Algorithmic Decision Making and its Impact on Society.

Thomas Walker

December 2021

Abstract

Algorithms and Artificial Intelligence (AI) are becoming increasingly common in our society and play an important role in the lives of many. Despite the countless benefits they may provide there can be threats to certain groups. Why do these risks occur and what can we do to mitigate these risks? The aim of this essay is to shed light on these questions using credit scoring algorithms as a case study. These algorithms are used to determine the creditworthiness of an individual (The World Bank Group, 2019). Therefore, they play a key role society and it's important we understand any threats they may cause.

1 Algorithms, how do they work?

Algorithms and AI encompass a field of programs that are optimised to a specific task. In the topic of credit scores, techniques optimised for making predictions, or performing classification are deployed. Algorithms are designed to receive data, and then iteratively train on the data to become aligned to the goal of the operator. There are two main types of data used regarding credit scoring algorithms, there is traditional data and alternative data. Traditional data includes transaction data and credit checks whereas alternative data includes utilities, social media, and online transactions (The World Bank Group, 2019). Different algorithms emphasise different variables, however, there is a collective focus on factors such as having long credit history and having various types of credit (Masonmac, 2021).

Intuition would indicate that having access to vast quantities of data would increase the accuracy of these algorithms. Although these systems can analyse multiple variables, if they are developed poorly they can make decisions worse (Fernandez & Menajovsky, 2019). If the data is inadequate problems such as overfitting may occur. This is where algorithms exacerbate intricate patterns that are not particularly representative. A significant issue as we may not even realise these connections which become proxies for sensitive variables (Balogh & Johnson, 2021). This can have detrimental financial implications for specific individuals.

AI methodologies fall into some broad categories. We have supervised learning techniques, that are trained using labelled datasets. Similarly, we have unsupervised learning techniques that use unlabelled data sets (The World Bank Group, 2019). On the one hand, unsupervised learning techniques may be used to limit a bias incurred when labelling a dataset. Therefore, the training process starts from a clean slate and the algorithm is free to infer relationships that it believes will help prediction. However, unsupervised techniques are very opaque and do not allow for regulations to be easily enforced. The complex associations developed make it difficult for the operators to be certain that the algorithm is making its decisions in a fair manner. In more niche financial communities there tends to be less regulation and little oversight to the deployment of these algorithms (Fernandez & Menajovsky, 2019). Therefore, having obscure systems may leave these oversights unidentified, causing unnecessary amounts of damage to individuals.

2 Societal evolution is hindered by the deployment of algorithms

AI has come to the forefront of technology at a rapid pace and is playing an important role in our society. They are discrete systems operating all around us (Dreyfuss, 2017). This AI boom is a consequence of better algorithms and faster computers (Lyons, 2017). The accelerated and widespread integration of AI means that it has large implications on culture and society. Often the media has skewed societies perception of AI, for instance through films. Furthermore, AI gets viewed in a negative way as people feel as if they are being subjected to these algorithms constantly and sometimes unknowingly.

In the past humans have processed applications for financial services, humans are susceptible to intentional and unintentional bias. Individuals make decisions predominantly on traditional data; however, their decisions can involve racial and gender discrimination (Balogh & Johnson, 2021). As historical data is then used to train these algorithms, they are inevitably going to perpetuate these biases. Since credit scoring algorithms play an increasingly significant role in society these sorts of injustices underlying the algorithms will hinder societal progress. If systems continue to show signs of historic societal norms, then we will not be able to leave those unwanted ideologies behind as they will be present in our everyday lives. It becomes more difficult for those challenged by society to promote change as they are disadvantaged, their loan applications are rejected, or they have a lower probability of receiving a mortgage. This can have consequences on someone's stability which then tends to indicate to an algorithm that these individuals should have low credit scores (Fernandez & Menajovsky, 2019), resulting in a negative feedback loop. This can diminish a groups significance in society. Currently about 54% of Black consumers, 41% of Hispanic consumers, 37% of white consumers and 18% of Asian consumers report on having no credit scores or credit scores deemed to be low (Balogh & Johnson, 2021), illustrating the discrepancies in race and ethnicity. Current algorithmic systems to determine credit scores are inconsistent.

