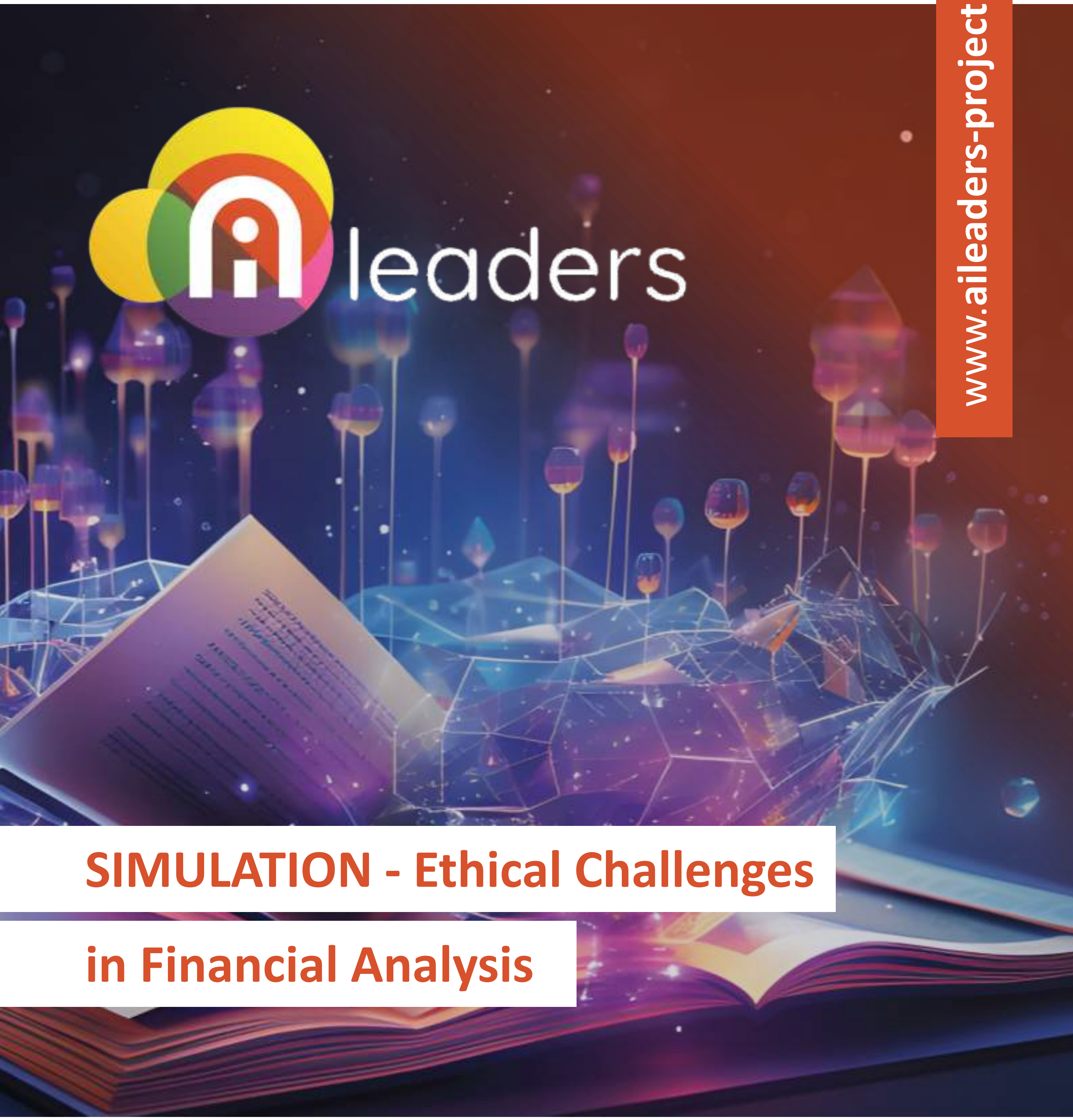


www.aileaders-project.eu

SIMULATION - Ethical Challenges in Financial Analysis

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

SIMULATION - Ethical Challenges in Financial Analysis

01 Abstract

Type of OER

Demo/Simulation using Google Colab
(RandomForest vs FairGBM Classifier)

Goal/Purpose

To provide students with a practical and critical exploration of how algorithmic decision-making in financial analysis and forecasting can reproduce structural inequalities by comparing outputs from fairness metrics. This simulation encourages reflection on the ethical dimensions of financial automation and promotes the development of fairer, more inclusive predictive models.

Expected Learning Outcomes:

By the end of the simulation, students will be able to:

01 Detect and interpret algorithmic bias in credit approval;

02 Reflect on the ethical implications of automated financial decisions.

03 Use fairness metrics to evaluate model outcomes;

Keywords:

- Machine Learning
- Classification
- Biases
- Fairness

Suggested Methodological Approach:

Problem-Based Learning

NOTE

Intermediate knowledge of **Python programming** is required to understand and work with the contents of this OER.

