Explainable Predictive Analysis of the Economic and Financial Performance of Italian Publicly Owned Companies

Idio Guarino*, Antonio Montieri†, Domenico Ciuonzo†, Antonio Pescapè†
*University of Verona, †University of Napoli Federico II
idio.guarino@univr.it, {domenico.ciuonzo, antonio.montieri, pescape}@unina.it

***Abstract*—This study examines the prediction of key economic and financial indicators for publicly owned Italian companies using historical time-series data. Four machine learning regression models—Linear Regression, Decision Tree, Random Forest, and XGBoost—are implemented with a sliding window approach to uncover patterns while addressing challenges like missing data and optimal window size. Performance is analyzed across groups defined by company characteristics (e.g., location, size, sector). An innovative eXplainable AI (XAI) methodology is introduced to interpret the prediction results, also aiding the design of simpler, more effective predictors. Results from 529 companies highlight the value of XAI in boosting prediction accuracy and streamlining the forecasting models.**

***Index Terms*—Forecasting, Financial Time Series Analysis, Financial Performance, Machine Learning, Explainable AI**

I. Introduction

In recent years, forecasting the financial performance of companies has become crucial for several stakeholders. *Investors* and *analysts* use these forecasts to identify opportunities, optimize portfolios, and manage risk. *Managers* rely on them for strategic planning, budgeting, and performance evaluation, while *government policymakers* use economic forecasts to guide policy for growth and stability. Accurate prediction trends are particularly valuable for stakeholders of *publicly owned companies* across different sectors, given their unique role in balancing public interest with financial sustainability. Indeed, effective forecasts provide company decision-makers with the insights needed to optimize financial management, resource allocation, and investments, all while ensuring accountability to the public. In this context, *Artificial Intelligence (AI)* and, particularly, *Machine Learning (ML)* [1, 2, 3] have found a booming field in which cutting-edge techniques are used to predict relevant economic and financial indicators associated with companies.

Accordingly, the *main contributions* of this work are as follows. Our study addresses these needs by developing and evaluating predictive models to forecast financial trends of Italian publicly owned companies using historical time-series data. We employ various regression algorithms based on ML and related data preparation techniques to create the most accurate model possible, also examining the role of categorical attributes in prediction accuracy. Secondly, we adopt a state-of-the-art methodology for the interpretability [4] of the forecasting results, highlighting advantageous means for analysts to investigate the performance of such companies powered by *eXplainable AI (XAI)*. These insights are valuable for investors, analysts, and companies, supporting better decision-making and financial strategies.

The rest of the paper is organized as follows: Sec. II reviews the works that have utilized ML techniques for predicting economic and financial indicators of companies and position our work against them; Sec. III describes the forecasting and explainability methodology employed in this study; Sec. IV outlines the experimental setup, including the dataset, the pre-preprocessing operations, and the evaluation procedure; Sec. V focuses on the findings emerging from the experimental evaluation conducted; finally, Sec. VI summarizes the takeaways and suggests future directions for the work.

II. Related Work

In the latest research developments, ML has garnered significant interest, with studies exploring its applications across various fields, including economic and financial prediction. Our work aligns with this trend, and the present section provides an overview of related research and outlines the positioning of this manuscript against the latter.

Numerous studies have focused on predicting various *economic and financial indicators and metrics* by using the values of other indicators as input variables for ML models. For instance, *Return On Equity (ROE)* and *Return On Assets (ROA)* are critical for assisting managers in making strategic choices and for investors in assessing a company's profitability. Their prediction based on the values of other indicators is the object of different studies leveraging ML-based approaches, such as Artificial Neural Networks (ANNs), Linear Regression (LIR), Support Vector Regression (SVR) [5] and Random Forest Regressor (RFR) [6]. Additionally, ANNs, LR, and SVR are applied to forecast *capital structure* [7], represented as the

This work was partially conducted within the course "Tecnologie ed Applicazioni per la Trasformazione Digitale" at the University of Napoli Federico II. We thank Iniziativa Cube S.r.l. for providing the dataset. This work is partially supported by the "Centro Nazionale HPC, Big Data e Quantum Computing – ICSC" and the "RESTART" Project funded by MUR, and the Marie Skłodowska-Curie Actions under the European Union's Horizon Europe research and innovation program for the Industrial Doctoral Network on Digital Finance (acronym: DIGITAL, project no. 101119635).

ratio of total debt to total equity, using independent variables like profitability, liquidity, solvency, and turnover ratios. In another work [8], *total revenue* prediction leverages LIR, k-Nearest Neighbors, SVR, and Decision Trees, with the latter showing the least prediction error and the highest R^2 value.

