Fighting discrimination in mortgage lending

March 30 2022, by Adam Zewe

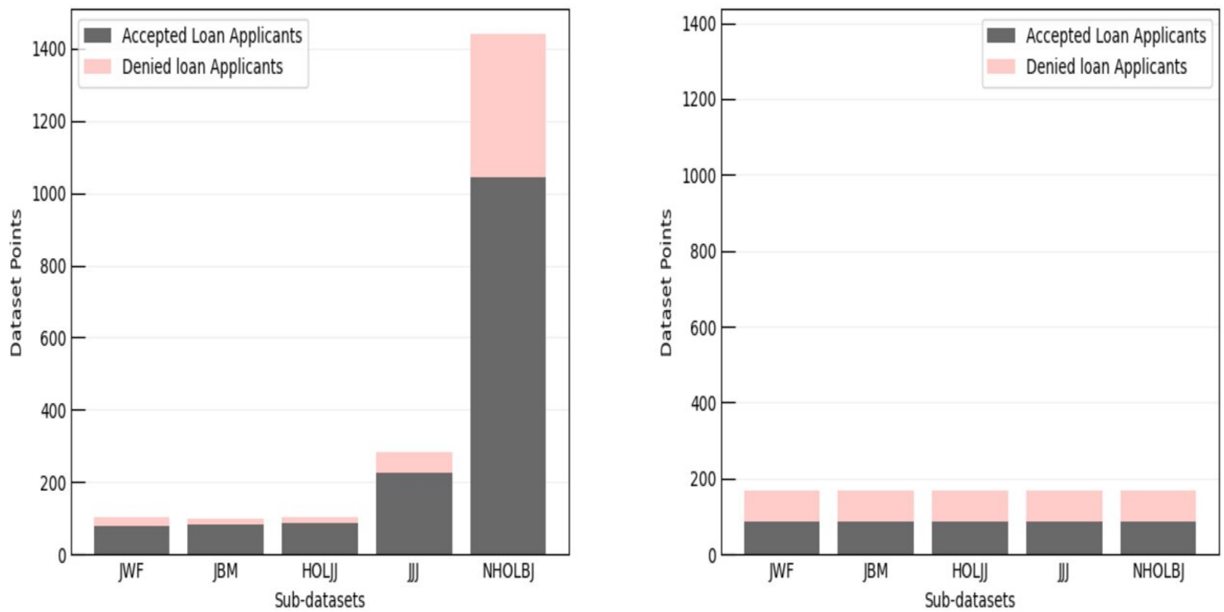

Figure 1. Class and label balancing before (left graph) and after (right graph) using DualFair. (Note: only 5 sub-datasets from the balancing procedure of Connecticut, a smaller sample of HMDA, are shown by the normal 27; note that J denotes joint, B denotes Black, W denotes White, F denotes female, M denotes male, HOL denotes Hispanic or Latino, and NHOL denotes Non-Hispanic or Latino. These denotations form sub-groups shown above. For instance, JJJ is a joint, joint, joint individual and HOLJJ is a Hispanic or Latino, joint, joint individual). Credit: *Machine Learning and Knowledge Extraction* (2022). DOI: 10.3390/make4010011

Although the U.S. Equal Credit Opportunity Act prohibits discrimination in mortgage lending, biases still impact many borrowers. One 2021 *Journal of Financial Economics* study found that borrowers from minority groups were charged interest rates that were nearly 8% higher and were rejected for loans 14% more often than those from privileged groups.

When these biases bleed into machine-learning models that lenders use to streamline decision-making, they can have far-reaching consequences for housing fairness and even contribute to widening the racial wealth gap.

If a model is trained on an unfair dataset, such as one in which a higher proportion of Black borrowers were denied loans versus white borrowers with the same income, credit score, etc., those biases will affect the model's predictions when it is applied to real situations. To stem the spread of mortgage lending discrimination, MIT researchers created a process that removes bias in data that are used to train these machine-learning models.

While other methods try to tackle this bias, the researchers' technique is new in the mortgage lending domain because it can remove bias from a dataset that has multiple sensitive attributes, such as race and ethnicity, as well as several "sensitive" options for each attribute, such as Black or white, and Hispanic or Latino or non-Hispanic or Latino. Sensitive attributes and options are features that distinguish a privileged group from an underprivileged group.

