Ethics and bias in AI: a potential challenge to fair economic progress

Bogdan Mirea[1]* and Giani-Ionel Gradinaru[2]

To cite this article:
Mirea, B., Gradinaru, G.I. (2026). Ethics and bias in AI: a potential challenge to fair economic progress. Romanian Journal of Economics, 62 (1), 99-110.

Abstract. *The swift integration of Artificial Intelligence (AI) into economic systems has fundamentally transformed and redefined decision-making processes across sectors such as finance, healthcare, and employment. However, this continuous innovation in emerging technologies also leads to the emergence of ethical challenges, particularly in terms of bias, transparency, and fairness. The objective of this study is to investigate the implications of the integration of artificial intelligence (AI) models in financial decision-making processes on the promotion of equitable economic growth. By focusing on a real-world use case in the banking sector, we highlight the ethical implications of biased AI decisions and contribute to the wider academic and societal conversation on responsible digital transformation. The study uses the publicly available German Credit Data on which we train an XGBoost classification model to predict the likelihood of a customer subscribing to a term deposit. To interpret the model's behaviour and identify the leading and most impactful factors driving its predictions, we apply SHAP (SHapley Additive exPlanations), a state-of-the-art interpretability technique. This approach allows us to analyze the marginal impact of each feature, such as age, job, marital status, or education on the model's outputs, thus revealing potential ethical concerns regarding sensitive variables. Our SHAP analysis reveals that certain socio-economic attributes significantly influence the model's predictions and certain categories are often favoured or penalized in ways that may not be economically or ethically justified. The results demonstrate that even highly accurate models can embed latent biases derived from historical data patterns. Furthermore, we show that omitting or transforming sensitive features alters the model's behaviour, but not always in predictable or fairways. These findings suggest that algorithmic decision-making in finance must be continuously audited for fairness and transparency to prevent systemic discrimination. This study integrates explainable machine learning, ethics, and economic policy. Unlike research focused mainly on accuracy, it highlights the interpretability and socio-economic effects of AI decisions. Using SHAP with XGBoost in a financial inclusion context, we propose a framework to detect and visualize bias in AI systems influencing economic access. The findings highlight the need for AI tools that supports progress while protecting ethical standards and encourage stakeholders to consider how AI shapes fairness and opportunity.*

Keywords*: AI ethics, algorithmic bias, economic fairness, credit scoring, SHAP, XGBoost, digital progress, machine learning fairness, responsible AI, financial inclusion*
JEL classification*: C45, C53, D63, G21*

[1] The Bucharest University of Economic Studies, Bucharest, Romania; Corresponding author: mireabogdan17@stud.ase.ro.
[2] The Bucharest University of Economic Studies, Romania; Institute of National Economy - Romania Academy, Bucharest, Romania; giani.gradinaru@csie.ase.ro

1. Introduction

Artificial Intelligence is truly transforming the global economy, reshaping decision making across different sectors like finance, where it powers applications such as credit scoring, customer segmentation or targeted marketing. In other sectors, like banking, AI powered systems offer a great processing efficiency, enabling businesses to analyse bigger datasets than ever before and deliver personalized services at scale. Predictive models can group potential customers for different financial products, expanding access to diverse populations. Yet, this technological step forward presents itself with many challenges. As these AI-powered systems expand in influence in economic outcomes, concerns about their bias, fairness and transparency have emerged as critical points to fair progress. Biased algorithms, usually trained on historical data that retains societal inequities can propagate discrimination and favour certain demographic groups while marginalizing others. This gap between AI's push for offering tailored economic solutions and its risk of perpetuating disparities is the focus in this study.

The main objective of this study is to assess how AI-based financial decision models influence equity and fairness in economic systems, with a focus on identifying and mitigating potential algorithmic biases. By focusing on a real-world use case in the banking sector, we highlight the ethical implications of biased AI decisions and contribute to the wider academic and societal discussion on responsible digital transformation.

The integration of AI in finance sectors as a decision maker raises ethical concerns that are not intensely addressed. One example is if a model keeps denying loans for younger or lower-income individuals, it has the potential to deepen existing financial gaps, therefore threatening economic fairness. Bias in AI is derived from skewed training data, a bunch of adjacent variables like city codes that can correlate to ethnicity or model design choices. These types of issues are not even just technical ones, they have broader social and economic consequences, with the potential to accelerate inequity in access to financial products as these types of approaches are becoming more popular. To address these concerns, explainable Artificial Intelligence (XAI) techniques like SHAP (*SHapley Additive exPlanations*) can bring a lot more transparency into AI modeling behaviours and reveal how variables like age, gender or education constrain predictions.

This study investigates the results of AI financial decision makings for equitable economic growth. Using a Bank Credit dataset, we trained XGBoost classification models to predict the propensity of customers subscribing to different loan types. We use SHAP for different XGBoost models to interpret the predictions and to identify the main features that shape the predictions with the purpose to identify any bias. While doing so, we highlight the impact of these social economical attributes and the ethical concerns surrounding them, such as whether subgroups are unfairly favoured or disadvantaged. Our paper aims to add to the growing discussion on the topic of responsible decisions made by autonomous Ais, offering a hands-on approach and a practical framework for analysing bias and encouraging fairness in financial institutions.