A wealth of studies have concentrated on *predicting stock market trends*. For instance, the work in [9] examines KOSPI market data, the principal stock market index of South Korea, employing Principal Component Analysis (PCA) alongside Deep Neural Networks—specifically Autoencoders and Restricted Boltzmann Machines—to provide insights into the application of ML for stock market analysis and forecasting. Similarly, the study in [10] compares RFR and Logistic Regression in predicting the financial performance of publicly listed Thai companies, highlighting the superiority of ML techniques over traditional statistical methods. The research in [11] aims to forecast the global Halal tourism stock index through text analysis and deep-learning methodologies, employing SHapley Additive exPlanations (SHAP) to interpret the models. More recent contributions [12, 13] focus on developing and analyzing stock market traffic predictors based on large language models that have been fine-tuned for forecasting stock performance, with the added goal of ensuring human-level interpretability in their outputs.

Earnings prediction plays a pivotal role for contemporary retail corporations, directly impacting their strategic decision-making processes. A number of studies have concentrated on sales forecasting, including [14, 15, 16]. For example, Pundir et al. [14] employs RFR and Vector Auto Regression (VAR) to project future revenues from retail sales data, revealing that VAR outperforms RFR, ARIMA, and other conventional methods. Dairu and Shilong [15] demonstrate the efficacy of XGBoost in managing large datasets, utilizing Walmart's sales figures to achieve superior forecasting accuracy. Also, Gurnani et al. [16] investigate various ML models for drugstore sales forecasting, concluding that the Hybrid ARIMA-ARNN model yields the most favorable results in terms of MAE and RMSE.

Forecasting *financial distress* using historical data represents another domain addressed via ML, as illustrated in [17]. Similarly, the study in [18] engages in financial time-series forecasting, estimating the growth rate of *free cash flow* through nine ML models and ARIMA. These models exploit a dataset consisting of past lags, the mean and standard deviation of the target variable, as well as relevant financial ratios.

Positioning. Our work focuses on using ML for predictive analysis in the economic-financial sector, based on time series and using (possibly) lightweight methods. Also, our second contribution is to empower predictive analysis with a recent XAI technique well suited for time series forecasting as opposed to other alternatives, e.g., SHAP [19]. Building on existing research demonstrating effective algorithms, we aim to gain more insight into Italian publicly owned companies.

III. Methodology

This section describes the prediction methodology proposed in this work. Specifically, in Sec. III-A, we formulate the prediction task and associated solution. Then, in Sec. III-B we describe our XAI methodology to interpret (and improve) the considered "black-box" financial predictors based on ML.

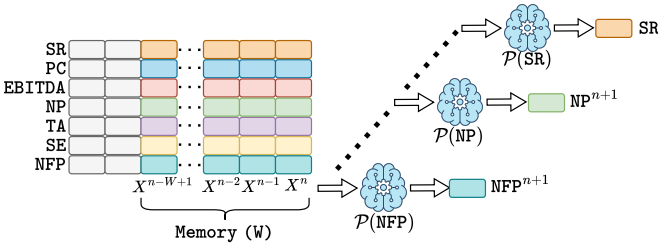


Fig. 1. Input-output construction: x^n is the n-th vector element of the multivariate time series of indicators extracted from the current company; W is the size of the memory window (*input*); $n+1$ is the index of the value to be predicted (*output*). Our methodology results in a dedicated model for each predicted indicator (i.e. P models in total).

A. Prediction Methodology

Problem Statement & Proposed Solution. We aim to predict P financial indicators for a company based on the history of the same indicators. Consider a multivariate time series $\boldsymbol{X} = \{\boldsymbol{x}^n\}_{n=1}^N$, where $\boldsymbol{x}^n = \{x_p^n\}_{p=1}^P$ represents the values of each p financial indicator at time instant n.

Our goal is to predict the value of the p-th indicator at time $n+1$, denoted as $\hat{x}_p^{n+1}$. Hence, we apply a predictor function $\mathcal{M}(\cdot)$ to a sequence of past observations collected over a memory window of size W as depicted in Fig. 1: $\hat{x}_p^{n+1} = \mathcal{M}_p\left(\boldsymbol{x}^{n-W+1}, \ldots, \boldsymbol{x}^{n-1}, \boldsymbol{x}^n\right)$. We use an *incremental sliding window* with unit stride to build the model's input. If the observations are fewer than the window size W ($n < W$), left-padding fills the window to ensure consistent input length for the predictor, enabling predictions from the first observed value. Once $n \geq W$, the window slides forward with each time step, always using the most recent W observations for continuous updates. Notably, we design a separate predictor for each financial indicator, following a single-task approach [20] with P different predictors $\mathcal{M}_1(\cdot), \ldots, \mathcal{M}_P(\cdot)$.