The researchers used their technique, which they call DualFair, to train a machine-learning classifier that makes fair predictions of whether borrowers will receive a mortgage loan. When they applied it to mortgage lending data from several U.S. states, their method significantly reduced the discrimination in the predictions while maintaining high accuracy.

"As Sikh Americans, we deal with bias on a frequent basis and we think it is unacceptable to see that transform to algorithms in real-world applications. For things like mortgage lending and financial systems, it is very important that bias not infiltrate these systems because it can emphasize the gaps that are already in place against certain groups," says Jashandeep Singh, a senior at Floyd Buchanan High School and co-lead author of the paper with his twin brother, Arashdeep. The Singh brothers were recently accepted into MIT.

Joining Arashdeep and Jashandeep Singh on the paper are MIT sophomore Ariba Khan and senior author Amar Gupta, a researcher in the Computer Science and Artificial Intelligence Laboratory at MIT, who studies the use of evolving technology to address inequity and other societal issues. The research was recently published online and will appear in a special issue of Machine Learning and Knowledge Extraction.

Double take

DualFair tackles two types of bias in a mortgage lending dataset—label bias and selection bias. Label bias occurs when the balance of favorable or unfavorable outcomes for a particular group is unfair. (Black applicants are denied loans more frequently than they should be.) Selection bias is created when data are not representative of the larger population. (The dataset only includes individuals from one neighborhood where incomes are historically low.)

The DualFair process eliminates label bias by subdividing a dataset into the largest number of subgroups based on combinations of sensitive attributes and options, such as white men who are not Hispanic or Latino, Black women who are Hispanic or Latino, etc.

By breaking down the dataset into as many subgroups as possible, DualFair can simultaneously address discrimination based on multiple attributes.

"Researchers have mostly tried to classify biased cases as binary so far. There are multiple parameters to bias, and these multiple parameters have their own impact in different cases. They are not equally weighed. Our method is able to calibrate it much better," says Gupta.

After the subgroups have been generated, DualFair evens out the number of borrowers in each subgroup by duplicating individuals from minority groups and deleting individuals from the majority group. DualFair then balances the proportion of loan acceptances and rejections in each subgroup so they match the median in the original dataset before recombining the subgroups.

DualFair then eliminates selection bias by iterating on each data point to see if discrimination is present. For instance, if an individual is a non-Hispanic or Latino Black woman who was rejected for a loan, the system will adjust her race, ethnicity, and gender one at a time to see if the outcome changes. If this borrower is granted a loan when her race is changed to white, DualFair considers that data point biased and removes it from the dataset.

Fairness vs. accuracy

To test DualFair, the researchers used the publicly available Home Mortgage Disclosure Act dataset, which spans 88% of all mortgage loans in the U.S. in 2019, and includes 21 features, including race, sex, and ethnicity. They used DualFair to "de-bias" the entire dataset and smaller datasets for six states, and then trained a machine-learning model to predict loan acceptances and rejections.

After applying DualFair, the fairness of predictions increased while the accuracy level remained high across all states. They used an existing fairness metric known as average odds difference, but it can only measure fairness in one sensitive attribute at a time.

So, they created their own fairness metric, called alternate world index, that considers bias from multiple sensitive attributes and options as a whole. Using this metric, they found that DualFair increased fairness in predictions for four of the six states while maintaining high accuracy.

"It is the common belief that if you want to be accurate, you have to give up on fairness, or if you want to be fair, you have to give up on accuracy. We show that we can make strides toward lessening that gap," Khan says.

The researchers now want to apply their method to de-bias different types of datasets, such as those that capture health care outcomes, car insurance rates, or job applications. They also plan to address limitations of DualFair, including its instability when there are small amounts of data with multiple sensitive attributes and options.

While this is only a first step, the researchers are hopeful their work can someday have an impact on mitigating bias in lending and beyond.

"Technology, very bluntly, works only for a certain group of people. In the mortgage loan domain in particular, African American women have been historically discriminated against. We feel passionate about making sure that systemic racism does not extend to algorithmic models. There is no point in making an algorithm that can automate a process if it doesn't work for everyone equally," says Khan.

More information: Arashdeep Singh et al, Developing a Novel Fair-Loan Classifier through a Multi-Sensitive Debiasing Pipeline: DualFair, *Machine Learning and Knowledge Extraction* (2022). DOI:

10.3390/make4010011

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology