Rethinking gender scoring to gain fairer creditworthiness assessments

Ethics & Trust in Finance
Global edition 2022-2023

First Prize

Josiah Wamwere-Njoroge
Kenya
Postgraduate Student, Kenya School of Law, Nairobi (Kenya)*

* The views expressed herein are those of the authors and do not necessarily reflect those of the Organization he is affiliated with or of the Jury.

Introduction

In this era of digital transformation, societies find themselves at the intersection of ethics, finance, and technology (Davis, Kumiega & Vliet, 2013). As we embrace the digital revolution, one theme that demands immediate attention is the ethical implications surrounding the algorithmic scoring of consumers in creditworthiness assessment procedures. In the realm of finance, where numbers and calculations reign supreme, we often assume that decisions are made objectively, devoid of any biases or discriminatory practices. However, the emergence of algorithmic decision-making in creditworthiness assessment procedures has challenged this assumption, revealing a world where gender-based algorithmic scoring can have profound implications for consumers.

Automated decision-making has the potential to bestow countless benefits on society, empowering financial institutions to make informed decisions based on vast amounts of data. Yet it also carries the inherent risk of undermining people's rights and freedoms. Discrimination, disguised within the algorithms, can silently permeate creditworthiness assessments, leading to unjustified denials of services and goods.

In developed economies, regulatory frameworks have been established to safeguard against the utilization of certain data types in credit risk analysis. For instance, in the US, regulations prohibit the inclusion of race data and zip code data, while the UK protects category data (OECD, 2021, p. 45). Additionally, the US Equal Credit Opportunity Act (ECOA) of 1974, along with its amendments in 1976, have played a significant role in shaping credit practices by addressing issues of discrimination in credit access. (Bumacov, Ashta & Singh, 2017, p. 549). By explicitly prohibiting discrimination based on factors such as race, religion, gender, marital status, and age, the ECOA aimed to foster fair and unbiased lending practices. In the EU, the use of gender in decision-making processes is also explicitly prohibited by law. This prohibition stems from the broader framework of antidiscrimination provisions that aim to ensure equal treatment in various aspects, including goods and services (Andreeva & Matuszyk, 2019, p. 1288). One of the key legislative measures in this regard is the European Equal Treatment in Goods and Services Directive (2004/113/EC), which was enacted by the European Council in 2004.

The prohibition of gender scoring, which refers to the exclusion of borrower gender data in credit-scoring models, plays a crucial role in the fight against discrimination. It sends a powerful message that gender should never be a determining factor in creditworthiness assessments. If algorithmic systems are allowed to utilize gender as a determining factor, they risk reinforcing regressive stereotypes and exacerbating gender disparities in economic opportunities. This would not only undermine social progress but also inhibit economic growth by excluding talented individuals from accessing the resources they need to thrive. By eliminating gender as a scoring factor, we lay the foundation for a more inclusive and thriving society, where talent and potential can flourish without unjust impediments.

Yet it is crucial to acknowledge that banning gender scoring may not be a foolproof solution to ensure non-discrimination by algorithms (Anderson, 2022, p. 39). A counterintuitive consequence of the prohibition of gender scoring is the potential negative impact on gender equality. Research has suggested that eliminating gender scoring could inadvertently result in a decrease in credit access for creditworthy women (Bostic & Calem, 2003). This unintended consequence is contrary to the very principles we seek to uphold—equality and empowerment. As a result, a critical question emerges: How can we strike a balance between eradicating discrimination and ensuring equal opportunities for all?

At the heart of this essay lies a fundamental inquiry: Is there a more effective approach to tackling discrimination in algorithmic credit lending decisions than solely relying on the prohibition of gender scoring? There is a pressing need to find a better approach, one that transcends the mere prohibition of gender scoring and addresses the nuances of non-discrimination in algorithmic decision-making for credit lending. The essay proposes a discrimination definition that centers on balancing error probabilities within the algorithmic system. In simpler terms, the goal is to ensure that the algorithm's errors, both false positives and false negatives, are distributed equally between genders.

This perspective challenges the conventional belief that the prohibition of gender scoring alone is sufficient to eliminate discrimination. The proposed approach emphasizes the evaluation of error distribution as a means to comprehensively assess the fairness of algorithmic decisions. Central to this approach is the pursuit of parity in the probabilities of positive credit decisions for creditworthy individuals, irrespective of their gender. A more equitable system can begin to take shape when the algorithm demonstrates an equal likelihood of approving creditworthy men and women. Furthermore, it is imperative to ensure equal probabilities for non-creditworthy individuals, both women and men, in receiving negative credit decisions.