The remainder of this paper is organized as follows: Section 2 presents the literature review and theoretical background. Section 3 describes the methodology. Section 4 discusses the results. Section 5 concludes with the main findings, limitations, and directions for future research.

2. Literature review

The rapid adoption of artificial intelligence (AI) across multiple industries has changed decision-making algorithmic reasoning now integrated in a multitude of sectors like healthcare, financial, educational, and other systems, and processes that were previously defined by human judgment and now are automatized. From medical advice and loaning decisions to the background checking of job candidates and ad recommendation systems, AI increasingly determine very personal outcomes that directly affect the lives of people and the social communities. While this redirection promises benefits of increased efficiency and scalability, it also faces us with inherent ethical challenges of bias, accountability and transparency. Greater reliance on black-boxed algorithmic systems in high-risk domains calls for performant examination of the way that fairness and interpretability are

theoretically proposed and used, with an increased importance where decisions have economic and social impacts (Wachter et al., 2021).

A lot of AI bias and unfairness is rising as the spread of this way of operating is more used in every sector of industry. While this expansion of AI use in businesses of all kinds is compared to a new Industrial Revolution, Professor Veldkamp from Columbia University discovered that the adoption of AI made people more productive overall, while the companies with finance specific that adopted AI actually hired more people with AI specific skills (Abis & Veldkamp, 2023). Of course, the major differences between the Industrial Revolution and the AI Revolution are the special uncertainty feeling that major world events are fuelling, AI and the shifting employment conditions being a leading cause in these high uncertainty levels as reported by key financial institutions like the International Monetary Fund in their annual report (International Monetary Fund, 2024).

AI systems often copy or amplify the contextual injustices rather than reducing them (Bircan & Özbilgin, 2025). Defined as Weapons of Math Destruction in the homonym book written by Cathy O'Neil, new scoring algorithms called e-scoring are replacing traditional methods like FICO, which was used to assess a risk level for individual loan offering based on financial variables. The new e-scoring models take into account all types of personal data, they create a less transparent but often more performant credit risk models, reintroducing the bias that the classical FICO model was able to avoid (O'Neil, 2016). This new perspective accentuates the social risk of opaque algorithmic decisions in finance. Completely eliminating social-economic factors and "debiasing" data could reduce the prediction power and make models obsolete (Hassani, 2021). Loan decisions are heavily influenced by social bias present in financial data, causing the continuation of discrimination and even the increasing of it under AI models if data represents gender or ethnic differences, regulatory step-ups being the right thing, not excluding those sensitive data variables.

A growing portion of research accentuates the importance of including transparency and interpretability into AI-driven credit risk models, particularly to address fairness concerns. A comprehensive framework for developing credit scoring systems that are both ethically responsible and technically sound is proposed to ensure it (Szepannek & Lübke, 2021). Their approach uses both causal inference techniques with explainable AI, framing the principle of counterfactual fairness, the idea that a model should have the same outcome even if a protected attribute like gender or age is different. Using SHAP (Shapley Additive Explanations), they demonstrate through simulations and real data that fairness enhancement can be achieved without drastic sacrifices in predictive accuracy. This research is providing a blueprint for incorporating fairness limitations into real-world credit models while maintaining model interpretability.

Building on similar grounds, (Jammalamadaka & Itapu, 2022) examine the creation of a fair and interpretable credit scoring system using the German Credit dataset. Their study uses an XGBoost classifier and introduces a Disparate Impact Remover as a pre-processing technique to constrict bias, particularly related to gender and age. The authors highlight that this combination provides a fairer distribution of outcomes between protected and unprotected populations and keeps the model's performance. Their results show that, with model tuning and fairness-aware improvements, complex machine learning models can meet both ethical and technical standards.

Purificato et al. (2022) extend these principles further by suggesting an end-to-end responsible AI pipeline tailored to the loan approval process for banks. Their pipeline spans data handling, model training, and post-deployment monitoring all with built-in explainability tools and fairness checks in between. Explainable AI systems are emphasized in the paper not only for regulatory reasons but to win customers' trust. By making it easier for internal stakeholders and job applicants to understand and critique model outputs, their approach echoes broader efforts to institutionalize algorithmic transparency in finance.

SHAP is a tool encouraged by researchers as a mean to improve the transparency of black-box algorithms and as a way to provide the main stakeholders with a clear understanding of how features can influence the models and perceive potential bias in fintech (de Castro Vieira et al., 2025). The explorations of ML in financial systems and the ethical challenges of bias and fairness are aligned with the push for accountable AI that can be easily checked for discrimination. The credit scoring systems become more data-driven, and it should be carefully assessed with regulatory guidance because the traditional anti-discrimination techniques face new challenges in this era of complex machine learning,

leading to unintentional harmful results. Shapley values were used to quantify the contribution of features to biased outcomes and showed that there is a noticeable change in performance across demographic groups. A propensity score matching technique was used to improve fairness by adjusting the comparisons amongst groups so there is more equitable scoring. Their findings highlighted that bias comes without explicitly using protected attributes and credit models are vulnerable to becoming unfair and propagating bias (Babaei & Giudici, 2024). Another method is proposed using Shapley-Lorenz values, combining game theory and economics' Lorenz curves for inequality to assess the fairness of models across groups. This approach helps discover if any groups are receiving worse scores in crediting algorithms in a systemic manner and offers a way forward on the problem of explainable outputs, enabling performance alongside equity in credit ratings (Chen et al., 2024). By applying the Shapley value method to cohorts of loan requesters, we could also identify fairness in models' outcomes for different groups of borrowers. An analysis of UK SME loan data revealed notable discrimination against certain groups – specifically startups, smaller businesses, businesses run by females and ethnic minority owners faced disparities in financing. The Shap method was able to quantify how many variables were contributing to unfavourable loans distributions (Lu & Calabrese, 2023). As well as an ethical methodological approach, the transparency behind models should be dual facing, providing the loan applicants with a clear understanding of how their personal data is used in assessing risk and have a clear view of the whole decision process.

Another perspective is the trade-off between the power of predicting risk and the profit of financial loaning institutions, while fairness criteria is still important and used. A study found that various fairness variables can usually be satisfied with a minimal reduction in predictive accuracy and profit, using an equalization between true positive rates across groups as a beneficial fairness metric in credit scoring (Kozodoi et al., 2022).

3. Methodology and data

To explore the implications of algorithmic bias in financial decision-making, this study employs the German Credit Data from the University of California, Irvine (UCI) Machine Learning Repository. The dataset contains 1,000 entries, each representing a credit applicant, and includes a binary target variable indicating credit risk status: good or bad. This dataset is widely used in academic research for modeling and evaluating credit scoring systems due to its structured, real-world nature.

The dataset consists of a mix of categorical and numerical features that span demographic, financial, and behavioural attributes. Key variables are presented in Table 1.

Table 1. Main variables description

Employment status	Categorized by skill level or job type
Credit amount	Loan size in Deutsche Marks
Duration	Number of months for repayment
Housing status	Type of housing arrangement: own, rent, or free
Savings and checking account levels	Indicators of financial activity and buffer (savings level and account status)
Purpose of credit	Reason for the loan (e.g., car purchase, education, business)
Instalment rate	Monthly instalment as a percentage of disposable income
Residence duration	Number of years the applicant has lived at their current address
Telephone ownership	Whether the applicant has a telephone

Source: own processing.

This diversity of features provides a solid foundation to train a credit risk model while also examining fairness concerns, particularly related to sensitive socio-economic attributes such as age and gender.

The original dataset required several preprocessing steps to ensure data quality and enable effective model training. First, we addressed the combined field "Personal status and sex", which

encoded both gender and marital status. This variable was parsed into two distinct fields: Gender and Marital Status. This refinement was necessary to isolate gender as a potential sensitive attribute and ensure clarity in interpretability analysis.

Next, all categorical variables were numerically encoded to allow compatibility with gradient boosting algorithms. Any ambiguous or missing entries were examined and cleaned, ensuring the dataset's integrity prior to modeling.

During this stage, we also identified two sensitive features that are known to correlate with potential bias in credit decisions:

Gender: Derived from the parsed personal status field.

Age: A continuous variable representing the applicant's age in years.

These variables were carefully monitored throughout the modeling process to assess their influence on prediction outcomes. Additionally, proxy variables such as job type, savings account status, or housing were flagged for later SHAP analysis to detect indirect discrimination.

These features were chosen not only for their predictive value in credit risk modeling but also for their relevance in understanding socio-economic dynamics. Certain features such as telephone ownership, foreign worker status, or residence duration may act as proxies for financial exclusion or social vulnerability, further motivating the need for interpretability tools.

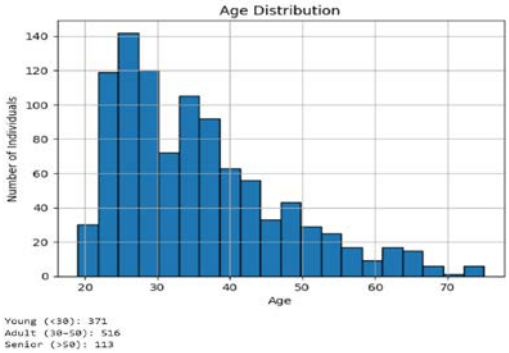
XGBoost, short for Extreme Gradient Boosting, is a machine learning algorithm that builds predictive models by combining many decision trees. It works by learning from the errors made by previous trees and gradually improving the overall prediction. What makes XGBoost particularly effective is its ability to handle large datasets quickly and its built-in techniques for preventing overfitting, which helps the model remain accurate on new, unseen data. It has become one of the most popular tools in data science, especially when working with structured data, such as customer information in banking or credit applications.

To make sense of XGBoost's predictions, especially since it can be quite complex, SHAP is often used. SHAP is a method that explains the contribution of each input feature to a specific prediction. It does this by calculating how much each feature increases or decreases the result, like how players in sporting event team are participating and contributing to the results. In the context of this study, SHAP helped reveal which customer characteristics such as age, employment, or account status had the most influence on whether someone was classified as a good or bad credit risk. More importantly, it helped identify whether certain sensitive features like gender or age were having a disproportionate impact, pointing to potential biases within the model. Through SHAP, the decision-making process of the model becomes more transparent and interpretable, allowing for a critical evaluation of fairness.

4. Research results and comments

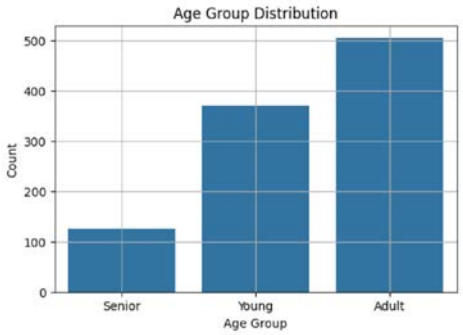
As a first step in the research, we used exploratory data analysis as a way of anticipating any bias or significant differences between groups, divided by variables that have a potential to influence the results of a propensity model.

Figure 1. Age distribution across database



Source: own processing in Python.

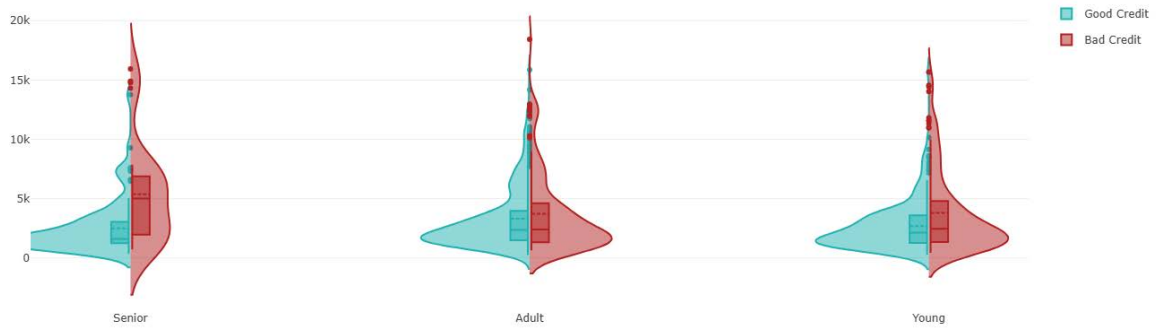
Figure 2. Age groups distribution



Source: own processing in Python.

In Figure 1, the age distribution in the database is not following a normal distribution, accentuating that younger people have higher credit needs, while older people to use tend to use fewer financial products. In the database, the defined age groups we used to split and perform the scoring modeling were also not equal, as an exact measurement would not permit a logical split in 3 groups depending on age, therefor, a category of young people, under 30 years old contains 371 individuals, the largest group would be comprised of 516 people accounting for the Adult category and the smallest group being the Senior one, with only 113 individuals. These reduced and different sample sizes were used to keep a decent and acceptable age difference, the overall database being too small to offer the change for a performant model to be trained. We sacrificed the comparability of models with the ones generally used in the industries that adopted AI modeling, keeping a strict bias analysis on these models, accentuating the changes for these different groups we defined.

Figure 3. Distribution of credit borrowed by age group

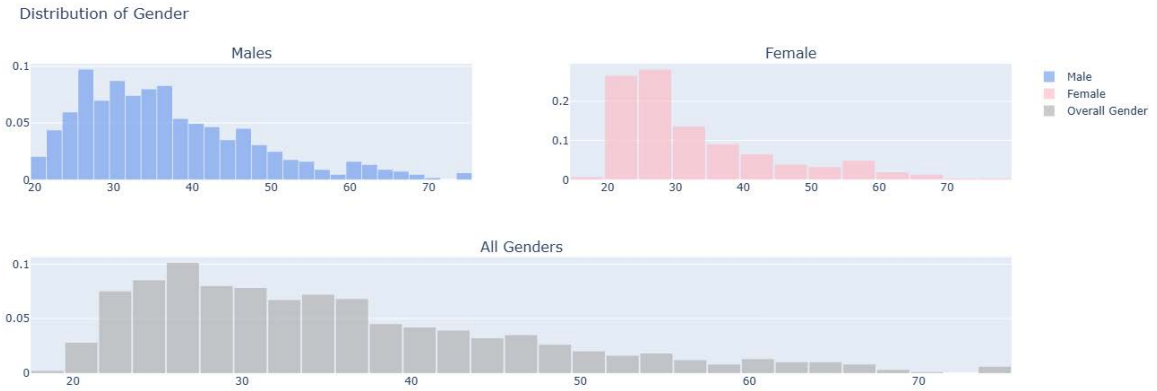


Source: own processing in Python.

For the credit distribution of bad or good creditors, we see a similar trend amongst our main age groups, the main difference being that senior creditors are less likely to request a loan if they already have a bad credit history, a plausible affirmation as they already know their financial capabilities or even have a history with crediting, thus suggesting the seniors that previously had difficulties in paying back in the past will not become new creditors.

Going further with the analysis of another sensitive attribute, we don't necessarily see a different distribution when comparing males and females' creditors, suggesting that crediting products are accessed by people based on needs, and in a developed society these needs are not immensely different between genders.

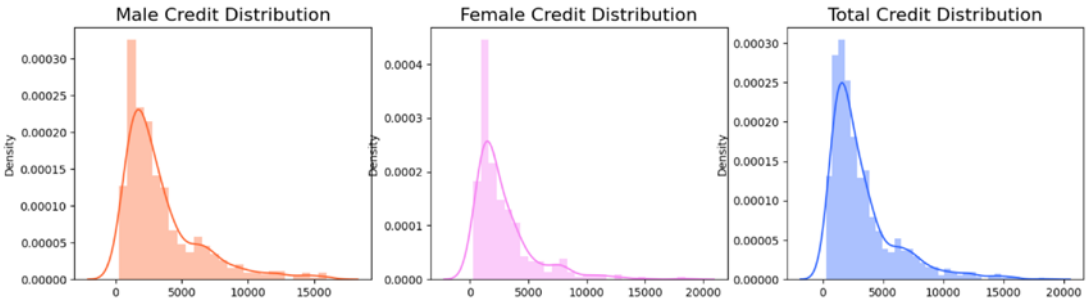
Figure 4. Distribution of gender



Source: own processing in Python.

With this new split, we can observe that around 50% of women are aged 20 to 30 years while the male has a more evenly distributed population.
When looking at the credit distribution based on gender, we can see an interesting spike on low amounts credits, with women having over 40% low amounts loans while the men had around 35%.

Figure 5. Distribution of credit split by gender



Source: own processing in Python.

Another interesting factor in crediting scoring was the purpose of the credit, with options ranging from vacation payment and domestic appliances to buying cars, furniture or even telecom equipment or paying for education or repairs.

Figure 6. Credit amount split by loan purpose

Source: own processing in Python.

The credited sums are for cars buying, furniture and housing equipment as well as technologic devices and business covering expanses.

For this study, the XGBoost algorithm was selected due to its strong predictive performance and its ability to handle datasets with both numerical and categorical features. It also integrates well with SHAP (SHapley Additive exPlanations), which provides insight into how each input feature contributes to a given prediction. The analysis was conducted using Python, with libraries such as xgboost, shap, pandas, and visualization tools like matplotlib and seaborn.

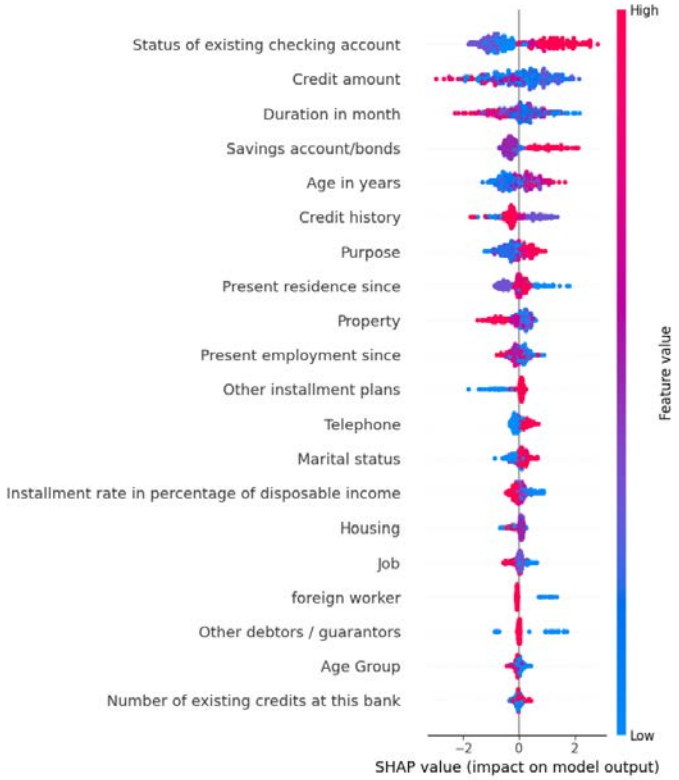
The training strategy was structured around several models. The first model was trained on the full dataset to serve as a baseline. To explore potential disparities, the data was then split by gender, creating separate models for male and female groups. Additionally, three separate models were developed based on age categories, young, adult, and senior, to allow a more granular analysis of how age-related factors might influence the model's predictions.

In the overall model, several features stood out as having a significant impact on the predicted creditworthiness of individuals. As presented in Figure 7, one of the strongest indicators was the status of the existing checking account. Individuals with a healthy, positive account balance were more likely to be classified as having good credit, while those with no account or a negative balance tended to be associated with a higher likelihood of bad credit.

Similarly, the credit amount played an important role. Lower loan amounts were generally linked to better credit scores, suggesting that smaller credit obligations may signal more responsible financial behaviour. The duration of the loan in months also influenced the predictions: shorter durations were associated with a higher chance of being labelled as good credit, likely reflecting lower risk over time.

The presence of savings or bonds contributed positively as well. Individuals with higher balances in their savings accounts or investment bonds were more likely to be predicted as good creditors, indicating that saving behaviour is seen as a strong financial indicator. Finally, age emerged as a notable factor. Older clients were more frequently associated with good credit predictions, whereas younger individuals tended to be predicted as having a higher credit risk.

Figure 7. SHAP plot of the general model

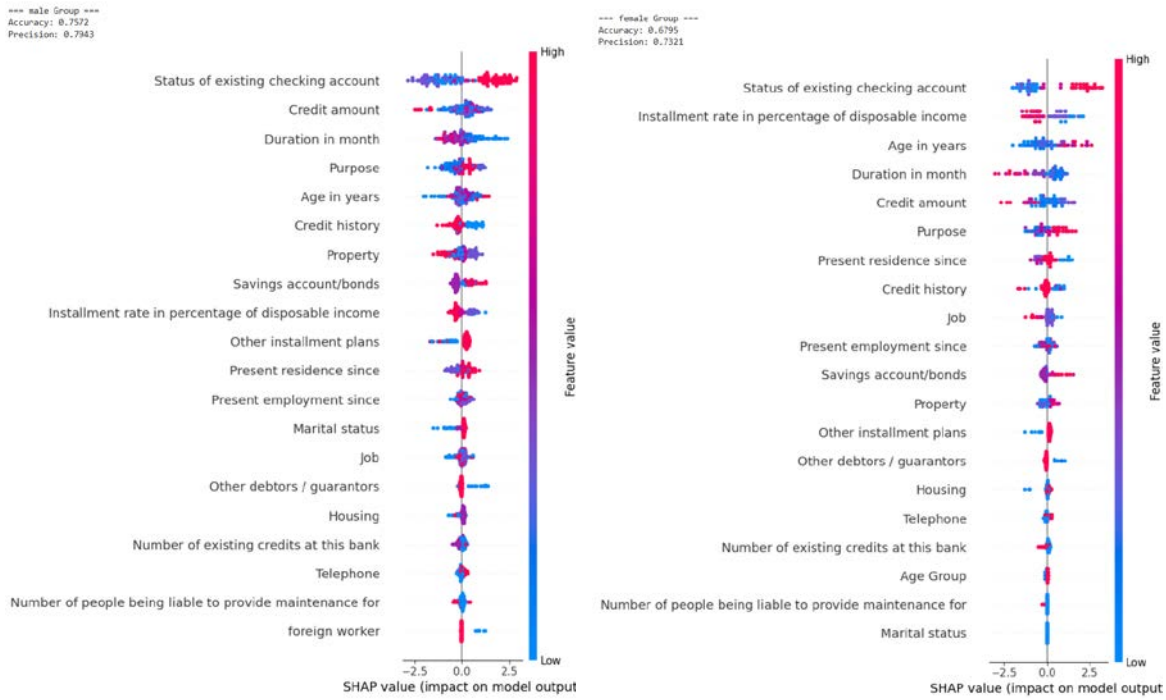


Source: own processing in Python.

For the next iteration, we split the database taking into account the two genders available.

While examining the models separately by gender, distinct patterns emerged in terms of feature importance. For female clients, disposable income had a noticeably stronger influence on the model's predictions. Women with higher disposable income were more likely to be classified as good creditors, suggesting that income stability plays a more central role in assessing creditworthiness for women. Age also proved to be a more significant predictor for women than for men, indicating that the model relies more heavily on age related trends when evaluating female credit applicants. In contrast, the purpose of the loan appeared to carry more weight in predictions for male clients, implying that the reason for seeking credit was more relevant in determining outcomes for men.

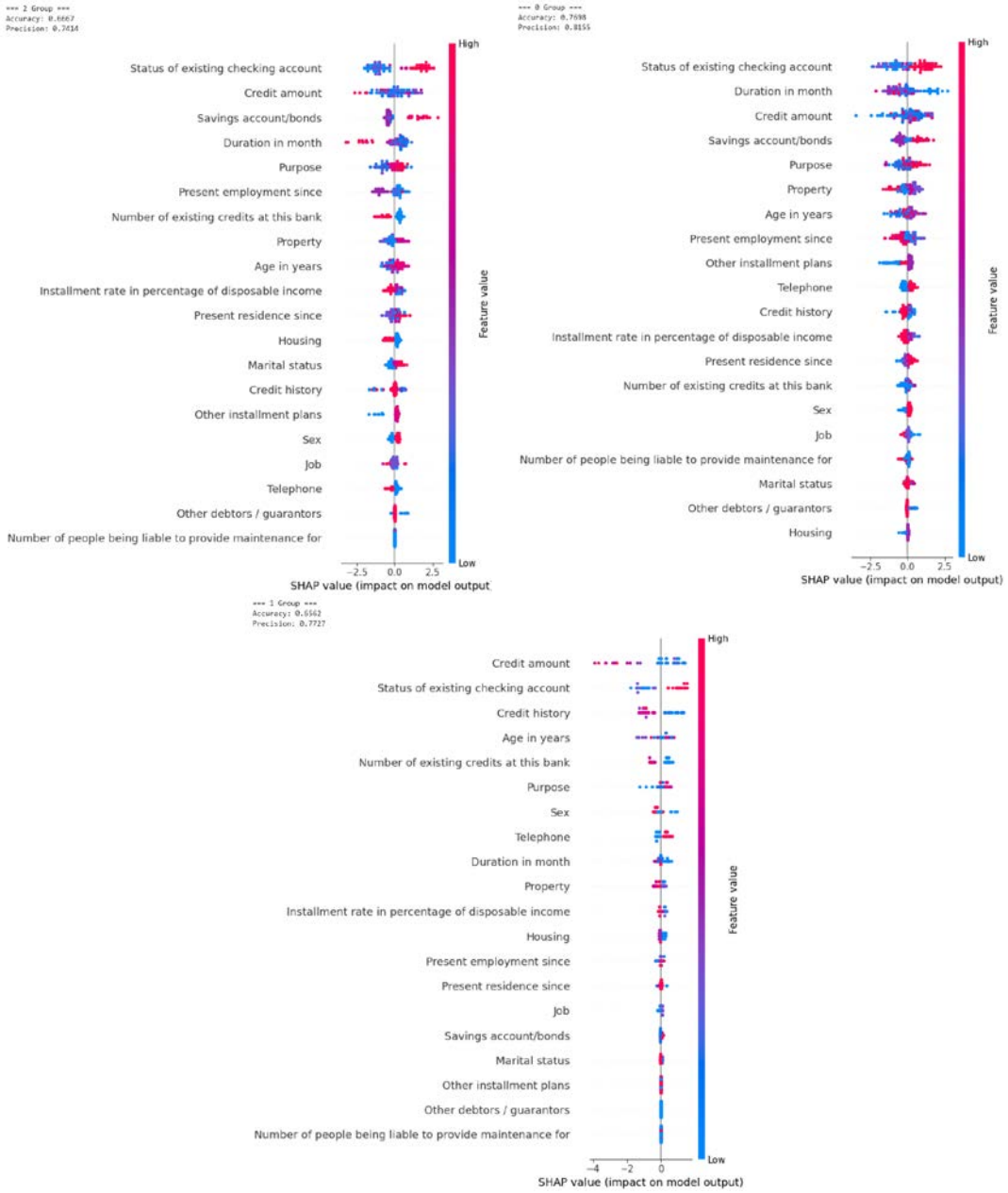
Figure 8. SHAP plot of the two models split by gender (male left, female right)



Source: own processing in Python.

When analyzing the models based on different age groups, some clear distinctions in predictive factors emerged. For the senior group, age itself became one of the most influential variables in determining creditworthiness. However, the most critical factor overall for this group remained the credit amount, indicating that the size of the loan continues to play a central role in the model's decision-making. A comparison between the young and adult groups revealed a notable shift in feature importance. For younger individuals, the credit amount carried more weight in predicting outcomes, suggesting that smaller or more manageable loans are seen as indicators of lower risk. In contrast, for adults, the duration of the credit agreement became more relevant, implying that the length of the repayment period was a stronger signal of reliability in this age group.

Figure 9. SHAP plot of the three models split by age group (Young left, Adult right, Senior bottom)



Source: own processing in Python.

The results highlight that creditworthiness predictions are shaped not only by financial variables, but also by demographic differences, with gender and age influencing which factors the model prioritizes.

5. Conclusion

This study explored the ethical implications of AI in making financial decisions with the use of the Bank Marketing dataset and the combination of XGBoost modeling and SHA. Despite the predictive accuracy of the model, our study discovered that even datasets with personal and sensitive data can initiate biases, particularly on demographic bases like age and gender. These biases tend to matter more in how different characteristics influenced predictions in different groups, like how income and age were more influential for women, but loan purpose was more predictive for men.

The transparency that can be provided using SHAP allows us to see the contribution of each variable to individual predictions, which gives a better look into the internal reasoning of the model, as

previously described by Szepannek & Lübke (2021) and de Castro Vieira et al. (2025). This interpretability is not just crucial for technical validation but also for rendering systems accountable where they have an impact on people's access to financial services.

Economically, when women or younger individuals are systematically disadvantaged by such models, access to credit becomes restricted - a pattern also identified by Lu and Calabrese (2023) in their analysis of biased lending systems. This can perpetuate economic inequality and limit upward mobility in the long term, which goes against the inclusive purpose that AI is meant to enable.

The key points from our research are that bias may be present even in freely available databases, that methods like SHAP for explainable AI are effective at highlighting this type of bias; and that demographic factors do affect model behaviour in a significant manner.

Following these results, we recommend incorporating fairness audits throughout the model development lifecycle, revisiting the use of sensitive or proxy variables, and exploring the implementation of fairness-aware machine learning models. Such measures can promote more transparent and equitable AI systems, as also emphasized by (Purificato, Lorenzo, Fallucchi, & Luca, 2022).

While this study provides valuable insights into the ethical implications of AI in financial decision-making, several limitations should be acknowledged. The analysis relies on a single publicly available dataset, which may not fully capture the complexity and diversity of real-world financial behaviours. In addition, the study focuses on one specific modeling approach, XGBoost combined with SHAP which, although effective, may not generalize to all machine learning frameworks.

For future work, we suggest reanalysing larger and more representative datasets that better reflect actual economic conditions to demonstrate the real impact of historical data bias. Expanding the interpretability toolkit beyond SHAP, by incorporating methods such as LIME or counterfactual explanations, could uncover additional patterns of bias and alternative reasoning pathways within models. Ultimately, further research should also examine legal and regulatory frameworks to guide the ethical and responsible application of AI, especially in areas where its influence on economic opportunity is most significant.

***Authors' contribution**: Introduction, B.M, G.I.G; Literature review, B.M, G.I.G; Methodology and data, B.M; Research results and comments, B.M, G.I.G; Conclusion, B.M, G.I.G*

***Funding:** This paper was co-financed by The Bucharest University of Economic Studies during the PhD program and was funded by the EU's NextGenerationEU instrument through the National Recovery and Resilience Plan of Romania - Pillar PNRR-III-C9-2022 – I8, managed by the Ministry of Research, Innovation and Digitalization, within the project entitled „CauseFinder: Causality in the Era of Big Data and AI and its applications to innovation management", contract no. 760049/23.05.2023, code CF 268/29.11.2022.*

References

Abis, S., & Veldkamp, L. (2023). The Changing Economics of Knowledge Production. *The Review of Financial Studies, 37*(1), 89-118.

Babaei, G., & Giudici, P. (2024). How fair is machine learning in credit lending? *Quality and Reliability Engineering International, 40*(6), 3452-3464.

Bircan, T., & Özbilgin, M. F. (2025). Unmasking inequalities of the code: Disentangling the nexus of AI and inequality. *Technological Forecasting and Social Change, 211,* 123925. https://doi.org/10.1016/j.techfore.2024.123925.

Chen, Y., Giudici, P., Liu, K., & Raffinetti, E. (2024). Measuring fairness in credit ratings. *Expert Systems with Applications, Volume 258,* 125184. https://doi.org/10.1016/j.eswa.2024.125184.

de Castro Vieira, J. R., Barboza, F., Cajueiro, D., & Kimura, H. (2025). Towards Fair AI: Mitigating Bias in Credit Decisions - A Systematic Literature Review. *Journal of Risk and Financial Management, 18*(5), 228. https://doi.org/10.3390/jrfm18050228.

Hassani, B. (2021). Societal bias reinforcement through machine learning: a credit scoring perspective. *AI Ethics 1*(3), 239-247.

International Monetary Fund. (2024). High uncertainty and the unknown. International Monetary Fund. *https://www.imf.org/external/pubs/ft/ar/2024/in-focus/high-uncertainty-and-the-unknown/*.

Jammalamadaka, K. R., & Itapu, S. (2022). Responsible AI in automated credit scoring systems. *AI and Ethics, 3*(2), 485–495. http://dx.doi.org/10.1007/s43681-022-00175-3.

Kozodoi, N., Jacob, J., & Lessmann, S. (2022). Fairness in credit scoring: Assessment, implementation and profit implications. European Journal of Operational Research, 297(3), 1083-1094. https://doi.org/10.1016/j.ejor.2021.06.023.

Lu, X., & Calabrese, R. (2023). The Cohort Shapley value to measure fairness in financing small and medium enterprises in the UK. *Finance Research Letters, 58*. https://doi.org/10.1016/j.frl.2023.104542.

O'Neil, C. (2016). *Weapons of Math Destruction.* Crown, New York.

Purificato, E., Lorenzo, F., Fallucchi, F., & Luca, E. W. (2022). The Use of Responsible Artificial Intelligence Techniques in the Context of Loan Approval Processes. *International Journal of Human–Computer Interaction, 39*(7), 1543-1562. https://doi.org/10.1080/10447318.2022.2081284.

Szepannek, G., & Lübke, K. (2021). Facing the challenges of developing fair risk scoring models. *Frontiers in artificial intelligence, 4*, 681915. https://doi.org/10.3389/frai.2021.681915.

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law. *West Virginia Law Review, 123*(3). https://researchrepository.wvu.edu/wvlr/vol123/iss3/4.