To combat these issues, we must resolve underlying issues in society to provide these algorithms with clean data to make representative predictions. However, this is a difficult process and will require time and a multitude of different initiatives (Heaven, 2021). Alternatively, operators are running two models in parallel. One which functions in the same way as current credit scoring algorithms, whereas the other tries to predict race. Giving an indication to operators whether certain variables are being used as proxies for race and causing discriminative outputs (Balogh & Johnson, 2021). This is a solution which aids in the interpretability of a model, allowing people to understand the sensitivity to certain features (The World Bank Group, 2019) and formulate plans to improve the algorithms fairness. This solution allows for the credit scoring algorithm to maintain its complexity whilst still being interpretable. Inevitably as these systems start to use unstructured data sources (social media) (The World Bank Group, 2019), these systems will become less interpretable so these types of solutions are vital. There is a balance between progressing AI research and protecting individuals from the repercussions its integration into society may have (Kerry, 2020).

Even once we try make algorithms interpretable it still raises the question as to whether what it sees is the same as what we see. Firstly, we could improve interpretability by using simpler models (The World Bank Group, 2019), this is effective as the relationship between variables are easier to understand. However, due to the sophistication of the credit scoring problem it is difficult to reduce it to a simple representation. Alternatively, take the technique of running two models in parallel. We can "see" what the algorithm "sees" but what can we really determine from this? There is a huge amount of context that comes into play when we are seeing (Hanson & Lund, 2018), something that an algorithm doesn't necessarily have despite the vast quantities of data. Seeing is something that is done in a calculated and premeditated way and hence relies on experience (Hanson & Lund, 2018). Algorithms are mainly unimodal; they are developed for a certain modality (Center for Research on Foundation Models, n.d.). Therefore, these models lack a lot of the context and do not necessarily see the same things we see according to (Hanson & Lund, 2018). If we observe the two models producing satisfactory outputs, these outputs may not be made with correct reasoning. Which may cause downstream issues.

3 Who is responsible for algorithmic discrimination, and what should they do about it?

Terah Lyons strongly believes that policy making must involve everybody, everyone needs to develop AI, and everybody should benefit from it (Lyons, 2017). This idea originates from the fact that these algorithms rely heavily on the general population to function, they require access to lots of data. The population are in general oblivious to the fact that they are data points for these algorithms to be trained. The lengthy privacy policies and terms and conditions are a burden and people just accept them without reading it, according to (Kerry, 2020) this burden should be shifted toward the operator. Forcing operators to become more transparent with their systems to provides clarity to its users, helping to generate a sense of trust between the users and the operator. Motivating individuals to aid the cause of fairer algorithms, we need everyone to feel as if they have the responsibility to develop these systems (Lyons, 2017). Creating an inclusive environment will change the philosophy of how algorithms are designed. Rather than being coded to be fair, the algorithms will be fair from the start. The latter, I believe, is a fundamentally different algorithm.

Individuals being unaware as to how these algorithms operate with their data is part of the larger issue of transparency. The EU enforces transparency in algorithms through its GDPR law. Putting an emphasis on explaining algorithms decisions and allowing people to gain clarity on the algorithm's outputs (GDPR EU, n.d.). However, it can be rather restrictive in terms of improvements as it puts a limit on the intelligence of these algorithms. This ceiling on complexity may restrict an algorithm's ability to make accurate predictions.

Government action that cooperates with the technology industry is required (Lyons, 2017). However, government actions are slow initiatives that often lag present societal demands. We therefore look toward the operators to implement measures to improve accountability. One such measure involves carrying out risk assessments, a versatile solution enforced by the operator or a third party. It can be carried out in proportion to the significance of the decisions the algorithms make (Kerry, 2020). This is a pre-emptive measure that can prevent algorithms threats to certain groups of people. It provides a definitive answer to who is responsible for mishandling of threats of a deployed algorithm. This puts pressure on companies to operate fair algorithms as there are possibility for consequences to be enforced. However, it is critical governments understand the field of technology (Lyons, 2017), so that there is clarity and consensus in the policies being issued.

We all hold a collective responsibility to mitigate algorithmic discrimination and develop better practices for future algorithms.