ML Models Considered. We apply *four supervised ML models* for financial prediction: (i) Linear Regression (LIR): a simple algorithm assuming a linear relationship between independent and dependent variables; (ii) Decision Tree (DT): a model based on a tree-like structure, where each node represents a decision based on an attribute, and leaves represent outcomes; (iii) Random Forest Regressor (RFR): an ensemble method combining multiple decision trees to improve prediction accuracy; (iv) XGBoost (XGB): a gradient boosting algorithm that sequentially adds decision trees, correcting previous errors, known for its efficiency and high performance.

B. Interpreting Predicted Financial Indicators via SAGE

This section details the methodology for exploring interpretability in regression, using *SAGE* [4], a model-agnostic approach based on Shapley values. SAGE measures (global) feature importance by comparing the ML model accuracy with

TABLE I
DESCRIPTION AND FORMULATION OF ECONOMIC AND FINANCIAL INDICATORS COMPOSING THE DATASET.

Indicator	Description	How is it calculated?
Sales Revenues (SR)	The total income a company generates from sales of goods or services in a period before deducting expenses	*Number of Units Sold* $\times$ *Average Price Per Unit Sold*
Production Costs (PC)	The expenses of a company to produce goods or provide services, e.g., materials, labor, and overhead costs	*Raw Materials Cost* + *Direct Labor Cost* + *Overhead Cost*
EBITDA	Earnings Before Interest, Taxes, Depreciation, and Amortization	SR $-$ PC $-$ *Operating Expenses*
Net Profit (NP)	The amount of money a company earns after subtracting all expenses (e.g., operating costs, taxes, and interest) from its total revenue	SR $-$ *Total Expenses*
Total Assets (TA)	The sum of all owned resources of the company, and includes both *current* (e.g., cash and inventory) and *non-current* (e.g., property and equipment) assets	SE + *Liabilities*
Shareholder's Equity (SE)	The residual value of the assets of a company after subtracting all the liabilities	TA $-$ *Total Liabilities*
Net Financial Position (NFP)	The measure of the financial stability of a company	*Total Cash* $-$ *Total Debt*

and without each feature, using a loss function $\ell(\cdot,\cdot)$ to assess how much the feature improves predictive performance.

Formally, given a model $\mathcal{M}(\cdot)$ (we drop the subscript p for brevity) predicting a response variable y (in our case $y \rightarrow \hat{x}_p^{n+1}$) based on input features $\boldsymbol{x}$ (in our case $\boldsymbol{x} \rightarrow \{\boldsymbol{x}^{n-W+1}, \ldots, \boldsymbol{x}^{n-1}, \boldsymbol{x}^n\}$), SAGE evaluates the performance of $\mathcal{M}(\cdot)$ on subsets of features $\boldsymbol{X}_S \triangleq \{X_i | i \in S\}$ for different $S \subseteq D$, where $D = \{1, \ldots, n\}$.

To this end, SAGE introduces the function $v_{\mathcal{M}} : \mathcal{P}(D) \rightarrow \mathbb{R}$—where $\mathcal{P}(D)$ is referred to as the power set—which quantifies the amount of predictive power $\mathcal{M}(\cdot)$ derives from the subset of features $\boldsymbol{X}_S$. This function is defined as: $v_{\mathcal{M}}(S) = \mathbb{E}[\ell(\mathcal{M}_\varnothing(\boldsymbol{X}_\varnothing), Y)] - \mathbb{E}[\ell(\mathcal{M}_S(\boldsymbol{X}_S), Y)]$ where $\mathcal{M}_S(\boldsymbol{X}_S) = \mathbb{E}[\mathcal{M}(\boldsymbol{X}) | \boldsymbol{X}_\mathcal{S} = \boldsymbol{x}_\mathcal{S}]$ is the conditional expectation function. It represents the model's prediction when only the features in $\overline{S} \triangleq D \backslash S$ are known and thus the influence of the missing features is properly marginalized. Conversely, $\mathcal{M}_\varnothing(\boldsymbol{X}_\varnothing) = \mathbb{E}[\mathcal{M}(\boldsymbol{X})]$ corresponds to the mean prediction.

The function $v_{\mathcal{M}}$ is used to derive Shapley values, quantifying the contribution ϕ_i of each feature X_i, to the overall performance of the model. Specifically, SAGE assigns feature importance using the Shapley values of the model-based predictive power, $\phi_i(v_{\mathcal{M}})$, which is referred to as *SAGE values*.

