A Materialist Analysis on the Role of Algorithmic Models in Mortgage Lending Discrimination and Commentary on Solutions in Informed Model Development

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Aidan Hijazi-Klop

Spring 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Pedro A. P. Francisco, Department of Engineering and Society

Introduction

Throughout much of the western world there is an ever-worsening housing crisis, of which the housing market in the United States is a particularly bad exemplar. And even more sinisterly the existing housing inequality in the U.S. – borne as a "result of [a wealth of] historically accumulated injustices... including redlining, racial covenants,…, and predatory inclusion" (So et al., 2022, p. 1) – is only worsening as this housing crisis progresses. But housing inequality in of itself, at least on a surface level, is not a socio-technical issue; where this topic becomes a socio-technical issue is in the role of machine learning in modern lending algorithms. This insight necessitates a far more specific inquiry: what role does algorithmically based loan worthiness decision making play in the modern landscape of housing inequality? This case of an engineered system actively perpetuating, or even accelerating, the degradation of equality within the U.S. socio-economic landscape is just one of many; in the words of economic geographer, David Harvey, "capitalists… pursue the expansion of value through exploitation without regard to the social consequences" (Harvey, 2006, p. 424). And it is symptomatic of the broader issue that many engineered innovations were not intended to produce societal progress, or in the extreme case only to create progress for a select class. It is a mathematically provable fact that these models are actively producing biased loan decisions and it is the historical context of the data the models are trained on, the significance of equitable housing access as a health indicator, and the broader dialectical context that will be dissected here.

At this point, it is pertinent to take a brief aside to fully introduce the theoretical framework in use here. This vein of dialectics was introduced in the prologues of the seminal work *Kapital* (Marx, 1867), and then later formalized by the German philosopher Joseph

Dietzgen in series of articles titled *The Religion of Social Democracy*, published in *Volkstaat*[1] from 1870 to 1875. Dialectical materialism, in short, is a philosophical framework which aims to explain societal conflict and events as a product of the conflict between and contradictions of different socio-economic classes and their interests. In applying this framework to technological systems, of which loan worthiness decision models are an example, technology can be interpreted as a product of conflict between the social and economic forces that are drivers of modern decision making and interaction. The examination of this relationship between algorithmic loan worthiness decision making and the social and economic context of housing inequality and capitalist interests of lenders will be what constitutes the bulk of this work.

Research Methods

The conduction of this dialectical examination will occur primarily through use of the method of literature review. This choice comes from the fact that there is a substantive body of work on the issue of housing inequality and loan worthiness decision models across legal, financial, and historical domains that if properly analyzed and synthesized can provide a comprehensive understanding of the issue at hand; there are also a handful of a select works that attempt to approach the issue from a computational perspective. While a primary focus of this work is simply raising awareness of the truly degenerative impacts engineering can have on issues of inequality, the domain context that will be produced by literature review will also double as a base for authoring of policy recommendations, the second, smaller fork of research methodology in use here. The final block of analysis of this paper will seek to render concrete suggestions for public policy that incorporates pertinent legal and financial context as well as computational specific domain knowledge with the hope that more technically informed legislation can significantly mitigate this issue.

