**Literature Review**

Access to credit has long been recognized as a critical driver of economic growth, enabling individuals to seize opportunities for upward mobility.

Conversely, the exclusion from financial access can have profound implications, particularly for marginalized communities, leading to a perpetuation of poverty and limited avenues for progress. The detrimental effects of unequal credit access based on factors such as race and gender are especially noteworthy, as they exacerbate the existing disparities, leaving previously disadvantaged groups further crippled and deepening the gaps in welfare and opportunity (King and Levine, 1993).

The link between credit accessibility and economic growth has been a subject of interest for researchers and policymakers alike. A functioning and inclusive financial system plays a vital role in providing the necessary capital to finance entrepreneurial projects, investments in human capital, and infrastructure development, which are all instrumental in fueling economic advancement. Studies have revealed that access to credit can substantially enhance productivity, facilitate innovation, and encourage risk-taking, leading to a multiplier effect on economic growth and overall prosperity (Beck et al., 2007).

The credit allocation decision-making process involves evaluating the creditworthiness and risk profile of potential borrowers to determine whether they qualify for the requested credit and, if so, under what terms and conditions. For this purpose, several aspects are considered, and the process must comply with legal and regulatory requirements to ensure that lenders adhere to fair lending practices to prevent discrimination based on factors such as race, gender, nationality, or religion.

Over the past three decades, consumer lending has witnessed a significant shift from traditional interview-based underwriting to data-driven models for evaluating and pricing credit risk. In the early 1990s, most lenders relied on a single "house rate" and conducted borrower interviews for screening purposes (Johnson, 1992).

Soon after, processes started to become automated, and statistical methods have been used consistently in various applications. The first notable example of the statistics applied to credit risk analysis can be traced back to 1959, when Fisher built a model using ordinary least squares (OLS) to explain the determinants of a bond’s risk premium.

The introduction of credit scoring, together with the computerization of the industry, has increased the availability and affordability of credit, by allowing lenders to assess more quickly the creditworthiness of applicants. At the same time, credit scoring is also a powerful instrument that can make evaluation more objective and prevent discrimination, which has negative effects both on fairness and on economic outcomes by misallocating available resources (Avery, 2009).

Generally, it is fair to say that efforts to predict credit-related variables using statistical methods have been quite successful over the years, and with the decline in storage and computing costs, lenders progressively adopted estimates of default risk to determine individual loan prices. A major advantage of employing such techniques is that statistical methods are succinct and easy to explain, and yet bring quite satisfactory results. However, they are theoretically not suited to work with financial data, which is generally not compliant with standard assumptions, including multivariate normality assumptions (Huang et al., 2003).

More recently, technological advancements have allowed for the implementation of artificial intelligence methods capable of automatically extracting knowledge from data sets. This enables the learning of the model's specific structure directly from the data, thus increasing the explanatory power of the model itself. However, this leads to the emergence of complex and hard-to-explain models, shining a new light to the trade-off between a model's explanatory power and parsimony, in which the first enhances prediction accuracy, while the latter ensures the model's generalizability and interpretability (Huang et al., 2003).

This increased predictive power of new technologies has more recently been combined with new available sets of data, giving rise to FinTech, described by Schueffel (2017) as “a new financial industry that applies technology to improve financial activities”.

FinTech has given a push the whole financial industry, prompting a general innovation and modernization trend to maintain competitiveness.

For example, FinTech lenders can be distinguished from traditional lenders based on their efficiency. They have been shown to have lower default rates on comparable loans, while being 20% faster than their counterparts at processing mortgage applications. As mentioned, this is due to the use of more sophisticated algorithms for credit scoring and pricing of loans that are better able to fit and learn from the underlying data (Fuster, 2019).

Additionally, researchers found that fintech lenders seem to use different information for credit-related decisions than non-fintech lenders, and this can be likely traced back to using different dimensions of “big” data that not available to traditional lenders. More specifically, these variables include the so-called “digital footprint”, which refers to the trail of data and information that individuals create and leave behind when using digital devices and the internet, ranging from the type of device and operating system used, the email and the behavior on a website (Buchak et al., 2018) to location, payment, and social media data (Koren, 2016).