The implications of this analysis extend beyond theoretical discussions. They have real-world consequences for individuals seeking fair access to credit and financial opportunities. By redefining how we evaluate discrimination in algorithmic decision-making, we can pave the way for more inclusive and equitable credit lending practices.

The subsequent sections of this essay unfold as follows: Section 1 undertakes an examination of the immense influence and application of algorithms in creditworthiness assessments. Building upon this foundation, Section 2 delves into the complex challenges posed by the exclusion of gender as a scoring factor in algorithmic credit scoring, accompanied by a proposal that advocates for a comprehensive framework aimed at ensuring equitable and impartial outcomes in creditworthiness assessment algorithms. Subsequently, Section 3 discusses the multifaceted role of gender within algorithmic credit scoring, unraveling its intricacies and uncovering its significance in the broader context of credit evaluation. Finally, the concluding section provides a concise summary and synthesis of the key findings and implications derived from this analysis.

Algorithmic Decision-Making in Creditworthiness Assessment Procedures

The term "algorithmic decision-making" itself evokes a sense of mystery and intrigue, conjuring visions of complex codes and digital marvels. But what does it truly mean in the context of creditworthiness assessment procedures?

Cormen et al. (2022) define an algorithm as "a sequence of computational steps that transform the input into the output" (Cormen et al., 2022, p. 5). They therefore serve as guiding principles that enable computers to make decisions and execute tasks. However, algorithmic decision-making extends beyond the realm of rule-based automation alone (Kerrigan, 2022, p. 35). The applications of technology are expanding towards more autonomous systems, blurring the boundaries between rule-based and AI-based solutions (Fishman & Stryker, 2020, p. 15). Thus, the term "algorithmic decision-making" serves as a useful umbrella term that encompasses both rule-based and AI-based approaches.

While this essay primarily focuses on rule-based automation in the context of gender-based algorithmic scoring, the principles presented may also be applicable to AI-based algorithms. The fundamental goal of ensuring non-discrimination and fairness in algorithmic decision-making transcends the specific technological implementation. Whether the decision-making process is rule-based or AI-based, the ethical and trust-related challenges remain pertinent.

The Power of Algorithms

At the heart of algorithmic decision-making lies the pursuit of efficiency and data-driven insights (Werbach & Cornell, 2021, p. 37). These systems can incorporate an unprecedented amount of empirical data, surpassing the limits of human cognition and processing power (Portuese, 2022, p. 5). By automating the credit application process, these algorithms navigate the labyrinthine landscape of financial assessments, evaluating factors such as income, payment history, and outstanding debts with remarkable speed and precision (Nalic & Martinovic, 2020).

In this domain, the role of human decision-makers diminishes, making way for the silent wisdom of algorithms. But what does this mean for the consumers who seek financial support? How do these algorithmic systems assess their creditworthiness?

The Complexities of Creditworthiness Assessment: Navigating Human Behavior, Financial Factors, and Demographic Characteristics

Within the framework of a loan agreement, the temporal dimension emerges as a distinguishing factor. Unlike other types of transactions, lending entails a temporal discrepancy between the disbursement of funds and their subsequent repayment (Turnbull, 1998, p. 343). This inherent characteristic of loans necessitates a meticulous assessment of the borrower's dependability, ensuring that lending institutions can have confidence in the borrower's ability to fulfil the repayment obligations according to the predetermined terms (Genberg, 2020, p. 71) The evaluation of creditworthiness assumes paramount significance in this context, as it determines whether an applicant, who seeks credit, possesses the necessary creditworthiness to justify the requested loan amount (Anderson, 2022, p. 2). Essentially, this evaluation endeavours to address a fundamental question: Can the borrower be reasonably anticipated to uphold their financial commitments and fulfill the loan repayment obligations in a timely manner?

However, the assessment of creditworthiness is not a straightforward task. It requires navigating the complexities of human behavior, financial circumstances, and the ever-changing dynamics of life. Predicting the future with absolute certainty is impossible, leading lenders to rely on tools such as inductive reasoning, profiling and credit scoring to make informed decisions (Hallinan & Borgesius, 2020, p. 9). These tools, when used responsibly, can provide valuable insights into a person's creditworthiness, providing lenders with a consistent and standardized approach to assess creditworthiness.

As highlighted by Kamp, Körffer & Meints (2008), the parameters employed in credit scoring can be categorized into three main groups: contract-related parameters, financial criteria, and demographic characteristics (Kamp, Körffer & Meints, 2008, p. 207). Each of these categories provides insights into an applicant's creditworthiness.

Contract-related parameters encompass factors such as the number of credit cards and loans, as well as the duration of previous contractual relationships between the borrower and lenders (Vercammen, 1995). These parameters offer insights into the borrower's financial history and behavior, serving as indicators of their ability to meet their financial obligations. Financial criteria, on the other hand, delve into an applicant's assets, income, and expenses, providing an understanding of their financial capacity to repay a loan (Cowan & De Gregorio, 2003, p. 165). These parameters form the backbone of creditworthiness evaluation, enabling lenders to assess an individual's financial stability.

However, it is the third category of parameters, namely demographic characteristics, that has garnered significant attention, particularly concerning gender-based algorithmic scoring. Alongside gender, variables such as address, age, number of children, residential area, education, occupation, employer name, nationality, and religion have historically been used in credit scoring models (Anderson, 2022, p. 84).

Lenders do not possess unfettered discretion when it comes to determining the information they utilize in their creditworthiness assessments. The fundamental principles of the rule of law demand that decisions be rooted in accurate and relevant information, tailored to each unique situation. Thus, the question arises: how do we discern which factors hold relevance in specific circumstances?

One approach is to consider variables that have demonstrated their ability to predict loan repayment with a reasonable level of certainty (Abdullah et al., 2020, p. 81.). These variables, backed by empirical evidence, can provide valuable insights into an applicant's creditworthiness. By relying on well-established indicators, lenders can minimize the risk of making erroneous judgments and ensure that their assessments are grounded in a reliable foundation.

This is where algorithmic scoring systems come into play. By harnessing the power of data and automation, algorithms aim to streamline the creditworthiness assessment process, bringing efficiency, consistency, and objectivity to the table (Genberg, 2020, p. 71). In recent years, algorithmic decision-making has gained traction as a tool to streamline and automate creditworthiness evaluations. Algorithms, driven by vast amounts of data and intricate calculations, analyze numerous attributes to generate a credit score for each applicant. These attributes can include income, employment history, debt-to-income ratio, and even personal characteristics such as gender. It is here that the potential for bias and discrimination comes to light.

Beyond Gender Exclusion: A More Comprehensive Approach to Address Bias in Algorithmic Credit Scoring

Various approaches have been proposed to mitigate algorithmic discrimination. One such approach, as suggested by Lepri et al. (2018) involves avoiding the use of sensitive attributes, like gender, in the decision-making process. The rationale behind this approach is to eliminate direct gender-based scoring, thus reducing the potential for discriminatory outcomes (Lepri et al., 2018, pp. 615-618).

Unravelling Algorithmic Discrimination: The Challenges of Excluding Gender as a Scoring Factor

At first glance, it may seem like a straightforward solution. By refraining from considering gender as a scoring factor, the algorithm's decisions would be free from explicit gender biases. However, the complexities of the issue become apparent when we realize that the impact of gender can still permeate the algorithm through its correlation with other variables used in the scoring process (Anderson, 2022, p. 39).

To illustrate this, consider a hypothetical scenario where gender is excluded as a direct scoring factor. However, if other variables that are correlated with gender, such as occupation or educational background, are still taken into account, the algorithm may indirectly incorporate gender biases.

This is because certain occupations or educational paths may have historically favored or disadvantaged individuals of a particular gender. As a result, the algorithm might inadvertently perpetuate gender-based inequalities, even without explicitly considering gender as a scoring attribute. For example, if historically more women have worked in lower-paying occupations or have had limited access to higher education, these factors might indirectly affect their credit scores. In such cases, the exclusion of gender as a direct factor does not fully eliminate the influence of gender on the algorithm's decisions.