Background

To understand the modern significance of housing inequality and its relational to current statistical models, the history of marginalized communities facing persistent barriers in accessing safe, affordable, and stable housing being a defining feature of American society must be understood. The roots of this inequality can be traced back to the early 20th century, when discriminatory practices such as redlining - a policy "established in 1934…[by the Federal Housing Administration (FHA)]…[who refused] to insure mortgages in and near African-American neighborhoods" (Gross, 2017) and racial covenants, "which restricted home-buying opportunities for black families in white neighborhoods" (Faber, 2020, p. 4) were used to prevent Black Americans and other minority groups from accessing quality housing. These practices were enforced both by law and cultural practice of white American and remained in place for decades, resulting in concentrated poverty and substandard living conditions. This inequality was only exacerbated by additional federally implemented policies by the FHA aimed at creating a middle class and promoting homeownership. Notably the GI Bill, which was at least presented as legislation meant to provide greater access to housing and financial security for Americans of all backgrounds. In reality, however, this bill - as an extension of the "Home Owner's Loan Corporation's (HOLC) racist practices" - was almost exclusively discriminatory in application, with white Americans receiving preferential treatment. Through the GI Bill the federal government "heavily subsidized the segregation of white people and capital from black people, establishing the socio-spatial arrangement of goods and services (e.g., jobs and schools) that, in the postwar United States, [ensuring]… racialized places provid[ing] opportunities for white families' upward mobility and the dis-proportionate immobility of disadvantaged black families" (Faber, 2020, p. 25). By machinations of both public and private entities historical housing inequality can be explicitly explained as a result of the capitalist interests of the white ruling class in the United States in the post-manumission socio-economic landscape. Some redemptory work was attempted towards the end of 1960s working off the momentum of the Civil Rights movement with the passing of the Fair Housing Act which sought to address housing inequality by broadly prohibiting discrimination in housing markets. However, the Fair Housing Act only "banned racial discrimination in the sale and rental of housing, it took no action to stop discrimination in mortgage lending," (Randolph, 2015, p. 8) thus only really addressing the symptoms of the situation rather than the root cause. With the passing of the Fair Housing Act, and a smaller selection of successive bills, the conversation on housing equality shifted out of the popular discourse in the United States really until more recent times where the conversation was significantly reopened, in part due to inquiries into lending practices. The historical conflict between social classes in the United States is perfectly encapsulated in this situation, with the capitalist interests of developers, lenders, and political institutions taking clear precedence over the basic rights, of which housing is one, of marginalized communities.

Inequality of any kind is sufficient justification in of itself to warrant analysis and solution development, but in the case of housing there are concerns beyond just equal financial accessibility. There is fairly exhaustive literature on this, but as neatly summarized in a one of a series of reports by the U.S. Department of Health and Human Services: "housing issues are associated with a comprehensive range of health consequences, whose impact on marginalized residents is especially severe given the likelihood of multiple simultaneous exposures and individual vulnerability due to chronic stress" (Swope, 2019, p. 18). From increased of exposure to toxins to statistically reduced access to quality medical care to "negative mental health effects" to more even something more abstract as "ontological security" (Rolfe, 2020), there is a distinct causal relationship between access to safe, quality housing and a myriad of health factors that are strongly related to life expectancy and mortality rates. In short, housing serves as a lynchpin in both socio-economic and physical wellbeing making the technically driven discrimination here that much more problematic, and that much more necessary to significantly analyze.

Before delving into the core of said analysis, a brief aside is necessary to detail, at least on a cursory level, how the decision models work, such that the discussion is properly contextualized. Predictive machine learning models as a class are built on the principle trying to learn patterns from large datasets, and then using those patterns to make predictions about new datasets. Such predictive models find use in a wider variety of applications, from predicting stock prices to customer behavior to medical diagnoses to, as pertinent here, predicting the likelihood of defaulting on a mortgage loan. The process of constructing a predictive model, or really any machine learning product, is to gather and preprocess the data. This work involves selecting the relevant variables, referred to as features, that are pertinent to your target problem to use in the model; for the case in question here such features include address, credit score, income, and age. From there the task is to clean the data to remove any inconsistencies or errors, splitting the data into training, evaluation, and testing sets, generally on an 80%/10%/10% split respectively. Once the data has been sufficiently prepared, the algorithm or architecture to be used in the predictive model is chosen and the model is trained using the specified dataset. There are a host of different architectures and algorithms that could be use, with a variety of strengths and weaknesses, the specifics of which are not pertinent to the conversation at head, both due to the generally non-technical focus of this work and the fact that many of the loan-worthiness prediction models are proprietary. During training, the model samples repeatedly from the training set and attempts to identify features that hold the most predictive value as it adjusts parameters to optimize over the target metric: generally some form of minimization of the error between the model's predictions and the true values. While this will be elaborated on from an anthropological standpoint later it is in this step that the issue arises regarding decision bias. If the data used to train the model, even after pre-processing has embedded bias, as housing data in this country does, that bias will be perpetuated into the model's predictive nature as features that strongly correlate to specific outcomes will be racialized as a product of systemic injustices in this country.