SAGE values are the expectation of per-instance SHAP values applied to the model loss, namely: $v_{\mathcal{M}}^{(\boldsymbol{x},y)}(S) = \ell(\mathcal{M}_\varnothing(\boldsymbol{x}_\varnothing), y) - \ell(\mathcal{M}_S(\boldsymbol{x}_S), y)$ which measures the change in loss for a specific sample $(\boldsymbol{x}, y)$ when using the subset of features $\overline{S}$ compared to using no features. The *global importance* for each feature, $\theta_i(v_{\mathcal{M}})$, is obtained by averaging the per-instance SHAP values across all instances, namely: $\phi_i(v_{\mathcal{M}}) = \mathbb{E}_{\boldsymbol{X}Y}[\phi_i(v_{\mathcal{M}}^{(\boldsymbol{X},Y)})]$. In this work, we use the official SAGE Python package[1] to efficiently compute approximated SAGE values via the Monte Carlo method.

IV. EXPERIMENTAL SETUP

Here, we outline the experimental setup: Sec. IV-A covers the dataset and pre-processing, while Sec IV-B details the evaluation procedure and performance metrics.

[1]https://pypi.org/project/sage-importance/

A. Dataset and Pre-Processing Operations

Dataset Description. We exploit a private dataset initially comprising 673 samples (viz. Italian publicly owned companies), each described by 53 distinct features. After pre-processing, the dataset was refined to include 529 companies. Iniziativa Cube S.r.l., a medium-sized Italian consulting company, provided the data under an NDA. For each company in the dataset, we were provided with basic information—such as Company Name, Fiscal Code, Registered Office Address, ATECO[2] Code, ATECO Description, and (per-year) Employee Count—along with *seven economic-financial indicators* spanning 2016–2021. These indicators—including Sales Revenues (SR), Production Costs (PC), Earnings Before Interest, Taxes, Depreciation, and Amortization (EBITDA), Net Profit (NP), Total Assets (TA), Shareholder's Equity (SE), and Net Financial Position (NFP)—form the core of our predictive analysis. More details on their meaning and how they are calculated are reported in Tab. I. They represent key features and serve as targets, offering a comprehensive view of each company's financial performance over the six years.

Management of Missing Values. Handling missing values is crucial for ensuring prediction accuracy. Hence, we identified and filtered out the companies with more than 16 missing values, reducing the dataset from 673 to 565 companies. For these remaining companies, missing values were addressed through *interpolation*: using the previous year's data for the last year, the second year's data for the first year, and averaging adjacent years for intermediate gaps.

Management of Outliers. Outliers, especially for 2021 EBITDA, significantly affected the dataset. Then, we used a *percentile-based* approach, removing companies corresponding to the top 1% of extreme EBITDA values per year (2016–2021), reducing the dataset from 565 to 529 companies.

Management of Categorical Variables. We defined three categorical variables: (a) Number of Employees, (b) Geographic Zone, and (c) Sector. For the "Number of Employees", we created three groups: 0-9, 10-49, and 50+, with 50+ being the

[2]ATECO categorizes Italian economic activities into different sectors.

largest. For the "Geographic Zone", we used Google's Bard (now Gemini) AI[3] to classify companies into *North*, *Centre*, or *South* Italy based on their registered office addresses, with the North being the most represented and containing about half of the companies. For the "Sector", we grouped the companies into eight categories using the first two digits of their ATECO Code: (*i*) Agriculture (Agri), (*ii*) Energy (Ener), (*iii*) Service Management (SM), (*iv*) Infrastructure (Infr), (*v*) Pharmacy (Pharm), (*vi*) Transports (Tran), (*vii*) Service for Public Companies (SPC), and (*viii*) Consulting (Cons).

Input Data Formatting. To forecast economic and financial trends for public companies, appropriate data formatting is essential for ML model compatibility. By leveraging (a) sliding windows, (b) padding value handling, and (c) data scaling, we prepared the data for model training. We tested window sizes of $W = \{1, \dots, 5\}$ years and normalized each feature via the *Min-Max* scaler to ensure a common scale. This ensured that the minimum value for each feature, represented by the padding value, corresponded to 0 after scaling, resulting in a uniform data distribution across all features.

B. Evaluation Procedure and Error Metrics

We used a stratified 10-fold cross-validation to ensure proper dataset split into training and test sets [21]. This approach allows for a more stable evaluation of models, while maintaining computational efficiency. To achieve a balanced distribution of company profiles across the folds, we introduced a column representing the Cartesian product of ATECO categories and geographic zones (24 distinct profiles) for stratification. In such a way, companies are proportionally represented in the training and test sets for each fold.