Additionally, removing gender-based scoring from creditworthiness assessments brings forth another complex dilemma as lenders may struggle to maintain the desired level of overall risk assessment. However, research suggests that this decline in accuracy is unlikely to be significant in the lending context (Andreeva & Matuszyk, 2019, p. 1287). Lenders have access to a wealth of other variables that can effectively gauge an applicant's creditworthiness and maintain the desired risk levels. However, while the overall accuracy of credit decisions may remain intact, the concern lies in the potential uneven distribution of errors between genders. Without gender-based scoring, there is a risk that the algorithm may disproportionately make incorrect decisions for one gender over the other. This raises questions about fairness and equality in the creditworthiness assessment procedure.

For instance, studies have indicated that if gender-based scoring is removed, the algorithm may struggle to effectively identify creditworthy women compared to when gender is considered (Andreeva & Matuszyk, 2019). This disparity could result in qualified and deserving women being unjustly denied credit opportunities, undermining their financial prospects and perpetuating gender-based inequalities. Conversely, there is also the issue of the algorithm potentially failing to identify non-creditworthy men as frequently as it does non-creditworthy women. This imbalance in identifying risk may lead to an increased likelihood of lending to individuals who are not capable of repayment, thus exposing lenders to higher default rates and financial losses.

Striking a Balance: Fairness and Accuracy in Creditworthiness Assessment Algorithms

To tackle this challenge effectively, we need to explore more comprehensive approaches that go beyond the exclusion of sensitive attributes. It requires a deeper understanding of how correlations between variables can introduce biases and discriminatory outcomes. By identifying and addressing these underlying correlations, we can work towards creating fair and unbiased algorithms that truly uphold the principles of ethics and trust in finance.

While examining the input variables used in algorithms is one approach, it may not be sufficient on its own. A promising approach, as suggested by Lepri et al. (2018), involves focusing on the concept of statistical parity (Lepri et al., 2018, p. 616). The idea behind statistical parity is to ensure that the probability of every possible decision outcome is equal across all groups, regardless of sensitive attributes like gender. However, as Lepri et al. (2018) acknowledge, implementing statistical parity in certain contexts, such as lending, may pose challenges. If different groups, such as women and men, exhibit variations in their historical loan repayment behavior, enforcing statistical parity could potentially compromise the accuracy of algorithmic decisions (Lepri et al., 2018, p. 616). Striking a balance between fairness and accuracy becomes a delicate task.

Kleinberg, Mullainathan, and Raghavan (2016) shed more light on the complexities surrounding the establishment of equitable risk scores. These researchers encapsulate the viewpoints presented in existing literature by delineating three fundamental conditions that are crucial in ensuring fairness within algorithmic systems.

The first condition pertains to the calibration of the algorithm, encompassing the idea that if the algorithm assigns a particular characteristic (such as creditworthiness for a specific loan amount) to a certain group (for instance, women) with a probability denoted as x, then x proportion of individuals within that group should indeed possess that characteristic (Kleinberg, Mullainathan, and Raghavan, 2016, p. 2). Essentially, the algorithm's claims must align with the actual distribution of traits within the group. This condition emphasizes the need for transparency and accuracy, where the scores themselves convey their intended meaning without discrepancy when examined within each group.

The second requirement revolves around achieving equal average scoring across all groups for individuals with a specific characteristic, such as creditworthiness (Kleinberg, Mullainathan, and Raghavan, 2016, p. 2). Referred to as *the balance for the positive class* condition, it would require, for example, that the average score for creditworthy women should be equivalent to the average score for creditworthy men. This condition seeks parity in the treatment of individuals with similar traits, regardless of their gender. Consequently, it aims to eliminate any bias or disparity that may arise from the algorithmic decision-making process.

Correspondingly, the third condition, known as *the balance for the negative class* requirement, focuses on maintaining consistency in average scoring across all groups, even for individuals who do not possess the aforementioned characteristics (Kleinberg, Mullainathan, and Raghavan, 2016, p. 2). This suggests that the average score for non-creditworthy women should align with the average score for creditworthy men. By extending fairness beyond the positive class, this condition seeks to rectify any potential biases against non-creditworthy individuals based on gender.

Importantly, Kleinberg, Mullainathan, and Raghavan (2016) highlight that the positive and negative class balance requirements can be seen as extensions of the principle that the relative amounts of false positives and false negatives (i.e., Type I and Type II errors) should be equal across both groups (Kleinberg, Mullainathan, and Raghavan, 2016, p. 2). While statistical parity demands equal average scoring across all groups for all members, the focus should be on achieving balance specifically for creditworthy and non-creditworthy individuals within each group. This distinction recognizes the need to account for differences in group compositions and the subsequent impact on fairness considerations.

Nevertheless, it is worth noting that fulfilling all three conditions simultaneously is often a daunting task. Trade-offs and choices inevitably arise, requiring a prioritization of which condition holds greater significance in a given context (Gandy, 2010, p. 39). Altman, Wood and Vayena (2018), in line with these considerations, argue that unless the probabilities of error are explicitly calculated for each group separately (such as for women and men), differences are likely to emerge (Altman, Wood and Vayena, 2018, p. 16). Implicitly or explicitly, the creators of the algorithm must deliberate on which types of errors should be minimized and which groups should be assessed with greater accuracy (Altman, Wood and Vayena, 2018, p. 17).

This essay centers on an examination of the discriminatory elements inherent in algorithms, employing a lens shaped by the two balance requirements outlined by Kleinberg, Mullainathan, and Raghavan (2016). These balance requirements specifically relate to the distribution of Type I and Type II errors. In the subsequent section, a detailed exposition will be presented, clarifying the precise nature and implications of Type I and Type II errors.

Exploring the Complexities of Type I and Type II Errors in Algorithmic Decision-Making

The assessment of creditworthiness is a binary classification task (Zliobaite, 2015, p. 6). In its simplest form, applicants are categorized into two classes:

creditworthy or non-creditworthy, depending on their perceived ability to repay a loan. This classification becomes the pivotal point where the algorithm steps in, armed with its complex calculations and predictive powers. The algorithm's goal is to make accurate decisions, minimizing the chances of errors that could lead to financial repercussions.

Type I and Type II errors serve as vital indicators of the fairness and accuracy of algorithmic systems. Type I errors, often referred to as false positives, occur when the algorithm wrongly identifies an individual as possessing a specific characteristic or falling into a particular category (Florez-Lopez, 2010, p. 494). In the context of creditworthiness assessment, a Type I error would mean labelling an individual as creditworthy when in reality they are not likely to repay a loan on time (Li and Zhong, 2012, p. 187). On the other hand, Type II errors, known as false negatives, happen when the algorithm incorrectly fails to recognize a characteristic or category that an individual genuinely possesses (Florez-Lopez, 2010, p. 494). For instance, a Type II error might involve deeming an individual as non-creditworthy, despite their high likelihood of timely loan repayment (Kern, 2017, p. 4).

By examining the distribution of these errors, we gain insights into the potential biases that algorithms can perpetuate. If there is an imbalance in the distribution of errors between different groups, such as women and men, it raises concerns of fairness and discrimination. For instance, if the algorithm consistently makes more Type I errors for women compared to men, it implies that women might face unjust disadvantages in accessing credit, despite their creditworthiness. Similarly, if the algorithm commits more Type II errors for men, it suggests that men might encounter challenges in obtaining credit even when they are creditworthy.

From Bias to Balance: Measuring the Performance of the Algorithm

Binary classification tasks such as creditworthiness assessment can be evaluated using various metrics to gauge their success and identify potential errors. Sensitivity and specificity are two key metrics that provide valuable insights into the algorithm's performance and its ability to accurately classify applicants (Sharma, Yadav & Sharma, 2009, p. 53).

Sensitivity, also known as recall or true positive rate, provides insights into how well the evaluation method correctly identifies creditworthy applicants when they are indeed creditworthy. It answers the question: How often does the system make the right call? To obtain the sensitivity value, we divide the number of true positive assessments (correctly identifying creditworthy applicants) by the sum of true positive and false negative assessments. High sensitivity indicates that the algorithm correctly identifies a large proportion of creditworthy applicants, minimizing the chances of Type II errors—incorrectly classifying a creditworthy applicant as non-creditworthy (Sharma, Yadav & Sharma, 2009, p. 58). In essence, high sensitivity suggests that the algorithm is doing a good job of capturing and acknowledging the creditworthiness of individuals.

It is imperative to establish a robust and reliable assessment process that incorporates both sensitivity and effectiveness in differentiating creditworthy individuals from non-creditworthy ones. Merely prioritizing high sensitivity is insufficient to ensure the accuracy and success of the evaluation process. What we truly require are algorithms that strike a nuanced equilibrium, accurately identifying applicants who are creditworthy while also effectively recognizing those who may present a potential risk. This delicate equilibrium is the cornerstone of responsible lending practices, enabling the efficient allocation of resources and fostering the sustainability of the financial system.

To achieve this delicate balance, algorithms must take into account a range of factors beyond sensitivity alone. Specificity, a complementary measure to sensitivity, assumes a vital role in this equation. Also known as the true negative rate, specificity measures the algorithm's ability to correctly identify non-creditworthy applicants (Sharma, Yadav & Sharma, 2009, p. 58). It ensures that the evaluation process is not overly lenient, shielding non-creditworthy individuals from undue risks. By dividing the number of true negative assessments (correctly identifying non-creditworthy applicants) by the sum of true negative and false positive assessments, we can gauge the specificity of the algorithm.

To maintain the integrity of the creditworthiness assessment procedure, it is essential to strike a balance between sensitivity and specificity. A high sensitivity ensures that deserving creditworthy individuals are not unjustly denied opportunities, while a high specificity safeguards against the indiscriminate approval of non-creditworthy applicants (Sharma, Yadav & Sharma, 2009, p. 55). The interplay between these two metrics establishes a robust and reliable evaluation process—one that can be trusted by both lenders and applicants alike.

Dissecting Gender Disparities in Credit Repayment: Unravelling the Role of Gender in Algorithmic Scoring

The main premises in this essay are that gender serves as a predictor of credit repayment probability, and that the removal of gender scoring would introduce gender-specific implications concerning the occurrence of Type I and Type II errors. The ensuing discussion offers a succinct justification for these underlying assumptions.

Gender-based algorithmic scoring presupposes that certain characteristics associated with gender—whether cultural, social, or economic—play a role in determining an individual's creditworthiness. Over the years, various studies and investigations have shed light on the disparities in payment behaviors between genders. These findings consistently highlight that men tend to have a higher incidence of payment defaults compared to women (Guérin et al., 2011, p. 8; Karlan & Zinman, 2009; Majamaa, Lehtinen & Rantala, 2019, p. 236). Such empirical evidence underscores the importance of understanding these differences and their potential impact on creditworthiness evaluations.

It is therefore crucial to unravel the complex web of factors that contribute to this association and explore whether it solely reflects men's poorer creditworthiness or stems from other underlying dynamics. This association can be attributed not necessarily to men's inherent creditworthiness deficit, but rather to factors such as their potentially greater utilization of credit services in comparison with women. This difference in credit usage patterns can contribute to their overrepresentation in indebtedness datasets. It does not necessarily imply that men are inherently less creditworthy, but rather that their credit behavior and utilization may differ from that of women. When individuals have a greater reliance on credit, it naturally follows that instances of repayment difficulties may also be more prevalent. Nonetheless, considering these factors, it can be posited that the overrepresentation of men in indebtedness datasets provides some indication that gender is potentially associated with creditworthiness, particularly within the subset of individuals who actively seek credit.

When discussing gender differences, even at their best, these differences are typically average and often have very small effect sizes. Individuals within each gender span a wide spectrum of behaviors and characteristics. However, removing the gender variable from creditworthiness assessment can present challenges, even if the repayment probability is identical for both women and men. To illustrate this point, consider the hypothetical scenario presented by Elliehausen & Durkin (1989) within the credit card market where an equal distribution of individuals with exceptional creditworthiness exists among both women and men (Elliehausen & Durkin, 1989, p. 100). Moreover, the hypothetical assumes that the lender has identified years of employment as a reliable indicator of creditworthiness. However, an intriguing observation emerges as it becomes apparent that creditworthy women, despite possessing similar creditworthiness to their male counterparts, tend on average to have fewer years of employment.

In this scenario, the number of years of employment serves as a signaling variable for creditworthiness. However, the informational value of this variable differs between genders. A lower number of years of employment for women could indicate the same level of creditworthiness as a higher number of years for men. Thus, the assessment of creditworthiness would become dependent on gender in addition to the signaling variable, even if the predicted variable (average creditworthiness) is identical between genders. Consequently, removing the gender variable from the equation could potentially result in creditworthy women facing greater challenges in obtaining credit compared with equally creditworthy men.

These considerations form the basis of the underlying assumption in this essay: achieving a balance between Type I and Type II errors in lending would be difficult, if not impossible, without the inclusion of a gender variable. It is crucial to understand the implications of this assumption while interpreting research results and engaging in discussions surrounding gender-based algorithmic scoring in creditworthiness assessments. The inclusion of gender allows for a nuanced understanding of creditworthiness, taking into account the different contexts and characteristics that may influence repayment behavior.

Conclusion

At the outset of this essay, the following hypothesis was formulated: the prohibition of gender scoring in algorithmic decision-making may not effectively address discrimination. As the analysis progressed, this initial assumption underwent a process of refinement. It became increasingly apparent that the prohibition of gender scoring might unintentionally hinder the progress towards gender equality objectives. Consequently, there emerged a pressing need to explore an alternative framework for conceptualizing algorithmic bias, thus necessitating a fresh perspective to guide and shape the investigation.

The primary research question sought to identify a more robust approach, beyond the sole emphasis on input variables, for defining algorithmic bias in the lending context. In addressing this question, a review of the existing literature yielded a proposal for a more comprehensive approach to evaluate algorithmic bias. Rather than solely focusing on the selection of input variables, the suggested approach entails examining the distribution of error types across genders. This alternative perspective aims to encompass both the similarities and differences between genders, acknowledging the significance of equitable creditworthiness assessments for all applicants. By defining discrimination based on probabilities of errors, the objective is to ensure that evaluations are conducted with equal rigor for both women and men, thereby mitigating the disproportionate impact of adverse outcomes.

Of course, it is imperative to acknowledge the compelling arguments supporting the prohibition of gender scoring. Among these arguments is the emphasis on averting the reinforcement of gender stereotypes within society. Additionally, there exists apprehension regarding the potential perpetuation of societal disparities between different groups through accepting differentiation. Furthermore, the inherent risk of erroneous calculations underlying scoring systems cannot be overlooked. However, within the framework of this essay, if the prohibition of gender scoring and the balancing of error probabilities are regarded as mutually exclusive scenarios, the prioritization of balancing error probabilities should be regarded as paramount.•

References

Abdullah et al. (2020). Credit Risk Assessment Models of Retail Microfinancing: The Case of a Malaysian National Savings Bank's Branch. *International Journal of Financial Research*, 11, 73-83.

Altman, M., Wood, A., & Vayena, E., (2018). A Harm-Reduction Framework for Algorithmic Fairness. *IEEE Security & Privacy*, 16, 34. Berkman Klein Center Research Publication No. 2018-3, 1-20.

Anderson, R. (2022). *Credit Intelligence & Modelling: Many Paths through the Forest of Credit Rating and Scoring*. New York: Oxford University Press.

Andreeva, G., & Matuszyk, A. (2019). The Law of Equal Opportunities or Unintended Consequences?: The Effect of Unisex Risk Assessment in Consumer Credit. *Journal of the Royal Statistical Society Series A: Statistics in Society*,182, 1287–1311.

Bose, U. (2012). An Ethical Framework in Information Systems Decision Making Using Normative Theories of Business Ethics. *Ethics and Information Technology*, 14, 17-26.

Bostic, R.W, & Calem, P.S. (2003). Privacy Restrictions and the Use of Data at Credit Registries. In: M. Miller (Ed.), *Credit Reporting Systems and the International Economy* (pp. 311-334). Cambridge: The MIT Press.

Bumacov, V., Ashta A., & Singh, P. (2017). Credit Scoring: A Historic Recurrence in Microfinance. *Strategic Change*, 26, 543–554.

Cormen, T.H., Leiserson, C.E., Rivest, R.L., & Stein, C. (2022). *Introduction to Algorithms*. Cambridge: The MIT Press.

Cowan, K., & De Gregorio J. (2003). Credit Information and Market Performance: The Case of Chile. In: M. Miller (Ed.), *Credit Reporting Systems and the International Economy* (pp. 311-334). Cambridge: The MIT Press.

Davis, M., Kumiega, A., & Vliet, B. (2013). Ethics, Finance, and Automation: A Preliminary Survey of Problems in High Frequency Trading. *Science and Engineering Ethics*, 19, 851–874.

Elliehausen, G.E., & Durkin, T.A. (1989). Theory and Evidence of the Impact of Equal Credit Opportunity: An Agnostic Review of the Literature. *Journal of Financial Services Research*, 2, 89–114.

Fishman, N., & Cole, S. (2020). *Smarter Data Science: Succeeding with Enterprise-Grade Data and AI Projects*. Indiana: John Wiley & Sons.

Florez-Lopez, R. (2010). Effects of Missing Data in Credit Risk Scoring. A Comparative Analysis of Methods to Achieve Robustness in the Absence of Sufficient Data. *The Journal of the Operational Research Society*, 61, No. 3, Consumer Credit Risk Modelling; Transportation, Logistics and the Environment, pp. 486-501.

Gandy, O.H. (2010). Engaging Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems. *Ethics and Information Technology*, 12, 29–42.

Genberg H. (2020). Digital Transformation: Some Implications for Financial and Macroeconomic Stability. In: J. Beirne & D.G. Fernandez (Eds.). *Macroeconomic Stabilization in the Digital Age* (pp. 68-85). Tokyo: Asian Development Bank Institute.

Guérin, I., D'Espallier, B., & Mersland, R. (2011). Women and Repayment in Microfinance: A Global Analysis. *World Development*, 39, 758-772.

Hallinan, D., & Borgesius F.Z. (2020). Opinions Can Be Incorrect (In our Opinion)! On Data Protection Law's Accuracy Principle. *International Data Privacy Law*, 0, 1-10.

Kamp, M., Körffer, B., Meints, M. (2008). Profiling of Customers and Consumers - Customer Loyalty Programmes and Scoring Practices. In: Hildebrandt, M., Gutwirth, S. (Eds.) *Profiling the European Citizen*. Dordrecht: Springer.

Karlan, D., & Zinman, J. (2009). Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment. *Econometrica*, 77, 1993 – 2008.

Kern, A.M. (2017). Credit Score Analysis. Southern Illinois University Carbondale, OpenSIUC.

Kerrigan, C. (2022). Introduction to AI. In: C. Kerrigan (Ed.), *Artificial Intelligence: Law and Regulation* (pp. 30-36). Massachusetts: Edward Elgar.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent Trade-Offs in the Fair Determination of Risk Scores. arXiv preprint arXiv:1609.05807.

Lepri, B., et al. (2018), Fair, Transparent, and Accountable Algorithmic Decision-Making Processes. *Philosophy & Technology*, 31, 611–627.

Li, X., & Zhong, Y. (2012). An Overview of Personal Credit Scoring: Techniques and Future Work. *International Journal of Intelligence Science*, 2, 181-189.

Majamaa, K., Lehtinen, A.R. & Rantala, K. (2019). Debt Judgments as a Reflection of Consumption-Related Debt Problems. *Journal of Consumer Policy*, 42, 223–244.

Nalic J., & Martinovic, G. (2020). Building a Credit Scoring Model Based on Data Mining Approaches. *International Journal of Software Engineering and Knowledge Engineering*, 30, 147–169

OECD (2021). *OECD Business and Finance Outlook 2021: AI in Business and Finance*. Paris: OECD Publishing.

Portuese, A. (2022). Prologue: Algorithmic Antitrust—A Primer. In: A. Portuese (Ed.), *Algorithmic Antitrust* (pp. 1-38). Switzerland: Springer.

Sharma, P., Yadav, U., & Sharma, D. (2009). The Concept of Sensitivity and Specificity in Relation to Two Types of Errors and its Application in Medical Research. *Journal of Reliability and Statistical Studies*, 2, pp. 53–58.

Turnbull, S. (1998). Should ownership last forever? *The Journal of Socio-Economics*, 27, 341-363.

Vercammen, J.A. (1995). Credit Bureau Policy and Sustainable Reputation Effects in Credit Markets. *Economica,* 62, 461–478.

Werbach, K., & Cornel, N. (2021). Contracts: Ex Machina. In: M. C. Compagnucci, M. Fenwick, S. Wrbka (Eds.) *Smart Contracts: Technological, Business and Legal Perspectives* (pp. 7-28). Oxford: Bloomsbury Publishing.

White & Case (2017). Algorithms and Bias: What Lenders Need to Know. Retrieved from https://www.jdsupra.com/legalnews/algorithms-and-bias-what-lenders-need-67308/

Zliobaite, I. (2015). A Survey on Measuring Indirect Discrimination in Machine Learning. arXiv preprint arXiv:1511.00148.