These variables, together with more traditional information, have been shown to be useful for predicting default without decreasing the total number of accepted credit requests. This is due to a positive reshuffling effect, by which customers with favorable digital footprints and low credit score gain access to credit, while customers with unfavorable digital footprints and medium credit scores lose access. This effectively implies that lenders who can complement traditional credit bureau information with digital footprints will be able to make superior lending decisions (Berg, 2020).

Additionally, research shows that the informational content of digital footprint data is better modeled for prediction of borrowers' losses and defaults by machine learning techniques than traditional statistical methods.

This difference becomes even more notable in the presence of an external shock to the aggregate credit supply, where the improved performance of machine learning models compared to traditional methods can be pinpointed to the better ability to mine and model non-linear relations within the data (Gambacorta et al., 2019).

The unique potential of the digital footprint lies in the widespread availability of smartphones users around the world, including the billions of adults who still lack access to the traditional financial services, which makes it a powerful tool for promoting financial inclusion. Since it effectively serves both scorable and unscorable customers, the use of digital footprints thus has the potential to boost access to credit and lower geographical, racial and gender inequality.

As a matter of fact, the G20 High-Level Principles for Digital Financial Inclusion (2016), along with international standard-setting bodies' principles, provide a basis for country-specific action plans to leverage the potential of digital technologies. G20 members endorse these principles and encourage countries to incorporate them into their broader financial inclusion strategies, particularly in the realm of digital financial inclusion. The goal is to facilitate concrete and meaningful actions in advancing financial inclusion with the aid of digital financial services.

In recent times, the FinTech industry has shown a keen interest in harnessing this potential and is challenging conventional banking practices and devising inventive financing solutions, motivated also by the significant rise in accessible digital footprints in developing economies (Kendall, 2017). Provided that consumer interests are safeguarded, and concerns related to privacy, security, and ethical usage are adequately addressed, utilizing this data could prove to be a valuable approach in reaching unbanked individuals from economically disadvantaged backgrounds, offering them a diverse array of financial products (Kumar and Muhota, 2012).

As a matter of fact, it is essential to consider the US legal framework and analyze the impact of discrimination on US credit markets. For this purpose, the Home Mortgage Disclosure Act (HMDA) requires lenders to collect and disclose individual-level data, including race and gender, aiming to foster transparency, accountability, and academic scrutiny in a domain with a history of contentious discriminatory practices. Economists have consistently found evidence of racial discrimination (CITARE QUALUNO), although some argue that disparities in outcomes may be attributed to differences in default risk (Ladd, 1998).

Additionally, Blattner and Nelson (2021) uncovered a significant information disparity faced by lenders in the US credit markets when evaluating default risk among historically under-served groups. Widely used credit scores are shown to exhibit higher statistical uncertainty for these groups, and the researchers find that equalizing the precision of credit scores can lead to a reduction of approximately half in approval rate disparities and credit misallocation for disadvantaged groups. These findings emphasize the crucial role of addressing information disparities to promote fair and efficient credit market outcomes (Blattner and Nelson, 2021).

In the United States, credit risk evaluation operates under strict regulations and has not fully embraced algorithmic decision-making. However, even in mortgage lending, researchers are exploring the potential of machine learning algorithms to improve default predictions and enhance financial inclusion for those who might have been excluded under simpler decision-making processes (Fuster et al., 2019).

These techniques shine a new light on unfair price discrimination and equity, raising persistent policy concerns about equity across consumers (Avery et al., 2009; Traub, 2013). As machine learning advances, it becomes essential to address these issues to ensure fair and equitable credit access for all consumers. In fact, algorithmic decisions are often less transparent and harder to explain than a straightforward rules-based process, and although both human decision-making and rules-based processes are susceptible to unfair biases and inaccuracies, algorithms face heightened scrutiny due to their reduced transparency and potential for broad scalability. While a human decision-maker may exhibit varying judgments influenced by cognitive biases, an algorithmic decision based on bias can perpetuate discrimination on a larger scale, raising concerns about fairness and accountability (Lee and Floridi, 2020).