Following the completion of training, the model is then evaluated using the test data set split during pre-processing. This allows for developers to gauge model performance on unknown data, as a supposedly reasonable stand-in for real-world use. Following the evaluation process on a controlled data set the decision model, presuming target performance metrics are met, the model is then released for use in its specific domain, loan-worthiness being the one under scrutiny here. While the evaluation process does not explicitly contribute to the bias these lending decision models exhibit, the bias has to be evident in the evaluation metrics, meaning by indifference or intent there is insufficient consideration leveled during this phase of development.

The failure to address bias prior to the use of the models in an environment which tangibly impacts consumers sets the stage for policy discussions later in this work.

Results and Discussion

Issues with housing equality in as it relates to mortgage lending has only worsened - especially following predatory housing investments by firms including, but not limited to, Pretium and Blackstone (Fields, 2022) – . Mortgage lending, while only a small element of the housing issue in the U.S., is one directly influenced by machine learning “as pricing [and lending] increasingly relies on intelligent algorithms that extract information from…real-world mortgage data” (Gillis & Spiess, 2022, p. 1). With situations as statistically significant as “Latinx and Black borrowers [paying] 4.7–4.9 basis points more in interest for GSE and FHA home-purchase loans and 1.5–1.6 basis points more for FHA and GSE refinance loans,” (Bartlett et al., 2019, p. 2), the current system is intrinsically detrimental to accessibility. When risk analysis for potential lending candidates is conducted based on data that is biased by decades of public and private policies designed to reinforce racial inequalities, “the use of algorithms can also lead to inadvertent discrimination” (Bartlett et al., 2019, p. 1). And it certainly would be simple to lay blame on the financial institutions willingly perpetuating the use of these discriminatory algorithms, that would be ignoring the broader relationship these entities have with the regulatory system at large. With “federal housing laws in the United States [failing] to catch up to technology” (Rodriguez, 2020, p. 6) the fault seems to extend beyond the lenders as they are not the only parties ignorant of, or apathetic towards, the implications that go into the data being used. It only seems natural within the capitalist construction of the U.S. economy for lending institutions to willfully ignore the ethical considerations of data used in life-changing decisions, especially if those decisions can positively the lenders' bottom line. This leads to the contention that loan-worthiness decision models, either by conscious decision or at the very least with apathy, are contributing to the burgeoning crisis of housing inequality in the United States.

The introduction of such a position inevitably produces counter-arguments, which in large can be boiled down to a single proposition. This is that in absence of human judgement, and the subconscious biases that weigh on decision making, that algorithms will, in large, provide for more equitable lending than historical human driven decisions. On the surface this assertion seems to be sound; algorithms being a purely mathematical entity should not hold the same biases that humans do. Under the smallest amount of scrutiny, however, this argument quickly falls apart. While the data and analysis present in Bartlett et al. (2019) and others irrefutably shows a clear perpetuation of, and in some cases the exaggeration of, racial bias in lending decisions, statistically sound data analysis does not take the full role it should in the legislative process in the United States. As such this argument is worth dissecting from a materialist standpoint as well. The core of the counter-argument that these algorithms are not intrinsically biased ignores the models' fundamental architecture: machine learning models, with some dependence on training parameters and available computing power, are primarily products of the data they are trained on. Within this context, the failure of the counter-argument is remarkably clear when the statistical decision models themselves are considered as contradictory materialist artifacts. With reliance on factors like zip code and income level as key features of the training data, which are intrinsically tied to historical discriminative practices, it is impossible to argue the lending decision models are unbiased, at least in the form in which deployed for use against the public. If the conclusion, effectively indisputably, is that these decision models are in fact biased, it begs the question why do contrarian opinions on the use of such decisions models exist at all.

The answer to this query, under application of a materialist framework, also addresses the quandary of conscious mal-intent versus apathy by lenders in use of the decision models. Such opposing stances on the equitability of models in use for determination of loan-worthiness can realistically only exist in a space where an interested party aims to intentionally preserve bias in the housing market in the United States. The materialist influences which would drive such stances are twofold. Firstly, based on the product of economic interests, mortgage lenders may see radicalized lending practices as a way to maximize short term economic interests, as based off un-sound judgements on protected classes, lenders can charge higher interest rates, leading a greater extraction of capital over the short-term. Secondly, the class conflict present in the United States socio-economic hierachy tend to stratify along bounds of race. And given the role housing plays in social mobility, mortgage lending is a tool that the predominantly white ruling class in the United States can use to reinforce historic and unjust racial hiearchies. The contradictions of class interests are only exacerbated by use of machine learning based loan worthiness decision making through use of models engineered to reproduce inequality under a veneer of scientificism.