02 Introduction

AI-driven financial forecasting tools are increasingly being adopted to automate credit-related decisions, such as loan approvals and risk assessments (see, for example, Chen, 2020; Dastile et al., 2020; Chen et al., 2024; Heß & Damásio, 2025). These technologies offer considerable benefits in terms of efficiency, consistency, and scalability. However, when deployed in contexts historically shaped by structural inequalities, they raise significant ethical concerns.

A central issue is the potential for these systems to replicate—or even amplify—existing biases embedded in financial data. Algorithms that appear neutral may, when trained on biased or incomplete datasets, produce unfair outcomes that disproportionately impact already marginalised or vulnerable groups.

A particularly compelling example is that of women (Orser et al., 2006; Ongena & Popov, 2016; Moro et al., 2017; Beck et al., 2018; De Andrés et al., 2021, among others) and Black individuals (Chatterji & Seamans, 2012), who continue to face compounded barriers in accessing credit from traditional financial institutions due to intersecting forms of discrimination.

This simulation-based exercise invites students to critically explore the intersection of machine learning, financial forecasting, and fairness. Using a synthetic dataset that reflects realistic credit applications, students will investigate how attributes such as gender, occupation, education, and marital status influence loan approval outcomes. The dataset has been deliberately constructed to encode subtle patterns of bias, offering a hands-on opportunity to detect, analyse, and mitigate disparities through applied experimentation.

Rooted in the broader literature on financial inclusion and funding gaps, this exercise challenges students to reflect on how AI systems can either reinforce or confront systemic injustice. In doing so, they will not only gain technical competencies in fairness auditing and bias mitigation, but also develop a critical understanding of the ethical responsibilities associated with designing and deploying AI in financial ecosystems.

03 Tools Presentation

The **core scenario** presented in this simulation centres on a dataset of synthetic loan applications.

The data includes demographic and socioeconomic features such as:

- Gender
- Race
- Age
- Occupation
- Education level
- Marital status

*The target variable is **loan approval,** modelled as a binary classification problem.*

While the dataset is constructed to reflect realistic conditions, it intentionally incorporates hidden patterns of bias. For example, rural women entrepreneurs may face higher rejection rates than their male or urban counterparts, even when controlling for qualifications and risk factors. This simulates how historical and systemic barriers can be replicated in automated systems.

The objective is to investigate two primary questions:

01 ML algorithms disproportionately "recommend" rejection of credit applications submitted by Black and Women individuals in binary predictive models?

02 What structural or algorithmic factors may contribute to this disparity?

04 Simulation Execution

01 Access the Simulation Notebook

Go to https://tinyurl.com/2txxdk9

02 Run the Code

Execute all the cells in the notebook sequentially. Make sure no errors occur and that all outputs are correctly displayed. To do so, please click play at the top-left of each cell.

03 Explore the Dataset[1]

Review the dataset structure and contents. Pay close attention to key demographic variables such as gender, occupation, education, and loan outcome. Observe any potential imbalances or patterns that could indicate bias.

04 Analyse Fairness

Examine results. Compare approval rates across different demographic groups, particularly focusing on gender and race. Use fairness metrics provided in the notebook (e.g., demographic parity, disparate impact) to quantify potential bias.

05 Compare Results

Carefully compare model outputs before and after applying fairness interventions, if included.

06 Reflect on Findings

Based on the observed results, reflect on the ethical and practical implications of using AI in financial decision-making. Consider the trade-offs between accuracy, efficiency, and fairness.

1. What If Tool Tutorial -https://www.youtube.com/watch?v=jHojeFCc5HE

05 Conclusion

The data we relied on this simulation is widely used to study fairness and bias in machine learning due to its inclusion of sensitive attributes like gender and race. It is commonly used to predict whether an individual's income exceeds $50,000 per year based on features such as age, education, occupation, and marital status

Our simulation demonstrates how AI-based financial forecasting systems, if left unexamined, can perpetuate structural biases embedded in historical data.

It reveals that the predictive accuracy of credit rejection in binary models is higher for Black individuals and women than for white individuals and men.