Even when these algorithms are not explicitly fed with protected characteristics such as race, religion, gender, or disability, they can indirectly infer such information, leading to disparate treatment. Research on US mortgages has revealed that Black and Hispanic borrowers are disproportionately less likely to benefit from the implementation of machine learning in credit scoring models (Fuster et al., 2019). Moreover, recent studies have cautioned that flexible statistical technologies like machine learning may reduce overall loan approval rates for disadvantaged groups (Fuster et al., 2020) and continue to generate cross-group disparities in loan terms in FinTech underwriting (Bartlett et al., 2022). Additionally, credit scores have been linked to geographic misallocation in the US mortgage market (Hurst et al.2016), and important associations have been found between geography and interest rates paid by minority borrowers (Bartlett et al., 2022).

It is therefore of the utmost importance to isolate those variables which might become sources of discrimination, even though they might not appear to be related with any sensitive characteristic such as race or gender. This would ensure that minorities are not subject to discrimination in the form of disparate impact, a concept which centers on outcomes rather than intent, encompassing any policy or practice that disproportionately disadvantages a protected group, regardless of the intention to discriminate (Baum et al. 2015).

This thesis aims to investigate the performance of machine learning algorithms and their influence on discrimination in credit allocation, with a particular focus on the mortgage market. This market significantly contributes to the continuation of wealth disparities between generations, especially for historically disadvantaged minorities, who are less inclined to invest into homeownership and accumulate home equity (Charles and Hurst, 2003; Kuhn et al., 2020; Charles and Hurst, 2002). The central question addressed is whether machine learning techniques can enhance fairness and mitigate biases in credit allocation.

Moreover, there exists a literature gap in analyzing and comparing regression techniques for predicting interest rates, as most studies have primarily focused on classification problems such as default or credit scores. This research aims to bridge this gap by exploring and contrasting regression techniques' effectiveness in interest rate prediction.