At this point it has been sufficiently established that algorithmically driven discrimination is a substantive problem within the lending industry. But leaving this as only an anthropological work seems insufficient, especially in the context of active concern for responsibility of engineers for their engineered systems, discussion of solutions is pertinent. The first major introduction of a machine learning based approach was in So et al.'s (2022) work in which they coined the phrase “reparative algorithms;” a “reparative” algorithm in this case being any machine learning model, statistical or procedural in nature, that can in some way contribute to the rolling back or identification of bias enforced by these existing loan worthiness decision models. While So et al. only delved in theory into what these reparative algorithms could look like, substantive work is starting to be published, the most preeminent of which being a recent publication from MIT’s Computer Science and Artificial Intelligence Laboratory. A team of researchers produced a classifier of existing loaning algorithms which “can mitigate bias within data with multiple sensitive parameters and sensitive options while maintaining high levels of accuracy” (Singh et al., 2022, p. 12). Historically bias classification has focused on the data that goes into machine learning models, not the models themselves - with traditional bias classification methodology all being contingent on “retraining on the properly resampled data” (Vucetic, 2001, p. 1) - but the innovations produced by this lab completely shatter that tradition. This work sets the stage for tangible next steps as interested parties in the U.S. look forward with the hope of meaningfully correcting this issue.

Operating from a materialist standpoint, the implementation of policy to promote truly equitable artificial intelligence (AI) use and development requires an understanding of the underlying power structures and social relations that shape and drive the deployment of AI in the United Statest. Therefore, any policy proposal for promoting socially cognizant AI would need to address these structural factors. Broadly, legislation should on the democratization of machine learning development and ownership. Up until now, the vast majority of social significant AI research is concentrated in a select few large corporations and research institutions. Having AI under the control of a selection of purely capitalist entities limits the potential for equitable outcomes, as ultimate these corporations and research institutions will operate at the behest of the capitalistic interests. Thus the baseline for equitable AI must be a shift to public investment and ownership of these development and research processes, where equitable AI as defined by Algorithmic Justice League (2022) is AI provides agency to those who interact with it while respecting human life, dignity and fundamental rights. This is such that social good can be held above profit motives. To this end there needs to be meaningful investment in legislation on how companies can use machine learning models, which at the very least will look like independent review boards that will aim to limit corporate overreach. Additionally, there should be public funding, as well requirements for corporations to participate, in the development of AI that is intentionally engineered to address pressing social and environmental challenges, ranging from medical inequality to climate change. The feat of successfully shifting the use of AI towards social progress only becomes easier the more communal buy-in there is leading to the third policy point: international cooperation. While the focus of this paper is on the misuse of decision models within the use, development of any such tooling holds global implications, and the benefits and risks of AI are unequally distributed across countries and regions. In the spirit of promoting equitable AI, any legislation passed either by the United States government or policies agreed to by G20 countries other should instantiate rules around use of AI in neo-colonialist situations as well promoting access to AI in developing nations; this is all in the hope that AI is not used to continue the exploitation and disenfranchisement of developing nations across the globe. The final aspect that should be present in any substantive policy overhaul is a legislative side to the bias classification work previously discussed. In the hopes of avoiding the issues we see in the United States today with mortgage lending guidelines should be enacted that require use of novel bias classifiers like the one discussed above from Singh et al. (2022); moreover, companies and research institutions, in the vein of democratizing AI, should be required to have full transparency in their decision-making processes in data selection and outcomes. In summary, promoting equitable AI use from a materialist view requires writing policy that addresses, top to bottom, the underlying power structures, class conflicts, and capitalistic motives that are the drivers behind machine learning development at this time. Through democratizing AI and the addressal of bias and discrimination in machine learning outcomes and training data, social welfare can be made the priority and situations such as the gross lending discrimination borne out in the U.S. can not only be remedied but also avoided.

