Preventing Bias in Credit Limit Algorithms

Mary Odenthal

Data Science

Regis University

MSDS 640

Dr. Ghulam Mujtaba

08/24/2025

Abstract

This case study will address gender bias in credit limit recommendation algorithms. Often in financial situations, women have less representation due to them filling the role of full time caretaker in a familiy more often than men. Therefore, women have less representation as income-earning individuals, leading to biased data in finance. Using kernel density estimate plots, strategies for detecting bias will be explored. Over and under sampling examples will be shown to balance classes and potentially reduce bias.

Preventing Bias in Credit Limit Algorithms

Observed Bias in Credit Limit Algorithms

Bias is a serious ethical issue in algorithms, as it has the highest impact on minority groups which may already be disadvantaged. It is essential to prevent bias in systems that have a direct impact on an individual's well-being like medical or financial recommenders. In 2019 Apple partnered with Goldman Sachs to release a credit card. An algorithm was used to determine applicant's credit limit for the card. The card was exposed on social media as having a potential bias towards women, "tech entrepreneur David Heinmeier Hansson wrote that Apple Card offered him 20 times the credit limit as his wife, although they have shared assets and she has a higher credit score" Nedlund (2019).

However, other experts claimed to debunk the allegations of bias stating, "For example, women tend to make less money than men, and with income being a significant factor in determining credit limits, it wouldn't be surprising to see that women might end up having lower credit limits than men" Nedlund (2019). How do we determine when an algorithm is producing biased or logical results when we cannot see exactly how the algorithm is making it's determinations?

Detecting Bias in Machine Learning Algorithms

It is important to thoroughly test an algorithm for biased results so that bias can be properly addressed. In this case study, the dataset (Credit Card Regression) was used to create a credit limit recommendation system. The recommendation system was tested for bias by using kernel density estimate plots to compare the probability density curves of each gender in relation to their real and predicted credit limits.

In figure 1, the probability density of each gender in relation to credit limit is overlayed for comparison. The differences in the densities can represent potential bias due to uneven gender representation in the target variable, in our case credit limits. Specifically, the distribution has a higher probability of women represented in the lower end of the credit limits, (0-5,000). Accompanied by a higher probability of mean in the upper end of the

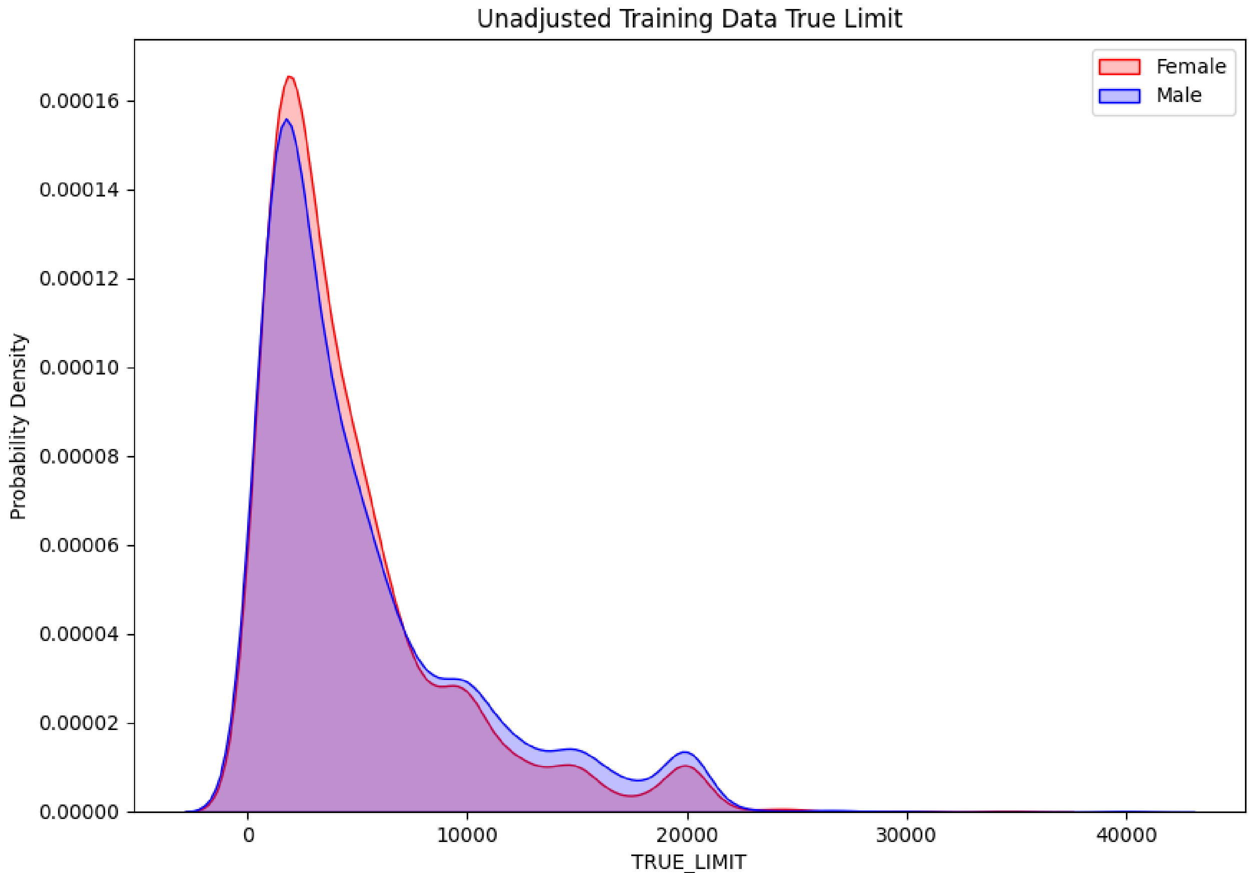

Figure 1

Real credit limit density in training data

credit limits, (10,000-20,000). When genders have differing representation in training data, the result will often be a model that is biased towards gender.

It is clear in figure 2, which shows the probability density of the ridge model's credit limit predictions by gender, that the probability density of a model is likely to closely follow the same density that was present in the training data. To further confirm the theory that models will follow the density of training data, a second type of regression model was trained and tested. Also using a lasso regression model, a kernel density plot was created, (Figure 3). The differences between the ridge and lasso model results were so small in every test case that we omitted the lasso model results from the rest of the study and only represent the ridge model results for comparison.

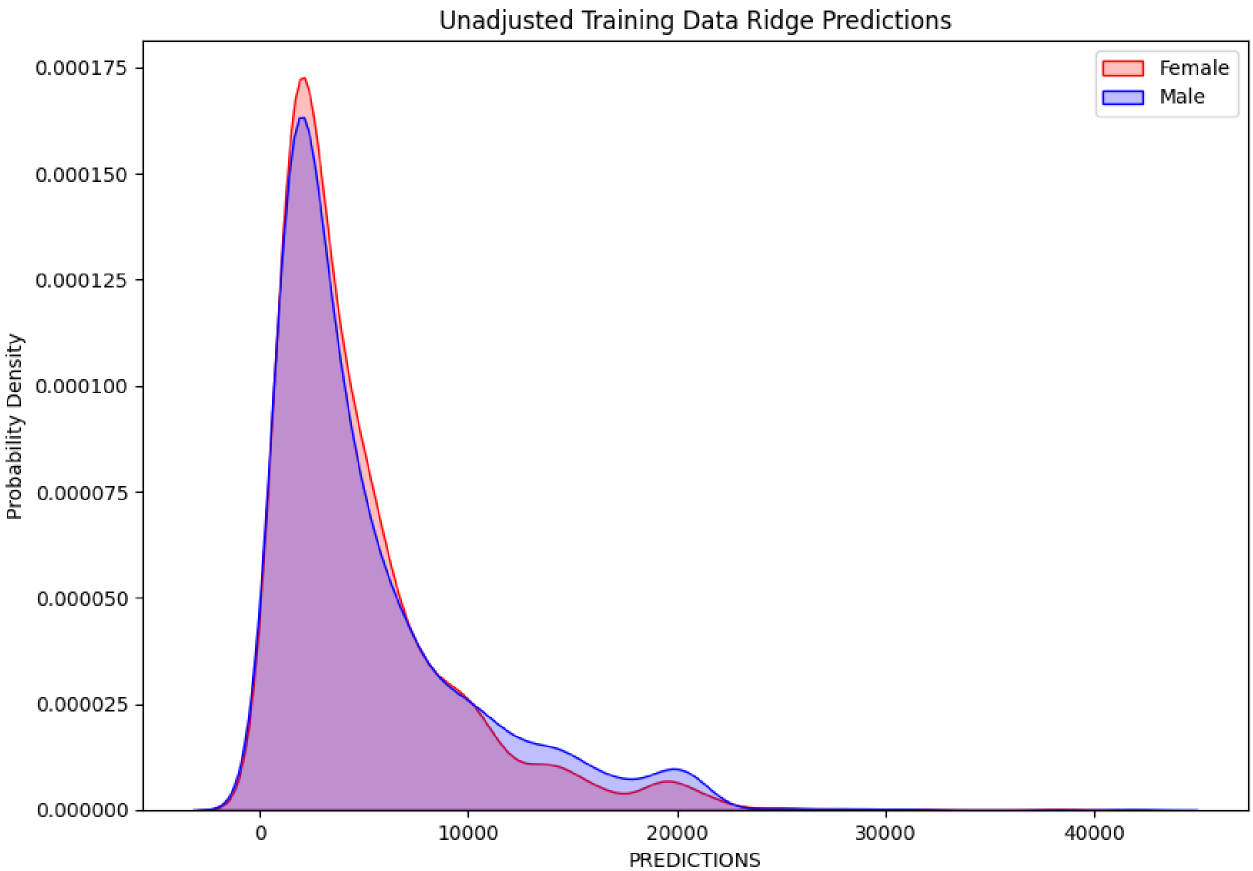

Figure 2

Prediction probability from a ridge model on unadjusted training data

The evidence shows that models will often mimic the probability densities of the data on which they were trained. Therefore, it is important to test and identify how different disadvantaged groups may be underrepresented in training data to prevent creating a model with biased results. Machine learning models often assume equal representation will be present which is why it is important to balance the classes which the model should not judge against. If classes are not balance, a model will interpret the underrepresented classes as a meaningful indicator. Because models are purely logical, they often amplify bias that is present in data. Testing and adjusting biased training data is an essential part of being an ethical machine learning engineer.

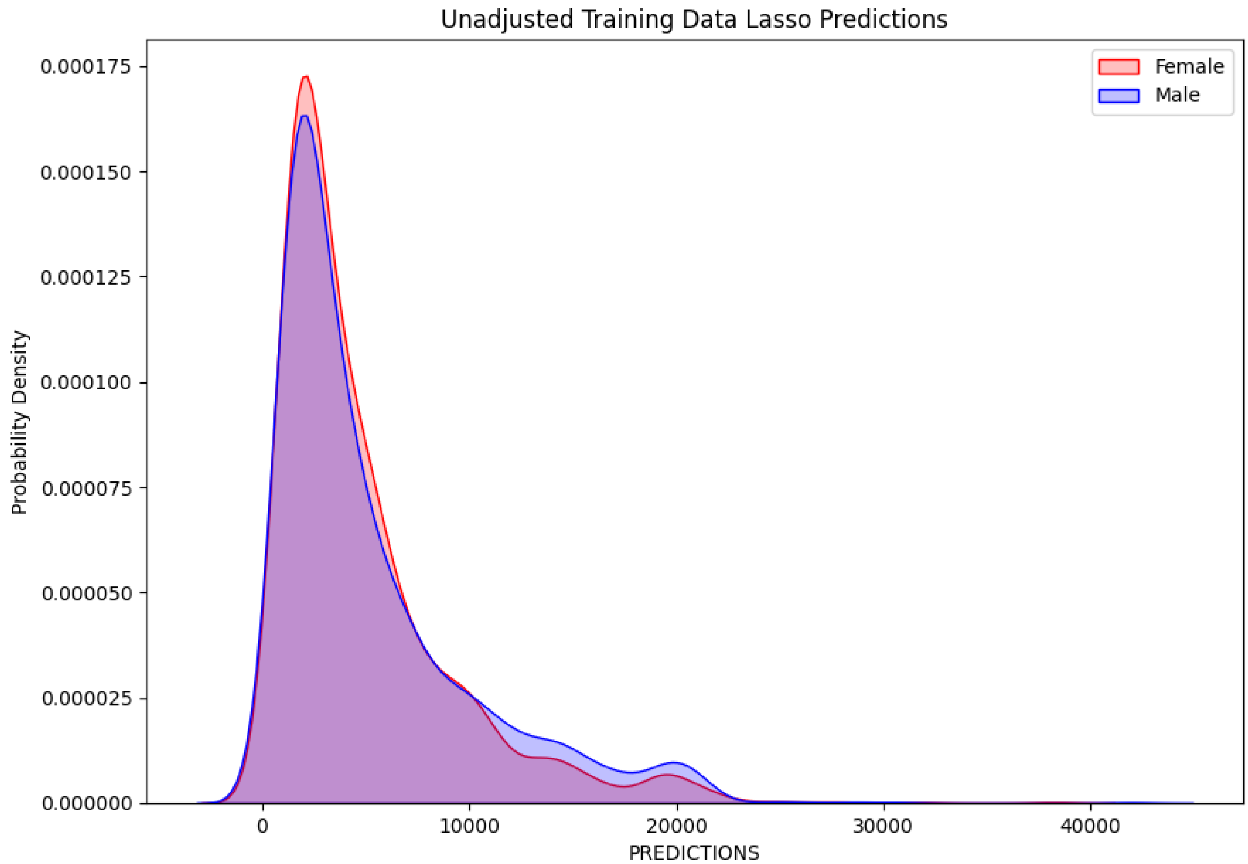

Figure 3

Prediction probability from a ridge model on unadjusted training data

Addressing Bias in Algorithm Results

Once a bias is identified, then it must be resolved. A common strategy for addressing bias is balancing minority classes in the training data. Sampling techniques including under and over sampling can be used to balance classes. By balancing classes, it is ensured that there is equals representation of classes in the training data. In this study, the gender classes were balanced to ensure an equal split of 50% male and 50% female. However, interesting results were discovered between under and over sampling technique.

Over Sampling to Balance Classes

First, examine the impact the over sampling had on the probability density. The differences between female and male probability densities has increased in terms of the upper

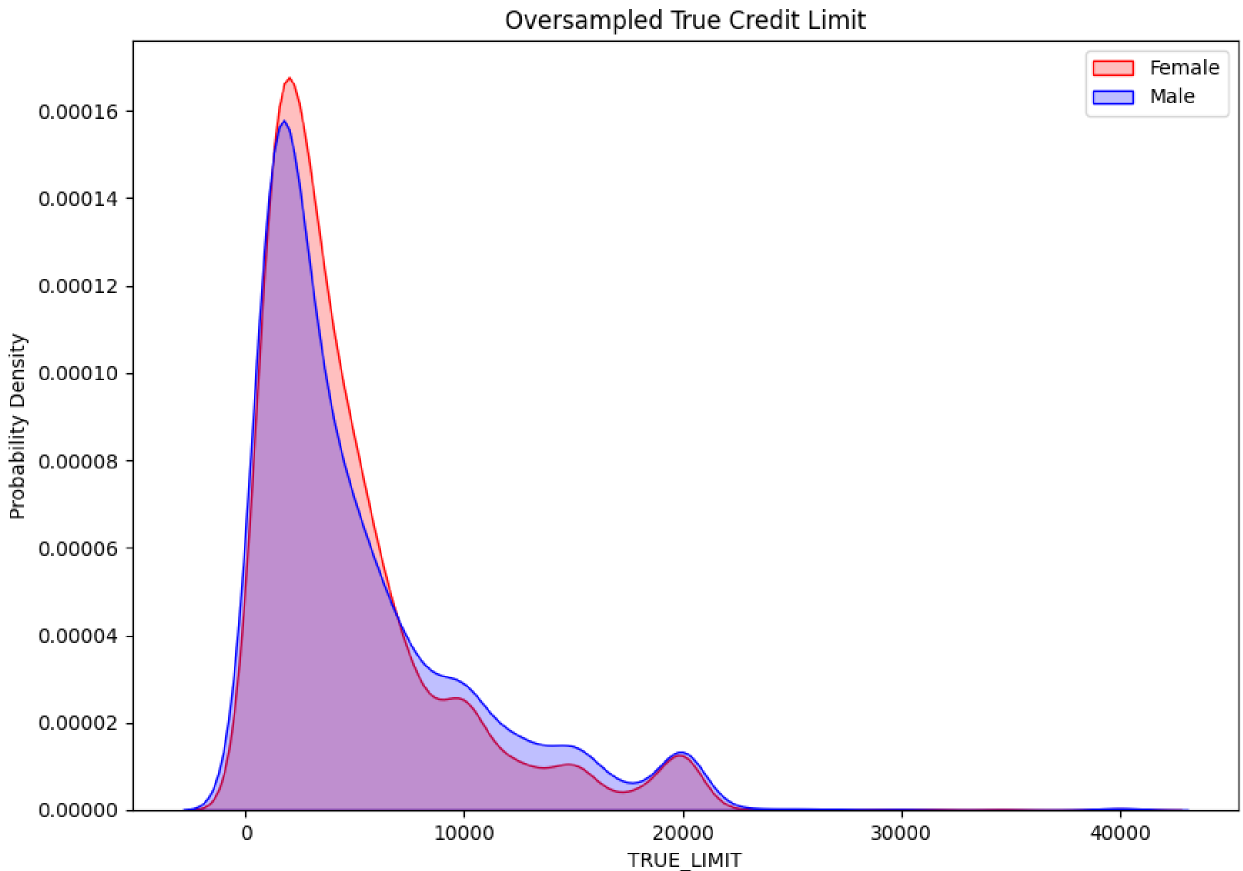

Figure 4

Prediction probability from a ridge model on unadjusted training data

and lower limits. Over sampling balances uneven classes by taking a larger number of samples from the minority class until it reaches the same size as the majority class. In the case of the credit limit data, over sampling has lead to an increased amount of female representation in the lower end of credit limits while further increasing the amount of male representation on the upper end of the credit limits.

It is important not to assume balancing the classes will solve biased data. The sampling strategy used can have a large impact on the success that balancing classes has on reducing bias. In the test case using over sampling, the ridge model performed with a higher rate of biased as reflected in the adjusted training data. Will under sampling perform better?

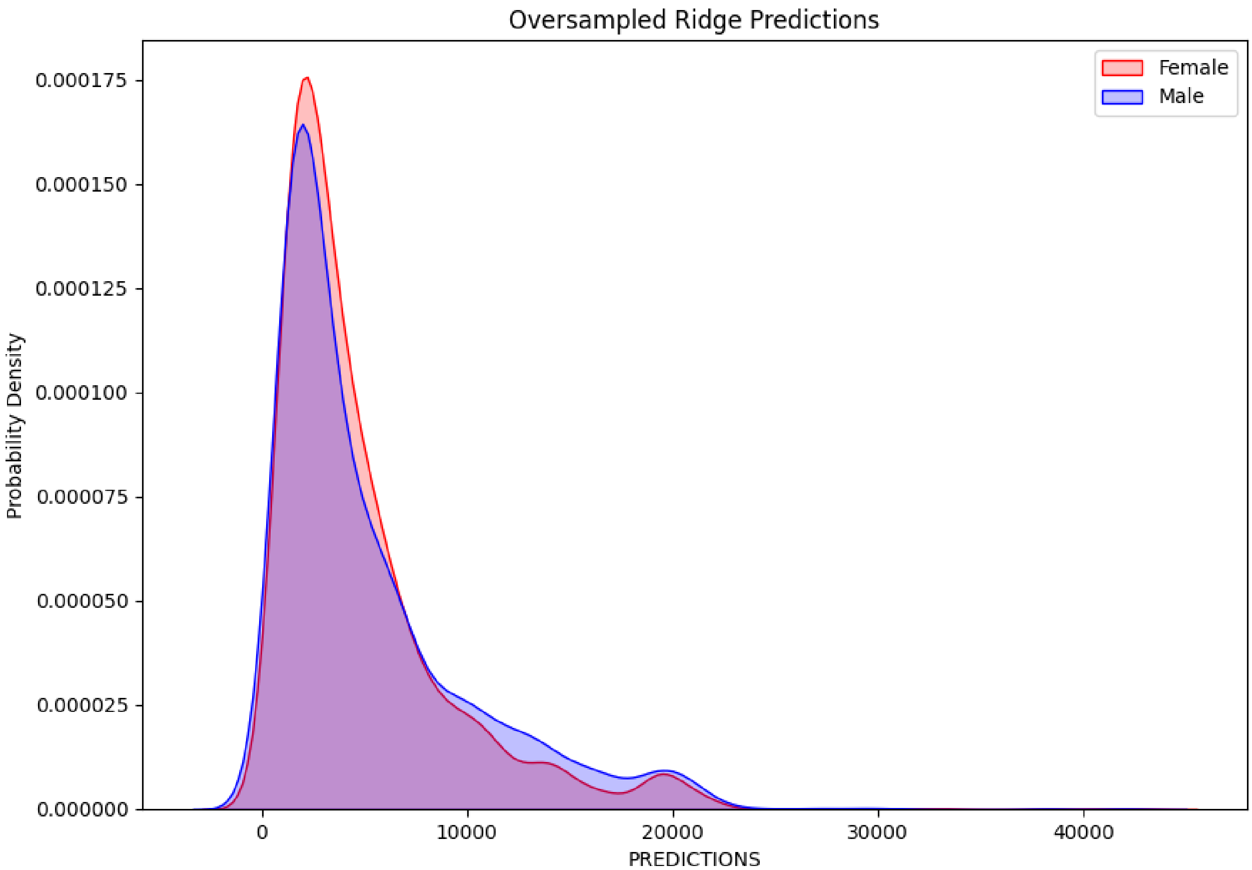

Figure 5

Prediction probability from a ridge model on unadjusted training data

Under Sampling to Balance Classes

Under sampling differs from over sampling in the method by which is balances classes. Under sampling reduces the majority class to the size of the minority class by sampling less than the whole majority. Through this method, classes can be balanced by increasing minority representation without reducing diversity in the samples. Over sampling reduces diversity in samples by repeating the same samples to inflate the minority classes. This can lead to reduce algorithm accuracy as more real, diverse samples lead to results that better reflect the real world.

Under sampling proves to be the better case for balancing the credit limit data. The differences in probability density between males and females are very small now with almost

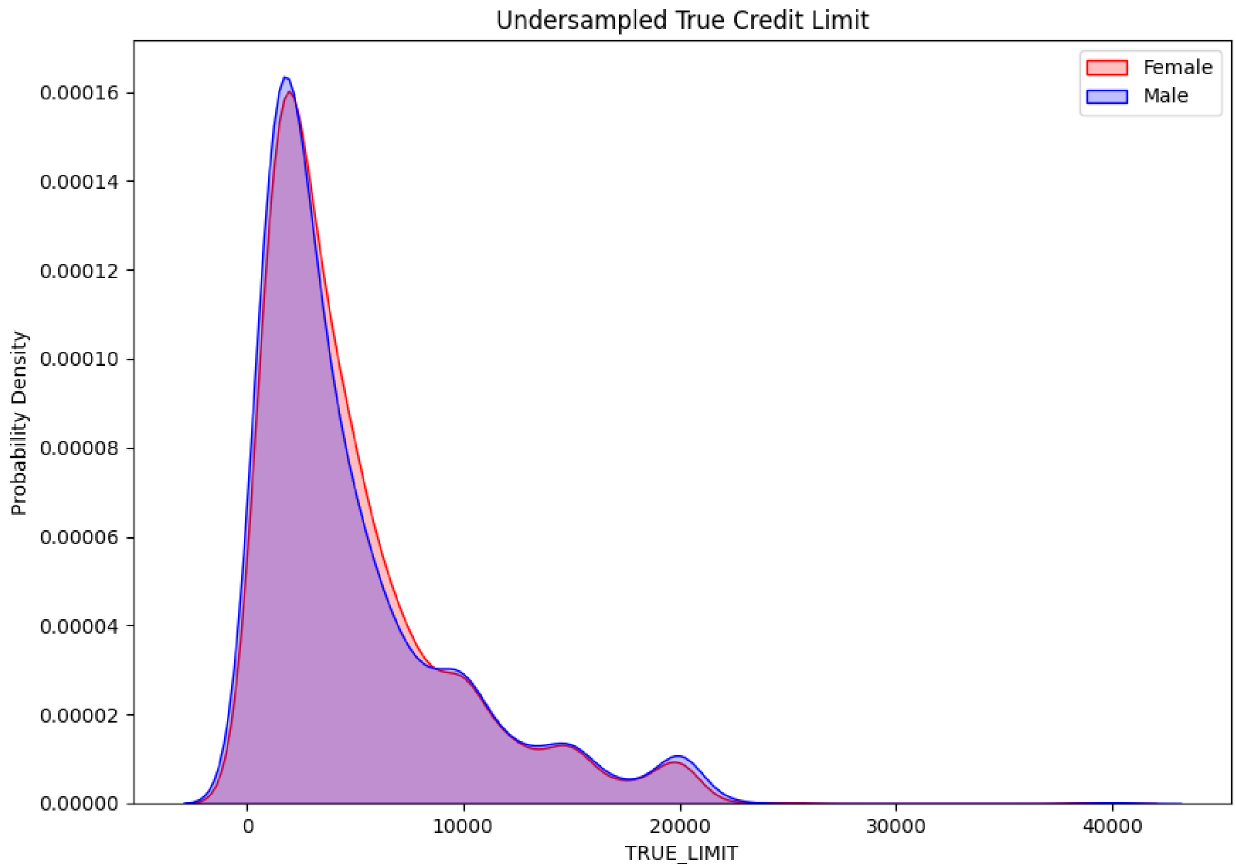

Figure 6

Prediction probability from a ridge model on unadjusted training data

equal representation at all intervals of credit limit. The improved representation is reflected in the next ridge model's results. The final results of the ridge model trained on under sampled data show the least bias towards gender.

