Designing an Ethical Algorithm

For ISEC3050 Ethics and Law in Data Analytics

By: Michael Laprise

Due: October 19th, 2025

Introduction:

For this assignment, we are tasked with outlining in sudo code an algorithm for a Bank Loaning service that features no room for discrimination of any kind.

For the typical loan process, the bank does the following calculations for their risk assessment:

- Credit history
- debt
- collateral
- earning power
- education
- experience

I will be sticking with these processes and going in depth with possible areas of discrimination (Systemic or accidental) that I try and fix.

To note:

```
1. While creating code that will check and correct itself for possible
discrimination is within good intentions, it is very much possible to
accidentally over correct and create discrimination against another group.
Example being: You are creating an algorithm that will filter candidates. You
company sees there is a history of female candidates being tossed aside too
often, so you do a quick correction of making it so there are an equal amount
of female and male candidates in the final selection phase. Afterwards you
realise that instead of having more women in the final selection phase, your
algorithm ends up tossing away equally qualified male candidates without a
proper reason, to simply fulfill it's instructions.
2. An area that tends to have some systemic discrimination is the data used to
train/create an algorithm. A good way to avoid this is to use up to date
recent records based on modern practices.
```

Sudo coding:

To note:

Final approval process involves signing and finalizing documentation which requires either in person or virtual meeting with a banker.

Start the process via gathering client info relevant to the business loan approval process and split into two core categories of discrimination risk.

Low risk are data values in which the algorithm will prioritize validating through and comparing within a range of previous loans to see if the business can handle a loan.

Low risk pool of data:

```
Business name and type, age of business, sector/industry, annual/quarterly
revenue and projections, net profit, debt and repayment records, collateral,
bank statements, personal income, personal debt, guarantees, business plan,
references, bills (rent, utilities, and more split), taxes and licenses, sale
transactions
```

High risk are the areas in which the client information may be important to collect and use for the loan process, such as for paperwork, but may lead to some discrimination if used in the algorithm. This data would need to be manually checked by a person if the loan requires a closer look at for clearance.

High risk pool of data:

```
Full name, address, contact, Credit history, personal assets, education, Age,
Gender, Ethnicity, Postal code, Marital status, citizenship/immigration status
```

Next, algorithm collects recent data from previous applications, using a timeframe of 5 years as a default, but adjustable by branch to their preferences and practices, and then creating a range in which the previous applications were successful, and thus for the remainder of the process, categorizing which were failed and success and using the flagged areas from previous applications to help construct the range, the whereabouts of previous applications data is unknown and only a substitute number is used that has no real world counterpart.

Next, the algorithm sorts the low risk pool and starts comparing relevant data for the loan process (Business name can not be measured for example)

If the algorithm sees all relevant criteria (Not counting areas that are blank on the client side, this applies for every check as to avoid disapproving a new company that has not done a tax report yet or a quarterly review) are above the threshold for approval (or within a small margin of 5%-10%), notify client for final approval process

elif the algorithm detects majority of criteria are above threshold, (roughly) 70/80% + but not 100%
if detected areas are within a margin (10-20)% of approval, proceed with final approval but flag areas for agent review
elif detected areas are below margin, flag client with problem and flag banker to verify with high risk data pool if client is able to proceed with loan acquisition

elif algorithm detects majority of criteria fail the check, proceed with notifying client for an automatic disapproval and to continue talks with agents if they disagree.

else if an outlier happens during the process, such as the algorithm outputs a majority of criteria is met but does not meet the threshold for automatic approval, proceed with data comparison of high risk pool data, such as credit history, personal assets, education level
Flag for agent final review if the client passes the high risk check (similar to low risk check but with larger threshold for success for bank related information (assets, credit))
proceed with loan pre-disapproval if these checks fail, flag client to encourage meeting with banker

Pre-approval process

During the pre-approval process, the algorithm will fill out information related to the loaning process in order to aid speed up the process, such as filling out client name from the high risk pool, or filling out relevant data related to their finances.

Finalisation

It is important that the final steps are done with an agent involved, such as signing papers. This makes it so a human can stop any accidental slip throughs in the system, such as someone who was not deserving of the loan but got approved by the algorithm because it saw they met the criteria through the data.

Algorithm analysis

This section will cover a six-step analysis of the algorithm to determine if it succeeds a non-discriminatory check

Fairness:

I believe, since this algorithm only uses data relevant to the financial records of the business, and comparing it to recent data in the regional branch, and has a margin of acceptance for each section, that this gives ample room for possible clients who may be disqualified from hard numbers while also keeping the algorithm regionally relevant.

Transparency:

The algorithm flags and notifies the client of each section that does not pass approval, and, due to paperwork, all relevant data used in the algorithm is shown on paper categorized for the client.

Privacy:

At all points of the process, data used for comparison from previous applications are only used for a hidden calculation for creating a range in which approval can be met. The numbers used for the checking do not come directly from previous applicants directly and thus said data is hidden from the current client. The current client's data is also secured in similar ways and is only shown to the final agent for the final approval if the data is relevant to the loan.

Data Quality:

To ensure data quality, the data is taken from a range of applications from the previous couple of years and is specific for each branch, making it regionally relevant.

Accountability:

No matter the process, the algorithm always transfer over to an agent for final step of the process, or for the customer to talk to and directly challenge the final conclusion.

Governance:

To ensure governance, the data is separated into risk factors of discrimination, in which can be altered by makers of the algorithm at later date, with the high risk factors being used only in extreme cases and flags human agents for manual review, or the high risk factors are used strictly for filling in relevant information.

Conclusion

While I hope I have achieved making an ethical algorithm, I am not perfect and there are probably holes in my approach. I am a very numbers oriented person and may not have seen certain areas in which discrimination can happen.

That being said, I made this algorithm in mind that high and low risk factors can be altered easily so that someone in the future can change them with the time, or even per region. The people in charge of this is a strict group and hopefully will need approval and oversight from various parts of the company to ensure no abuse of power from bad actors.