Output Summary

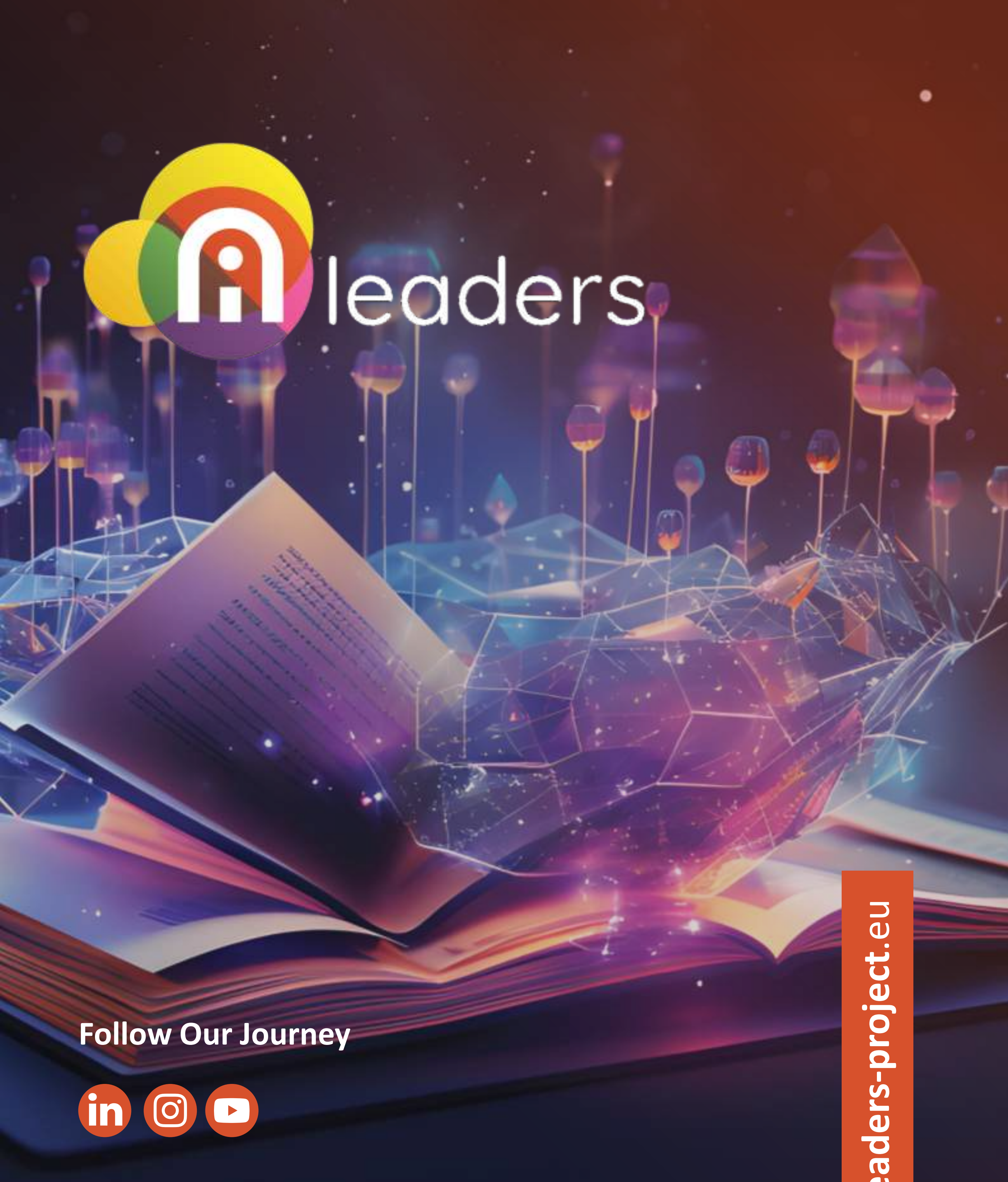
Accuracy in Predicting Credit Rejection Decisions —i.e., True "Negatives" (%)

Even in simplified models like the one at hand, the presence of biased outcomes - such as the lower percentage of false negatives in predicted rejection rates for Black individuals and women - highlights the ethical responsibility of developers and analysts to audit and refine AI systems. These results reignite the discussion on blind spots in credit scoring models (see, for example, Robb & Robinson, 2018).

By engaging directly with fairness metrics and mitigation strategies, students not only acquire technical tools for identifying bias, but also develop a deeper understanding of how such biases manifest and how they can be addressed. This exercise reinforces the principle that fairness in financial AI is not a secondary concern, but a core component of responsible and sustainable innovation.

06 References

- Beck, T., Behr, P., & Madestam, A. (2011). Sex and credit: Is there a gender bias in lending?. European Banking Center Discussion Paper, (2011-027).
- Chatterji, A. K., & Seamans, R. C. (2012). Entrepreneurial finance, credit cards, and race. Journal of Financial Economics, 106(1), 182-195.
- Chen, T. H. (2020). Do you know your customer? Bank risk assessment based on machine learning. Applied Soft Computing, 86, 105779.
- Chen, Y., Calabrese, R., & Martin-Barragan, B. (2024). Interpretable machine learning for imbalanced credit scoring datasets. European Journal of Operational Research, 312(1), 357-372.
- Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing, 91, 106263.
- De Andrés, P., Gimeno, R., & de Cabo, R. M. (2021). The gender gap in bank credit access. Journal of Corporate Finance, 71, 101782.
- Heß, V. L., & Damásio, B. (2025). Machine learning in banking risk management: Mapping a decade of evolution. International Journal of Information Management Data Insights, 5(1), 100324.
- Moro, A., Wisniewski, T. P., & Mantovani, G. M. (2017). Does a manager's gender matter when accessing credit? Evidence from European data. Journal of banking & finance, 80, 119-134.
- Ongena, S., & Popov, A. (2016). Gender bias and credit access. Journal of Money, Credit and Banking, 48(8), 1691-1724.
- Orser, B. J., Riding, A. L., & Manley, K. (2006). Women entrepreneurs and financial capital. Entrepreneurship Theory and practice, 30(5), 643-665.
- Robb, A., & Robinson, D. T. (2018). Testing for racial bias in business credit scores. Small Business Economics, 50, 429-443.



Follow Our Journey

www.aileaders-project.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.