Finally, model predictive capability is evaluated using various metrics. In most analyses, we consider the Root Mean Squared Error (**RMSE**) and Median Absolute Error (**MedAE**).

V. Experimental Evaluation

In this section, we first investigate the performance trend of ML predictors when varying the memory size W. Then, we delve into the prediction performance by evaluating the effect of the *Number of Employees*, *Geographic Zone*, and *ATECO Sector*. Lastly, we exploit our SAGE-based interpretability approach to *understand* the contribution of all current indicators to the performance of future indicators and *improve* them by retaining only informative features. Below, we use italicization to highlight key takeaways from the analyses.

How Does Memory Size Affect Prediction Performance? Herein, we examine how memory size W influences prediction performance to identify its optimal value. Figure 2 depicts the prediction error distribution (viz. RMSE) across all ML models as W changes for SR, EBITDA, and TA. The results for other indicators are omitted for brevity, as their trends are comparable to those presented.[4] Overall, *increasing W offers no evident or generalizable performance gains for the predicted indicators*, as smaller W often yields median errors comparable to larger W values. For instance, in the case of LIR, using only the indicators from the previous year ($W = 1$) yields a limited increase or decrease in the median RMSE—namely, $\approx +5\%$ for EBITDA and $\approx -5\%$ for SR—compared to considering data from all available years ($W = 5$). Conversely, only in very few cases the median error is reduced when using a larger memory (e.g., $\approx -23\%$ for TA using LIR with $W = 5$).

Comparing the performance of ML models, LIR *and* RFR *consistently achieve the best results*. Specifically, LIR ($W = 1$) achieves a median RMSE $\approx 28\%$ lower (i.e. 4.25 vs. 5.99 M€) for SR and $\approx 7\%$ lower (i.e. 2.13 vs. 2.28 M€) for EBITDA compared to RFR. For TA, this trend reverses and RFR reaches a median RMSE $\approx 32\%$ (i.e. 12.13 vs. 16.03 M€) and $\approx 19\%$ (i.e. 10.50 vs. 13.00 M€) lower than LIR, with $W = 1$ and $W = 5$, respectively. *As a consequence, in the following analyses, we will exploit either* LIR *or* RFR *with a memory of one ($W = 1$) or five ($W = 5$) years, depending on the specific indicator to be predicted.*

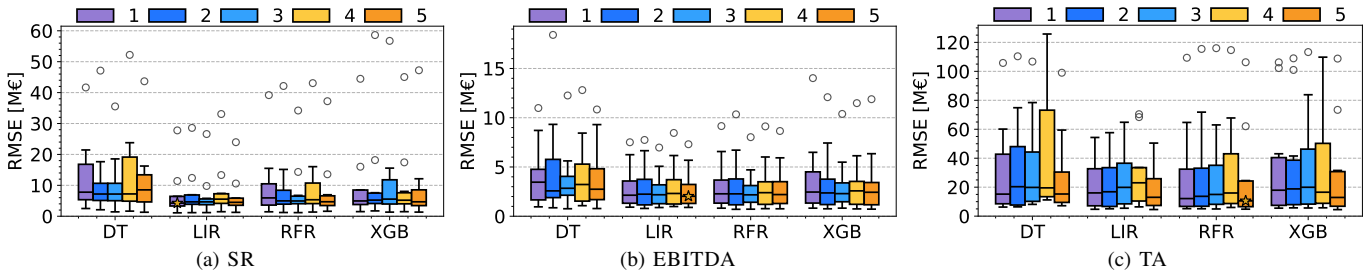
How Does Error Vary with the Number of Employees? We analyze how prediction error changes with the *Number of Employees*, categorized into three classes: 0-9, 10-49, and 50+. Fig. 3 shows the distribution of the absolute prediction error ($|\hat{x}^{n+1}_{(\cdot)} - x^{n+1}_{(\cdot)}|$) for NP and PC when using the best-performing LIR with $W = 1$ and $W = 5$, respectively. *The prediction error varies based on the number of employees and the financial indicator predicted.* For instance, in the case of NP (Fig. 3a), the maximum error is approximately an order of magnitude lower than that observed for PC (Fig. 3b). Similarly, median errors range within €$(200, 500)$K for NP, while they span from €300K to €2M for PC.

On the other hand, *companies with a larger number of employees (viz. 50+) consistently show higher prediction errors*, regardless of the predicted feature. In contrast, *the error is similar for medium (viz. 10-49) and small (viz. 0-9) companies*, making it difficult to establish a relationship between the error and company size/scale. Particularly, for NP, the error is $\leq$ €300K in 65% of cases for companies with up to 49 employees, whereas this holds for only 40% of cases for companies with 50+ employees. Notably, the error is $\leq$ €1M in at most 85% of cases for companies with 0-9 and 10-49 employees. A similar pattern is observed for PC, where the error is $\leq$ €1M in 70% of cases for companies with 0–9 and 10–49 employees, and in only 40% of cases for companies with 50+ employees.

How Does Error Change Based on Geographic Zone? We aim to analyze the dependence of the prediction error on the *Geographic Zone*. Fig. 4a depicts the prediction error ($\hat{x}^{n+1}_{(\cdot)} - x^{n+1}_{(\cdot)}$) on NFP (using LIR with $W = 5$) across three geographical zones in Italy, i.e. North, Centre, and South. By construction, negative values in the distribution correspond to underestimation (meaning the prediction is less than the actual value). In contrast, positive values indicate overestimation (meaning the prediction is greater than the actual value).

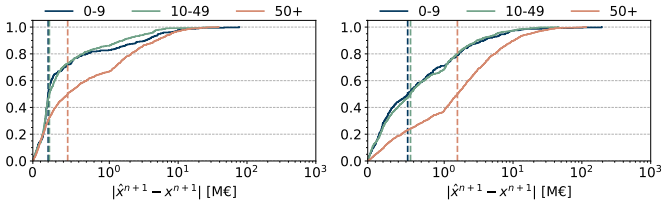
[3] https://gemini.google.com/

[4] In detail, (a) EBITDA $\approx$ NP, (b) NFP & PC $\approx$ SR, and (c) TA $\approx$ SE.



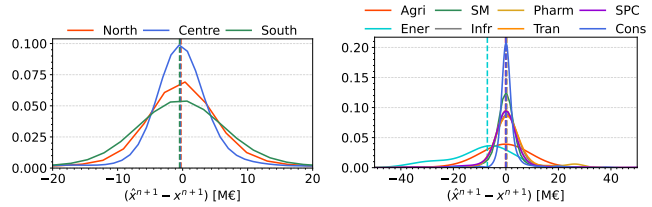
(a) SR (b) EBITDA (c) TA

Fig. 2. RMSE distributions of (a) SR, (b) EBITDA, and (c) TA as $W \in \{1, 2, 3, 4, 5\}$. Values are computed over 10 folds.



(a) NP (LIR w/ $W = 1$) (b) PC (LIR w/ $W = 5$)

Fig. 3. CDF of the absolute prediction error on (a) NP and (b) PC depending on the *Number of Employees*. Values on the x-axis are in the *symlog* scale. The **- - -** depicts the median value.



(a) NFP (LIR w/ $W = 5$) (b) SR (LIR w/ $W = 1$)

Fig. 4. KDE of the prediction error on (a) NFP grouped by *Geographic Zone* and (b) SR grouped by *Ateco Sector*. The **- - -** depicts the median value.

When comparing median values, *we observe slight under-confidence for all areas, with errors ranging from* €−469K *(South) to* €−258K *(North)*. Errors are *more dispersed in the South*, reflecting less reliable predictions, while errors are more concentrated elsewhere, indicating greater accuracy.

How Does the Error Change Based on ATECO Sector? Following the previous analysis, we examine how prediction error varies by *ATECO Sector*. Fig. 4b illustrates SR prediction errors across eight sectors (using LIR w/ $W = 1$). *Overall, we note that errors depend on the specific sector.* In detail, *the model is notably under-confident for companies in the Energy (Ener) sector*, with a median error of approximately €−7M. Conversely, *the model is slightly over-confident for companies in sectors like Infrastructure and Transports, and under-confident for those in Agriculture*, with median errors of €220-300K and €−140K, respectively. We note a median error for the remaining sectors ranging from €−67K to €44K. Finally, *residuals are more dispersed for Agriculture and Energy*, but more concentrated in other sectors, especially *Service Management* and *Consulting*.