4 Data density and how this affects the accuracy of algorithms

There has been an ongoing concern that current credit scoring algorithms disadvantage those with limited amount of credit history. For example, immigrants may have low repayments due to their limited access to local networks (Fernandez & Menajovsky, 2019). Therefore, algorithms have less data to learn meaningful patterns, resulting in credit scores for immigrants that are not as accurate. Credit scoring software typically favours wealthier white applicants and hinders minority groups. This skew comes from the fact that these minorities have fewer financial data points (Heaven, 2021). Leading to a cycle where those with minimal financial records are denied financial services and thereafter, they are denied future financial services due to lacking records. Credit scores are often used to make life changing decisions so these are big issues, and sometimes fairer algorithms will not fix them (Heaven, 2021).

One proposition to solve these issues is to give loans to boarder line candidates (Heaven, 2021). Allowing individuals to build up a credit history that is representative of their financial behaviours and situation. Companies could experience short term loses with this solution so may be hesitant to adopt it. However, the long-term benefits may mean that algorithms become more accurate which then benefits the operator. Perhaps this is the point that government intervenes to provide compensation to those that implement these solutions. Overtime datasets will become more representative and unbiased, helping the operators manage their finances with lower levels of risk as individuals will receive a fairer score. Furthermore, this will lead to a greater trust between the public and the algorithms as individuals feel as if the algorithms understand them. This goes to reversing the negative perception of AI and increasing the diversification of those in the field of AI.

On the other hand, we could allow algorithms to use more data about applicants (Heaven, 2021), helping predictions to become more accurate. However, it raises questions about privacy and interpretability as understanding the links between more variables is challenging. More data leads to a more granular analysis on individuals which may cause unanticipated inferences (Kerry, 2020). Initially, the individual consented for their information to be used to generate a credit score. Now the operator is in the situation where they have extracted additional information about the individual which no longer relates to their credit score. As AI evolves it will have a greater ability to intrude on the privacy of individuals, in extreme circumstances this can be used as a tool for authoritarian control (Kerry, 2020) which would only emphasise the societal issues discussed so far.

5 Conclusion

Due to the importance of credit scoring algorithms, they can have lasting impacts on and individuals. Work to ensure these algorithms do not misrepresent certain groups needs to be a sustained effort. This is critical as society is continually evolving and so the algorithms must evolve too. Government policy often lags this change in society; therefore, companies and individuals must act proactively and collectively to ensure that credit scoring algorithms are fair and give accurate decisions for everyone.

6 References

Balogh, S & Johnson, C. (2021) AI can help reduce inequity in credit access, but banks will have to trade of fairness for accuracy – for now. https://www.businessinsider.com/ai-lending-risks-opportunities-credit-decisioning-data-inequity-2021-6?r=US&IR=T [Accessed 13th December 2021].
Center for Research on Foundation Models, n.d. On the Opportunities and Risks of Foundation Models, Stanford University.
Dreyfuss, E. (2017). Algorithms and Society | STEAM: Ideas That Shape Our World. https://www.pbslearningmedia.org/resource/6e4fb1b0-2657-4a86-9cd8-fbee55af4af3/algorithms-and-society-steam-ideas-that-shape-our-world/ [Accessed 8th December 2021].
Fernandez, M. V. & Menajovsky, J. (2019) Algorithm Bias in Credit Scoring: What's Inside the Black Box? CGAP https://www.cgap.org/blog/algorithm-bias-credit-scoring-whats-inside-black-box [Accessed 12th December 2021]
GDPR EU, n.d. What is GDPR, the EU's new data protection law? https://gdpr.eu/what-is-gdpr/ [Accessed 20th December 2021]
Hanson, N. R. & Lund, M. D. (2018) Seeing and Seeing As. In: Perception and Discovery: An Introduction to Scientific Enquiry. Springer International Publishing pp. 67-82. Heaven, W. D. (2021) Bias isn't the only problem with credit scores – and no, AI can't help. https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/ [Accessed 11th December 2021]
Kerry, C. F. (2020) Protecting privacy in an AI-driven world. https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/ [Accessed 9th December 2021]
Lyons, T. (2017) The Role of Artificial Intelligence in Society [Video]. TED Conferences. https://tedxboston.com/videos/the-role-of-artificial-intelligence-in-society/
Masonmac (2021) Why Credit Scoring Models Show Different Scores. https://www.masonmac.com/why-credit-scoring-models-show-different-scores/ [Accessed 12th December 2021]
The World Bank Group (2019) Credit Scoring Approaches Guidelines, The World Bank Group