Conclusion

In the body of the preceding work a twofold argument has been leveled. Firstly, that housing injustice, specifically targeting LatinX and Black borrowers, has been actualized through the use of statistical decision models for loan worthiness; and the implementation and reliance on such models arises as a direct result of the contradictions of class interests between borrowers, the lending institutions, and the United States government. The goal of this discussion being to awareness among engineers and public officials to the severity and extent of the issue with the hope of prompting further, very necessary, dialogue among critical stakeholders. The second fork of the argument operates to this end: that solution crafting is not just something that exists in theory. It’s more than common in the digital age, especially when the conversation is in regard to machine learning, to concede defeat prematurely. The research and policy proposals laid out here paint a clear picture of ways that algorithmically driven housing injustice can actually be solved or the very least mitigated. In short, the hope is this work will serve as a trigger for meaningful future discourse and research in this space.

References

Algorithmic Justice League. (2022). *What Do We Mean by Equitable and Accountable AI.* The Algorithmic Justice League. https://www.ajl.org/learn-more.

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). Consumer-Lending Discrimination in the FinTech Era (Working Paper No. 25943). *Journal of Financial Economics*. https://doi.org/10.3386/w25943

Faber, J. W. (2020). We Built This: Consequences of New Deal Era Intervention in America's Racial Geography. *American Sociological Review*, *85*(5), 739–775. https://doi.org/10.1177/0003122420948464

Fields, D. (2022, January 4). Perspective | Tech and finance firms buying up homes doesn't bode well for everyone else. *Washington Post*. https://www.washingtonpost.com/outlook/2022/01/04/corporate-landlords-silicon-valley/

Gillis, T. B., & Spiess, J. L. (2022). Big Data and Discrimination. *The University of Chicago Law Review*, 31.

Gross, T. (2017, May 3). A "Forgotten History" Of How The U.S. Government Segregated America. *NPR*. https://www.npr.org/2017/05/03/526655831/a-forgotten-history-of-how-the-u-s-government-segregated-america

Harvey, D. W. (2006), *The Limits of Capital*. Verso Books.

Lee, M. S. A., & Floridi, L. (2021). Algorithmic Fairness in Mortgage Lending: From Absolute Conditions to Relational Trade-offs. *Minds and Machines*, *31*(1), 165–191. https://doi.org/10.1007/s11023-020-09529-4

Massey D. S. (2015). The Legacy of the 1968 Fair Housing Act. *Sociological forum (Randolph, N.J.)*, *30*(Suppl 1), 571–588. https://doi.org/10.1111/socf.12178

Rodriguez, L. (2020). All Data Is Not Credit Data: Closing the Gap Between the Fair Housing Act and Algorithmic Decisionmaking in the Lending Industry. *Columbia Law Review*, *120*(7), 1843–1883.

Rolfe, S., Garnham, L., Godwin, J., Anderson, I., Seaman, P., & Donaldson, C. (2020). Housing as a social determinant of health and wellbeing: Developing an empirically-informed realist theoretical framework. *BMC Public Health*, *20*(1), 1138. https://doi.org/10.1186/s12889-020-09224-0

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. https://doi.org/10.3390/make4010011

So, W., Lohia, P., Pimplikar, R., Hosoi, A. E., & D'Ignazio, C. (2022). Beyond Fairness: Reparative Algorithms to Address Historical Injustices of Housing Discrimination in the US. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 988–1004. https://doi.org/10.1145/3531146.3533160

Stolcke, A., & Droppo, J. (2017). Comparing Human and Machine Errors in Conversational Speech Transcription. *Interspeech 2017*. 141. https://doi.org/10.21437/Interspeech.2017-1544

Swope, C. B., & Hernández, D. (2019). Housing as a determinant of health equity: A conceptual model. *Social Science & Medicine*, *243*, 112571. https://doi.org/10.1016/j.socscimed.2019.112571

Footnotes

[1] *Der Volkstaat* was a newspaper of the German Social Democratic Workers Party that was published in Leipzig from October of 1869 to September of 1876.