Interpreting Model Prediction via SAGE. We use SAGE to assess how economic and financial indicators from previous years impact the prediction of the same indicators for the following year. Figure 5 reports the importance of indicators used to feed the model when predicting NP, SE, and TA. The results refer to the best-performing model for each target indicator, namely LIR ($W = 1$), LIR ($W = 5$), and RFR ($W = 5$) for NP, SE, and TA, respectively.[5]

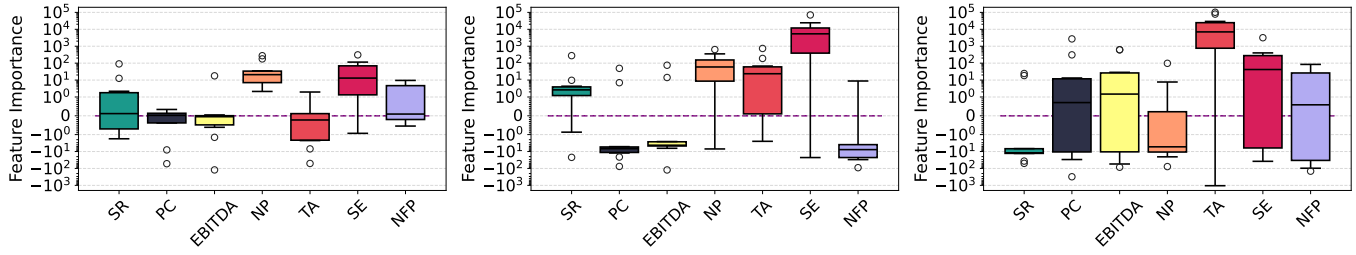
For each predicted indicator, the corresponding indicator from the previous years (depending on the memory W) consistently has a positive effect on the final decision of the model. Notably, this effect is significantly higher for SE (Fig. 5b) and TA (Fig. 5c) compared to other indicators. Also, both NP and TA positively impact SE, and vice versa, suggesting *a strong correlation between these indicators*. In particular, TA, which include all company resources (e.g., liquidity, property, equipment), are partly financed through SE. Similarly, NP, representing the company's actual earnings, contributes to the growth of SE when profits are retained rather than distributed as dividends. On the other hand, *the impact of other indicators varies depending on the target*. For instance, PC and EBITDA (resp. SR) positively (resp. negatively) influence TA but negatively (resp. positively) affect SE, with a negligible impact on NP.

Improving Model prediction via SAGE. Building on the previous findings, we assess how prediction performance changes when various economic-financial indicators, identified by SAGE as "non-informative", are occluded from the model

[5]When $W > 1$, we obtain the importance of each feature by summing the corresponding values over all the years considered to feed the model.

TABLE II
COMPARISON OF THE MODEL WITH "ALL" FEATURES (⊗) WITH THE MODEL USING ONLY SAGE "INFORMATIVE" FEATURES (✓) POSITIVELY AFFECTING THE PREDICTION OF NP, TA, AND SE.[6] VALUES REFER TO THE RMSE [M€] COMPUTED AS $mean/median/IQR$ OVER 10 FOLDS.

	NP (LIR w/ $W = 1$)	SE (LIR w/ $W = 5$)	TA (RFR w/ $W = 5$)
⊗	3.27/2.45/2.84	7.27/4.73/5.79	26.11/10.54/17.74
✓	3.22/2.44/2.59	7.13/4.72/5.80	25.89/9.96/16.87



(a) NP (LIR w/ $W = 1$) (b) SE (LIR w/ $W = 5$) (c) TA (RFR w/ $W = 1$)

Fig. 5. Distributions of SAGE-powered feature importance values when predicting (a) NP, (b) SE, and (c) TA. Values are computed over 10 folds.

inputs. Accordingly, Tab. II compares the performance—in terms of RMSE—of the previous models (⊗), with that obtained when considering only the features that positively affect the prediction (⊘) of NP, SE, and TA.[6]

Excluding "negative" financial indicators improve performance, varying by predicted indicator. The largest gain is for TA (error reduced by €≈ 1.04M), while NP and SE see smaller improvements of €−40K and €−50K, respectively.

VI. Conclusions

This work explored predicting economic and financial indicators for publicly owned Italian companies using ML models trained on historical time-series data. **Key findings** are outlined below: increasing memory (W) did not provide guaranteed benefits, while LIR and RFR consistently outperformed other models. Prediction errors varied based on the financial indicator and company size, with larger errors observed for NP and companies with over 50 employees. Regional differences in Italy affected NFP predictions, with wider errors in the South and more consistent results in the North and Centre. Sector disparities were also evident, with notable underestimation in the Energy sector and mixed results across other sectors. SAGE analysis showed that prior-year SE and TA strongly influenced predictions, while PC had little impact. Excluding certain indicators improved performance for specific targets. **Future work** could benefit from (i) classifying companies by life-cycle stage, (ii) improving the handling of missing values, (iii) exploiting multimodal data (e.g., textual reports), (iv) exploring dimensionality reduction to enhance model accuracy, and (v) leveraging the obtained findings to instruct large language models with tailored prompts for generating analyst-level explanations of the forecasts.

References

[1] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, "Machine learning in financial crisis prediction: a survey," *IEEE Trans. on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 421–436, 2011.

[2] F. Rundo, F. Trenta, A. L. Di Stallo, and S. Battiato, "Machine learning for quantitative finance applications: A survey," *Applied Sciences*, vol. 9, no. 24, p. 5574, 2019.

[3] Y. Tang, Z. Song, Y. Zhu, H. Yuan, M. Hou, J. Ji, C. Tang, and J. Li, "A survey on machine learning models for financial time series forecasting," *Neurocomputing*, vol. 512, pp. 363–380, 2022.

[6]NP = {NP, SR, SE, NFP}; SE = {SR, NP, TA, SE}; TA = {PC, EBITDA, TA, SE,NFP}.

[4] I. Covert, S. M. Lundberg, and S.-I. Lee, "Understanding global feature contributions with additive importance measures," *NeurIPS*, 2020.

[5] M. Kayakus, B. Tutcu, M. Terzioglu, H. Talaş, and G. F. Ünal Uyar, "ROA and ROE forecasting in iron and steel industry using machine learning techniques for sustainable profitability," *Sustainability*, vol. 15, no. 9, p. 7389, 2023.

[6] P. V. H. Son and L. T. Duong, "Research on applying machine learning models to predict and assess return on assets (ROA)," *Asian Journal of Civil Engineering*, vol. 25, pp. 4269–4279, 2024.

[7] J. C. Tellez Gaytan, K. Ateeq, A. Rafiuddin, H. M. Alzoubi, T. M. Ghazal, T. A. Ahanger, S. Chaudhary, and G. Viju, "AI-based prediction of capital structure: Performance comparison of ANN, SVM and LR models," *Comp. Intelligence and Neuroscience*, vol. 2022, no. 1, p. 8334927, 2022.

[8] P. Chakri, S. Pratap, Lakshay, and S. K. Gouda, "An exploratory data analysis approach for analyzing financial accounting data using machine learning," *Decision Analytics Journal*, vol. 7, p. 100212, 2023.

[9] E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," *Expert Syst. with Applications*, vol. 83, pp. 187–205, 2017.

[10] P. Supsermpol, V. N. Huynh, S. Thajchayapong, and N. Chiadamrong, "Predicting financial performance for listed companies in Thailand during the transition period: A class-based approach using logistic regression and random forest algorithm," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 9, no. 3, p. 100130, 2023.

[11] M. Abdullah, Z. Sulong, and M. A. F. Chowdhury, "Explainable deep learning model for stock price forecasting using textual analysis," *Expert Systems with Applications*, vol. 249, p. 123740, 2024.

[12] X. Yu, Z. Chen, Y. Ling, S. Dong, Z. Liu, and Y. Lu, "Temporal Data Meets LLM–Explainable Financial Time Series Forecasting," *arXiv preprint arXiv:2306.11025*, 2023.

[13] K. J. Koa, Y. Ma, R. Ng, and T.-S. Chua, "Learning to Generate Explainable Stock Predictions using Self-Reflective Large Language Models," in *ACM WWW'24*, 2024.

[14] A. K. Pundir, L. Ganapathy, P. Maheshwari, and M. N. Kumar, "Machine learning for revenue forecasting: A case study in retail business," in *IEEE IEMCON*, 2020.

[15] X. Dairu and Z. Shilong, "Machine learning model for sales forecasting by using XGBoost," in *IEEE ICCECE*, 2021.

[16] M. Gurnani, Y. Korke, P. Shah, S. Udmale, V. Sambhe, and S. Bhirud, "Forecasting of sales by using fusion of machine learning techniques," in *IEEE ICDMAI*, 2017.

[17] Y.-g. Song, Q.-l. Cao, and C. Zhang, "Towards a new approach to predict business performance using machine learning," *Cognitive Systems Research*, vol. 52, pp. 1004–1012, 2018.

[18] I. Evdokimov, M. Kampouridis, and T. Papastylianou, "Application Of Machine Learning Algorithms to Free Cash Flows Growth Rate Estimation," *Procedia Computer Science*, vol. 222, pp. 529–538, 2023.

[19] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *NeurIPS*, 2017.

[20] A. Montieri, G. Bovenzi, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapè, "Packet-level prediction of mobile-app traffic using multitask deep learning," *Computer Networks*, vol. 200, p. 108529, 2021.

[21] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: an empirical study on performance estimation methods," *Machine Learning*, vol. 109, no. 11, pp. 1997–2028, 2020.