	DEBTA	24.484	18.702	1.001	0.954	D vs C

For Trade sector, Table 13 shows several interesting patterns. DEBTA is identified as the most important predictor in RF4, RF7 and RF9, highlighting the relevance of debt-related indicators across multiple portions of the rating scale. In RF7 (D-C versus CC) and RF9 (D versus C), lower median values of DEBTA are associated with the higher rating classes, suggesting that indebtedness plays an important role in distinguishing firms located in the lower segments of the rating scale. By contrast, only limited differences are observed in RF4, where firms belonging to the A and AAA-AA categories exhibit similar median values of DEBTA.
FCCOV emerges as the most influential predictor in RF5. The median value of FCCOV is substantially lower for firms forecasted to belong to the BBB class (0.042) than for firms forecasted to remain in the BB category (0.142). This result suggests that the sustainability of financial charges relative to operating performance is strongly associated with future rating improvements within the intermediate segment of the rating scale.
In RF6, FAEQ represents the most important predictor. Firms forecasted to belong to the B class exhibit considerably higher median values than firms forecasted to remain in the CCC category (0.779 versus 0.400). This evidence highlights the importance of the relationship between fixed assets and permanent capital in distinguishing firms characterized by different levels of financial risk.
Finally, DEBPROD emerges as the dominant predictor in RF8, separating AAA-rated firms from AA-rated firms. The lower median value observed for AAA firms suggests that debt-related indicators remain relevant even within the highest rating categories.

Table 13. Mean and Median Values of the Most Important Predictor Across Rating Categories: Trade Sector

RF	Most Important Variable	Mean (Class 0)	Mean (Class 1)	Median (Class 0)	Median (Class 1)	Rating Transition
RF8	DEBPROD	7.465	6.931	0.216	0.179	AA vs AAA
RF4	DEBTA	1.635	4.173	0.472	0.448	A vs AAA-AA
RF5	FCCOV	4.902	1.966	0.142	0.042	BB vs BBB
RF6	FAEQ	6.258	8.822	0.400	0.779	CCC vs B
RF7	DEBTA	16.585	5.996	1.890	0.963	D-C vs CC
RF9	DEBTA	9.311	4.317	1.039	0.941	D vs C

Table 14 presents results for Services sector. DEBTA appears as the most relevant predictor in RF4, RF7 and RF9. In RF4, firms belonging to the highest rating segment (AAA-AA) exhibit slightly higher median values of DEBTA than firms rated A (0.450 versus 0.393). Conversely, in RF7 and RF9, lower median values of DEBTA are associated with the higher rating categories (CC and C, respectively), suggesting that the role of indebtedness may vary across different portions of the rating scale.
A similar pattern emerges for FAEQ, which is identified as the most influential predictor in both RF5 and RF6. However, while BBB firms exhibit lower median values than BB firms in RF5, the opposite result is observed in RF6, where firms rated B display higher median values than firms rated CCC. This finding highlights that the financial determinants of rating transitions are not constant across adjacent categories and may change according to the level of credit risk considered.
Finally, also for Services sector DEBPROD emerges as the most relevant predictor in RF8, distinguishing firms rated AAA from those rated AA. The lower median value observed for AAA firms suggests that debt-related indicators continue to play an important role even within the highest rating classes.

Table 14. Mean and Median Values of the Most Important Predictor Across Rating Categories: Services Sector

RF	Most Important Variable	Mean (Class 0)	Mean (Class 1)	Median (Class 0)	Median (Class 1)	Rating Transition
RF8	DEBPROD	87.217	23.589	0.304	0.254	AA vs AAA
RF4	DEBTA	1.192	4.605	0.393	0.450	A vs AAA-AA
RF5	FAEQ	24.348	18.095	1.221	0.892	BB vs BBB
RF6	FAEQ	31.937	26.551	1.045	1.333	CCC vs B
RF7	DEBTA	18.081	10.368	1.135	0.969	D-C vs CC
RF9	DEBTA	10.948	10.114	1.002	0.952	D vs C