**Bibliography**

* Avery, R. B., Brevoort, K. P., & Canner, G. B. (2009). Credit scoring and its effects on the availability and affordability of credit.*Journal of Consumer Affairs, 43*(3), 516-537. Retrieved from <http://www.econis.eu/PPNSET?PPN=715455176>
* Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the FinTech era.*Journal of Financial Economics, 143*(1), 30-56. doi:10.1016/j.jfineco.2021.05.047
* Baum, D. B., Judish, J. E., Stute, D. J., & Scalia, J. (2015). Supreme court affirms FHA disparate impact claims.
* Beck, T., Demirgüç-Kunt, A., & Levine, R. (2007). Finance, inequality and the poor.*Journal of Economic Growth, 12*(1), 27-49. doi:10.1007/s10887-007-9010-6
* Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of FinTechs: Credit scoring using digital footprints.*The Review of Financial Studies, 33*(7), 2845-2897. doi:10.1093/rfs/hhz099
* Blattner, L., & Nelson, S. (2021). How costly is noise? data and disparities in consumer credit.*arXiv Preprint arXiv:2105.07554,*
* Buchak, G., Matvos, G., Piskorski, T., & Seru, A. (2018). Fintech, regulatory arbitrage, and the rise of shadow banks.*Journal of Financial Economics, 130*(3), 453-483. doi:10.1016/j.jfineco.2018.03.011
* Charles, K. K., & Hurst, E. (2002). The transition to home ownership and the black-white wealth gap.*Review of Economics and Statistics, 84*(2), 281-297.
* Charles, K. K., & Hurst, E. (2003). The correlation of wealth across generations.*Journal of Political Economy, 111*(6), 1155-1182.
* Einav, L., Jenkins, M., & Levin, J. (2013). The impact of credit scoring on consumer lending.*The Rand Journal of Economics, 44*(2), 249-274. doi:10.1111/1756-2171.12019
* Fisher, L. (1959). Determinants of risk premiums on corporate bonds.*Journal of Political Economy, 67*(3), 217-237.
* Fuster, A., Goldsmith‐Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? the effects of machine learning on credit markets.*The Journal of Finance, 77*(1), 5-47.
* Fuster, A., Plosser, M., Schnabl, P., & Vickery, J. (2019). The role of technology in mortgage lending.*The Review of Financial Studies, 32*(5), 1854-1899.
* *G20 high-level principles for digital financial inclusion* (2016).
* Gambacorta, L., Huang, Y., Qiu, H., & Wang, J. (2019). How do machine learning and non-traditional data affect credit scoring? new evidence from a chinese fintech firm.
* Huang, Z., Chen, H., Hsu, C., Chen, W., & Wu, S. (2003). *Credit rating analysis with support vector machines and neural networks: A market comparative study* Elsevier BV. doi:10.1016/s0167-9236(03)00086-1
* Hurst, E., Keys, B. J., Seru, A., & Vavra, J. (2016). Regional redistribution through the US mortgage market.*American Economic Review, 106*(10), 2982-3028.
* Kendall, J. (2017, Jan 20,). Fintech companies could give billions of people more banking options.*Harvard Business Review* Retrieved from <https://hbr.org/2017/01/fintech-companies-could-give-billions-of-people-more-banking-options>
* King, R. G., & Levine, R. (1993). Finance and growth: Schumpeter might be right\*.*The Quarterly Journal of Economics, 108*(3), 717-737. doi:10.2307/2118406
* Koren, J. R. (2016). What does that web search say about your credit.*Los Angeles Times,*Retrieved from <https://www.latimes.com/business/la-fi-zestfinance-baidu-20160715-snap-story.html>
* Kuhn, M., Schularick, M., & Steins, U. I. (2020). Income and wealth inequality in america, 1949–2016.*Journal of Political Economy, 128*(9), 3469-3519.
* Kumar, K., & Muhota, K. (2012). Can digital footprints lead to greater financial inclusion?
* Ladd, H. F. (1998). Evidence on discrimination in mortgage lending.*Journal of Economic Perspectives, 12*(2), 41-62.
* Lee, M. S. A., & Floridi, L. (2021). Algorithmic fairness in mortgage lending: From absolute conditions to relational trade-offs.*Minds and Machines: Journal for Artificial Intelligence, Philosophy and Cognitive Science, 31*(1), 165-191. doi:10.1007/s11023-020-09529-4
* Schueffel, P. (2017). Taming the beast: A scientific definition of fintech.*Journal of Innovation Management, 4*(4), 32-54. doi:10.24840/2183-0606\_004.004\_0004
* Sethi, K., Gupta, A., Gupta, G., & Jaiswal, V. (2021). Comparative analysis of machine learning algorithms on different datasets.*International Journal of Engineering, Sciences and Research Technology, 10*(6 to 12) doi:10.29121/ijesrt.v10.i12.2021.4
* Singh, A., Singh, J., Khan, A., & Gupta, A. (2022). Developing a novel fair-loan classifier through a multi-sensitive debiasing pipeline: DualFair.*Machine Learning and Knowledge Extraction, 4*(1), 240-253. doi:10.3390/make4010011
* Traub, A. (2013). DiscreDiteD.*How Employment Credit Checks Keep Qualified Workers Out of a Job, 7*
* Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing price prediction via improved machine learning techniques.*Procedia Computer Science, 174*, 433-442. doi:10.1016/j.procs.2020.06.111
* Wang, D., Hong, D., & Wu, Q. (2023). Prediction of loan rate for mortgage data: Deep learning versus robust regression.*Computational Economics, 61*(3), 1137-1150. doi:10.1007/s10614-022-10239-5