Unlike traditional classification exercises, HERA performs a one-year-ahead forecasting task. Consequently, the observed differences in financial indicators should be interpreted as characteristics associated with firms' future rating outcomes rather than their contemporaneous rating positions. The results suggest that the determinants of rating transitions are not constant throughout the rating scale, as the same financial indicator may play different roles depending on the risk segment considered. While intermediate rating transitions are primarily associated with measures of financial sustainability and capital structure, debt-related indicators repeatedly emerge as key predictors in classifications involving both the highest and lowest rating categories. This evidence highlights the central role of indebtedness in shaping firms' future creditworthiness and further supports the adoption of a sector-specific hierarchical framework for corporate rating forecasting.

To provide a preliminary illustration of the relationship between financial indicators and forecasted rating outcomes, Figures 6 and 7 show the Partial Dependence Plot (PDP) for two RF of Production sector. Figure 6 reports the PDP associated with DEBPROD, the most important predictor identified for RF8. PDP measures the average effect of a predictor on the probability estimated by the Random Forest while accounting for the joint distribution of the remaining variables. The curve suggests a positive but non-linear relationship between DEBPROD and the probability of belonging to the positive rating class. In particular, the predicted probability increases rapidly for low values of the indicator and then stabilizes, indicating diminishing marginal effects for higher levels of DEBPROD. Notice that since HERA performs a one-year-ahead forecasting exercise, the observed pattern should be interpreted as the relationship between DEBPROD measured at time *t* and the probability of obtaining a specific rating outcome at time *t+1*.

Figure 6: Example of a Partial Dependence Plot for RF8 – Production Sector

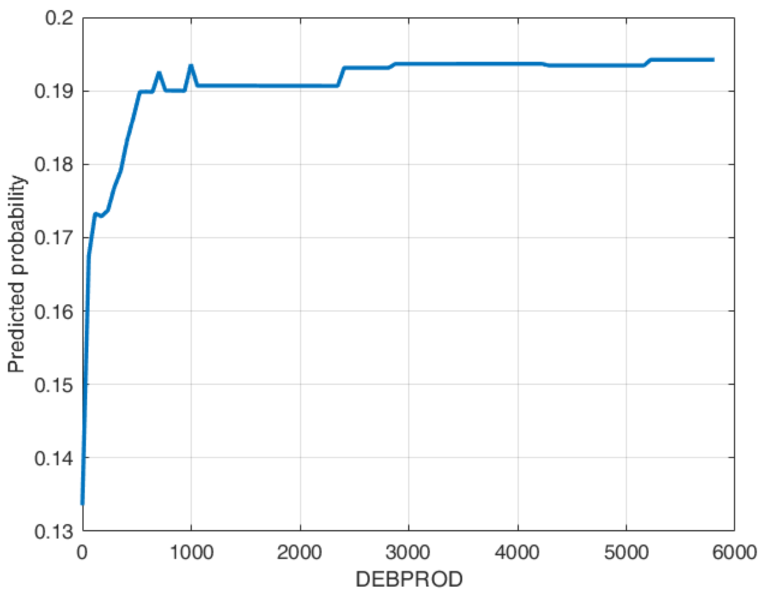
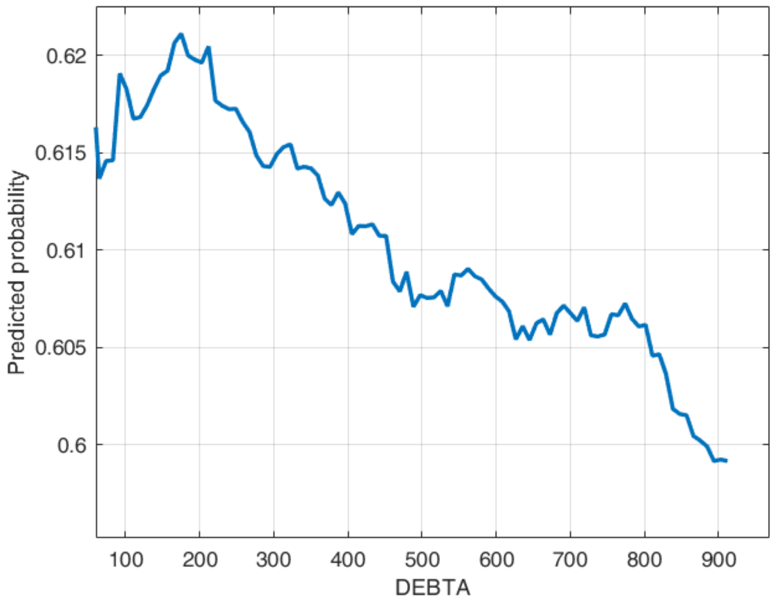


Figure 7 reports the Partial Dependence Plot associated with DEBTA, the most relevant predictor identified by RF9 in the Production sector. The results suggest a negative relationship between DEBTA and the probability of belonging to the C rating class. As DEBTA increases, the predicted probability gradually declines, indicating that higher levels of indebtedness are associated with a greater likelihood of being assigned to the D category. The figure also highlights the non-linear nature of the relationship, supporting the need for more advanced interpretability tools when investigating the economic mechanisms underlying rating transitions.

Figure 7: Example of a Partial Dependence Plot for RF9 – Production Sector



Both presented figures highlight that simple comparisons based on mean and median values should be interpreted with caution, as the relationship captured by the classifier may be non-linear and may depend on interactions with other financial indicators. For this reason, the results presented in Tables 12-14 should be regarded as preliminary evidence of the direction of the association rather than as a complete characterization of the effect exerted by each predictor.

5.3. Discussion

Tables 9, 10, and 11 report the variable importance scores obtained from the Random Forest classifiers composing the HERA framework for the Production, Trade, and Services sectors, respectively. Since each classifier corresponds to a specific rating transition within the hierarchical architecture, the reported scores identify the financial indicators that contribute most to distinguishing between adjacent rating categories. Consequently, the analysis provides insights into the financial dimensions that are most strongly associated with firms' rating movements.

Several interesting patterns emerge from the results. First, the importance of financial indicators varies considerably across the different classifiers, suggesting that the determinants of corporate ratings are not constant throughout the rating scale. Variables that are highly relevant for distinguishing firms belonging to intermediate rating classes are not necessarily the same as those driving classifications at the upper or lower ends of the rating spectrum.

Second, important differences emerge across sectors. In all three sectors, debt-related indicators and measures of financial sustainability appear among the most relevant predictors. However, the relative contribution of individual variables changes according to the economic environment in which firms operate, supporting the hypothesis that sectoral heterogeneity plays a significant role in the rating assignment process.

For firms operating in the Production sector, the results highlight the importance of debt sustainability and capital structure indicators. In particular, RF5 and RF6 assign substantial importance to variables such as FCCOV, DEBTA, DEBPROD, and FAEQ, suggesting that the relationship between indebtedness, financial charges, and long-term financing is particularly relevant for distinguishing firms belonging to adjacent rating classes.

A similar pattern emerges in the Trade sector, where FCCOV, DEBTA, and FAEQ again appear among the most influential predictors. The concentration of importance in a relatively small

number of variables suggests that financial sustainability and debt management play a central role in determining firms' creditworthiness within commercial activities.
The Services sector exhibits partially different dynamics. Although financial sustainability indicators remain important, capital structure variables assume a particularly prominent role. Specifically, FAEQ records the highest importance values within RF6, indicating that the relationship between fixed assets and permanent capital is especially informative for some rating transitions among service firms. Moreover, profitability indicators maintain a more stable contribution across the hierarchy than in the other sectors.
Overall, the evidence confirms that rating determinants are both sector-specific and transition-specific. The financial indicators that contribute most to a given rating transition differ according to the sector under consideration and the portion of the rating scale being analysed. These findings provide support for the sector-specific hierarchical architecture adopted by HERA and suggest that a single rating model may fail to capture important industry-specific financial characteristics. Consequently, the empirical evidence provides a positive answer to both research questions, highlighting the relevance of sector-specific forecasting architectures and confirming the existence of heterogeneous financial drivers across sectors and rating transitions.
Beyond their methodological relevance, these results have important practical implications. From a managerial perspective, the analysis identifies the financial dimensions that are most closely associated with specific rating transitions, thereby providing useful information for firms seeking to improve their future creditworthiness. From a policy perspective, the results may help public institutions and financial regulators identify the financial areas that deserve greater attention within different sectors of the economy. The observed heterogeneity across sectors suggests that support measures aimed at strengthening firms' financial resilience may need to be tailored to the specific characteristics of the industries involved rather than relying on a uniform policy framework.

6. Conclusion

The results provide strong evidence that the determinants of corporate ratings are both sector-specific and transition-specific. Financial indicators do not contribute uniformly to the rating assignment process; instead, their relative importance varies according to both the economic sector and the rating categories being compared.
The empirical findings support the adoption of a hierarchical sector-specific architecture rather than a single global rating model. By decomposing the original ten-class rating problem into a sequence of binary classification tasks, HERA is able to capture heterogeneous financial patterns that would otherwise remain hidden in a traditional multiclass framework.
Beyond its predictive capabilities, the proposed framework also contributes to the interpretability of rating systems by identifying the financial indicators that are most relevant for specific rating transitions. This feature may provide useful support to managers, analysts, and financial institutions interested in understanding the drivers of firms' future creditworthiness.
The empirical findings provide answers to the two research questions that motivated this study. Concerning RQ1, the results show that the sector-specific hierarchical architecture adopted by HERA is capable of producing satisfactory rating forecasts through a sequence of binary Random Forest classifiers calibrated by means of F1-score optimization. The obtained performance measures support the suitability of the proposed framework for one-year-ahead corporate rating forecasting. With regard to RQ2, the analysis of variable importance reveals that the financial determinants of rating transitions vary considerably across both economic sectors and rating categories. This evidence confirms that the contribution of accounting indicators to the rating assignment process is neither uniform across industries nor constant along the rating scale.
The preliminary PDP analyses further suggest that the relationship between financial indicators and future rating outcomes may be non-linear and influenced by interactions among predictors. While the variable importance measures adopted in this study successfully identify the main

drivers of rating transitions, they do not provide a complete characterization of the direction and magnitude of their effects.

Future developments of the HERA framework could therefore integrate explainable artificial intelligence techniques, such as SHAP values, capable of decomposing individual predictions and quantifying the contribution of each financial indicator to the forecasted rating outcome. Such an extension would allow a more detailed interpretation of the economic mechanisms underlying rating transitions and would enhance the usefulness of HERA as a managerial and policy-support tool.

A further research avenue concerns the computational optimization of the framework. Given the size of the dataset used in this study (more than 4 million observations), the estimation of 200 Random Forest classifiers at each hierarchical node entails substantial computational costs. More efficient implementations would be necessary to facilitate the deployment of HERA as a forecasting and decision-support platform capable of performing scenario analyses in operational settings.

More generally, although HERA has been developed and validated in the context of corporate rating forecasting, its methodological architecture is not inherently restricted to this specific application. The combination of a hierarchical structure, sector-specific modelling, and transition-oriented classification can potentially be extended to other forecasting problems characterized by heterogeneous populations and ordered state transitions. Examples include bankruptcy prediction, financial distress forecasting, credit risk migration analysis, ESG risk assessment, and other domains where prediction accuracy may benefit from decomposing complex multiclass problems into a sequence of specialized binary decisions.

The main contribution of HERA therefore lies not only in its predictive performance but also in its flexibility, scalability, and interpretability. By combining machine learning techniques with economic and sectoral information, the framework provides a structured approach capable of capturing heterogeneity across groups while maintaining transparency regarding the determinants of predictions. These characteristics make HERA particularly attractive for applications requiring both forecasting accuracy and explainability.

Looking ahead, a possible evolution of the framework could pursue several complementary objectives. First, the current Random Forest architecture could be enriched through the integration of advanced machine learning and deep learning models, allowing the framework to exploit both structured accounting data and additional unstructured information sources. Second, dynamic learning mechanisms could be introduced to adapt the models more rapidly to changing economic conditions and evolving risk profiles. Third, enhanced explainability modules based on SHAP values, counterfactual explanations, and scenario analysis could provide more actionable insights for decision makers. Finally, future versions of HERA could explore hybrid architectures combining hierarchical classification with ensemble and probabilistic forecasting approaches, transforming the framework into a more general decision-support platform for risk assessment and forecasting applications beyond corporate ratings.

7. References

Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), pp. 589-609.

Atiya, A.F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks*, 12(4), pp. 929-935.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), pp. 5-32.

Calabrese, G. G., Falavigna, G., & Ippoliti, R. (2024). Financial constraints prediction to lead socio-economic development: An application of neural networks to the Italian market. *Socio-Economic Planning Sciences*, 95, 101973.

Falavigna, G. (2012). Financial ratings with scarce information: A neural network approach. *Expert Systems with Applications*, 39(2), pp. 1784-1792.

Falavigna, G., Costantino, G., Furlan, R., Quinn, J. V., Ungar, A., & Ippoliti, R. (2019). Artificial neural networks and risk stratification in emergency departments. *Internal and Emergency Medicine*, 14(2), pp. 291-299.
Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), pp. 861-874.
Grice, J.S., & Ingram, R.W. (2001). Tests of the generalizability of Altman's bankruptcy prediction model. *Journal of Business Research*, 54(1), pp. 53-61.
Habibzadeh, F. (2025). Diagnostic tests performance indices: an overview. *Biochemia Medica*, 35(1), pp. 5-12.
Lessmann, S., Baesens, B., Seow, H.V., & Thomas, L.C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), pp. 124-136.
Ohlson, J.A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), pp. 109-131.
Platt, H.D., & Platt, M.B. (1991). A note on the use of industry-relative ratios in bankruptcy prediction. *Journal of Banking & Finance*, 15(6), pp. 1183-1194.
Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management,* 45(4), pp. 427-437.
Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32-35.
Zhu, Y., Zhou, L., Xie, C., Wang, G.J., & Nguyen, T.V. (2019). Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. *International journal of production economics*, 211, pp. 22-33.

8. Appendix A: Summary Results of the Artificial Neural Network Experiments

The initial stages of the research involved the estimation of several Artificial Neural Network (ANN) models aimed at reproducing firms' rating classifications. These experiments were conducted as preliminary benchmarks prior to the development of the final HERA architecture. Owing to computational constraints and the time-consuming nature of the estimation process, the comparison was limited to the Trade sector.
The ANN models were implemented using MATLAB's patternnet architecture. Unlike the more general feedforwardnet, which can be employed for both regression and classification tasks, patternnet is specifically designed for pattern recognition and classification problems. In the present application, the network consists of a two-layer architecture with 10 neurons in the hidden layer, a hyperbolic tangent sigmoid activation function (tansig) between the input and hidden layers, and a logistic sigmoid activation function (logsig) in the output layer, producing probabilities for the target class. Network training was performed using the Scaled Conjugate Gradient algorithm (trainscg), which is particularly suitable for large datasets because of its lower memory requirements compared with alternative optimization procedures.
To ensure comparability between ANN and Random Forest (RF) models, the available observations were partitioned into training (64%), validation (16%), and test (20%) subsets. Performance evaluation was conducted exclusively on the test set for both model families. The RF architecture adopted in this comparison replicates the specification employed throughout the paper and consists of 200 decision trees. Model performance was evaluated in terms of Accuracy, Classification Error, and Area Under the ROC Curve (AUC). The results are reported in Table A.1.

Table A1. Comparison of AUC results between RF (200) and ANN ([10]) – Trade sector

	AUC RF (200)	**AUC ANN ([10])**	**Winner**
Node1	0.8649	0.7296	RF
Node2	0.8969	0.7632	RF

Node3	0.7723	0.6486	RF
Node4	0.8197	0.5385	RF
Node5	0.7826	0.5291	RF
Node6	0.7474	0.5944	RF
Node7	0.8435	0.6868	RF
Node8	0.7765	0.6076	RF
Node9	0.7029	0.6084	RF

9. Appendix B: Youden Index

In Table B2 are presented performance of RF. Notice that, the reported thresholds correspond to the values maximizing the Youden Index on the training sample. The Youden Index is computed as the sum of Sensitivity and Specificity minus one. Results are reported as a robustness check and are not used in the calibration of the final HERA classifiers, which rely on F1-score maximization.

Table B2. Training-Sample Maximum Youden Index Values and Corresponding Optimal Thresholds

	Production	**Trade**	**Services**
RF1			
Youden index	0.9991	0.9995	0.9980
Threshold	0.45	0.49	0.49
RF2			
Youden index	0.997	0.998	0.999
Threshold	0.46	0.46	0.47
RF3			
Youden index	0.9973	0.9990	0.9971
Threshold	0.49	0.56	0.49
RF4			
Youden index	1	1	1
Threshold	0.41	0.41	0.41
RF5			
Youden index	0.9999	0.9999	0.9996
Threshold	0.49	0.46	0.46
RF6			
Youden index	0.9989	0.9995	0.9983
Threshold	0.5	0.55	0.48
RF7			
Youden index	0.9938	0.9964	0.9925
Threshold	0.71	0.73	0.74
RF8			
Youden index	1.0000	1.0000	1.0000
Threshold	0.31	0.32	0.31
RF9			
Youden index	0.9985	0.9985	0.9976
Threshold	0.52	0.52	0.56

ISSN (online) 2421-7158

CNR-IRCrES Working Papers

2026

N.3/2026 Il detective sociale: laboratorio di divulgazione delle scienze sociali per la scuola primaria. Antonella Emina, Anna Perin, Lisa Sella, Isabella Maria Zoppi.

N. 2/2026 Il calcio come infrastruttura sociale? Riflessioni ed esempi dal contesto italiano. Chiara D'Amato, Luca Tricarico.

N. 1/2026 Regolamentazione ambientale europea e impatti economici sulle imprese. Greta Falavigna, Giampaolo Vitali.

2025

N. 9/2025 Finanziamento delle comunità energetiche: attori, strumenti e aspettative dell'ecosistema ESG nel contesto italiano. Maksym Koltunov. Luca Tricarico, Adriano Bisello, Chiara D'Amato.

N. 8/2025 Importazioni, esportazioni e strategie green: il caso di studio della chimica in Italia. Greta Falavigna, Giampaolo Vitali.

N. 7/2025 Mapping international collaboration in Nanosciences and Nanotechnologies: a bibliometric and econometric study of South African research. Ugo Finardi, Jeanne C.M. Vallette d'Osia.

N. 6/2025 Dispersione salariale e performance nel calcio in Italia: un'analisi degli incentivi tra sforzo e risultato nella serie A (2010-2022). Mario Nosvelli, Matteo Colzato.

N. 5/2025 Designing a multiple case study approach for studying universities' community engagement. Andrea Orazio Spinello, Valentina Carazzolo, Ugo Finardi, Stefania Frongia, Valentina Ghibellini, Emanuela Reale, Andrea Vargiu.

N. 4/2025 Tools and Methods for Assessing Active Ageing and Inclusion: The Rima Experimental Project. Lisa Sella, Nicola Pollo, Valentina Lamonica.

N. 3/2025 Sinfonie urbane: il ruolo della ritmanalisi nella rigenerazione degli spazi. Il caso di Piazza San Venceslao a Praga. Vittoria Virgili, Marco Marucci.

N. 2/2025 From Single to Multi-Project Participation: SME Absorptive Capacity in Horizon 2020. Loreta Isaraj.

N. 1/2025 Invecchiamento e welfare responsabile: la "condizione anziana" nelle periferie di Roma Capitale. Cristiana Di Pietro.

2024

N. 5/2024 National frameworks for universities' community engagement: perspectives from Italy, France and the UK. Valentina Carazzolo, Ugo Finardi, Emanuela Reale, Andrea Orazio Spinello.

N. 4/2024 DAPIL – Digital archive of pilgrimage literature. Un archivio digitale per la letteratura medioevale di pellegrinaggio. Giuseppe Cuscunà.

N. 3/2024 Thesaurus e glossario ragionato della lingua della IA e della robotica. Thesaurus and reasoned glossary of AI and robotic language. Grazia Biorci.

N. 2/2024 Comprendere le trasformazioni di San Lorenzo, Roma. Un tentativo di ordine concettuale tra gentrificazione e politica urbana. Antonia Astore, Luca Tricarico.

N. 1/2024 How to measure the social acceptability of alternative environmental management solutions in wetlands and other ecosystems. Lisa Sella, Francesca Silvia Rota, Nicola Pollo.

2023

N. 5/2023 Laboratori creativi sperimentali a distanza per la valutazione. Isabella Maria Zoppi, Antonella Emina.

N. 4/2023 Eco-innovations and labor in the European automotive industry: an econometric study. Anna Novaresio.

N. 3/2023 Climate change adaptation planning: tools and methods for effective and sustainable decisions. Filippo Fraschini, Marco Pregnolato.

N. 2/2023 The effect of propensity to savings on rate of profit. Mario De Marchi.

N. 1/2023 UNESCO o non UNESCO? Quando il brand incentiva il turismo: il caso dei paesaggi vitivinicoli delle Langhe e del Monferrato. Orsola Borsani, Greta Falavigna.

2022

N. 5/2022 Logit statico, Logit dinamico e modelli hazard. Franco Varetto.

N. 4/2022 Evolution of Deep Learning from Turing machine to Deep Learning next generation. Greta Falavigna.

N. 3/2022 A Simulation Model of Technology Innovation of a Territory. Angelo Bonomi.

N. 2/2022 Technology and Environmental Policies. Angelo Bonomi.

Numeri precedenti/Previous issues

This study develops HERA (Hierarchical Engine for Rating Assessment), a sector-specific artificial intelligence framework designed to forecast corporate credit ratings one year ahead. The methodology decomposes the original multi-class rating problem into a hierarchical sequence of binary classification tasks estimated through Random Forest models. Separate architectures are developed for the Production, Trade, and Services sectors, enabling the framework to account for sector-specific financial characteristics and heterogeneous risk profiles.

The analysis evaluates the predictive performance of the proposed models and identifies the financial indicators most strongly associated with future rating transitions. Results show that the determinants of corporate ratings vary across both economic sectors and rating categories. Debt-related indicators emerge as the most important predictors for transitions involving the highest and lowest rating classes, while measures of financial sustainability and capital structure are more relevant for intermediate rating categories.

Beyond providing one-year-ahead rating forecasts, HERA offers insights into the financial drivers underlying future creditworthiness. The findings contribute to the literature on rating prediction and may be of interest to firms, financial institutions, and policymakers seeking a deeper understanding of the factors shaping corporate rating dynamics.

CNR - Consiglio Nazionale delle Ricerche

IRCrES - Istituto di Ricerca sulla Crescita Economica Sostenibile