Conclusion

Testing for representation of potential disadvantaged groups in training data is a very important step to preventing harmful bias in prediction models. An ethical machine learning engineer is one who actively seeks out potential bias in training data and test methods to resolve it. Class balancing is often a useful tool to reduce bias, but the strategy used to balance classes is important. Over sampling can yield less accurate and more biased results, so it should be careful validated. Under sampling was proven to be the most effective

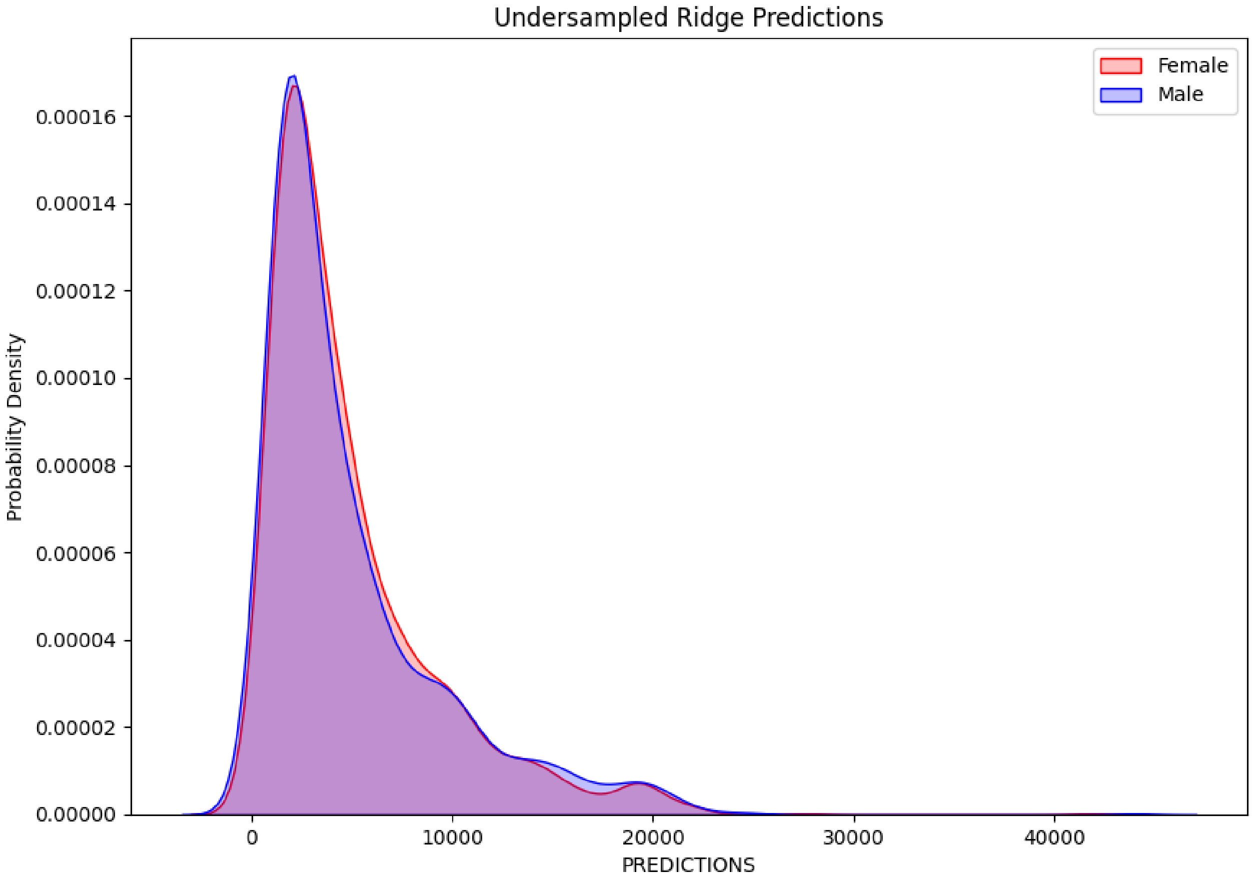

Figure 7

Prediction probability from a ridge model on unadjusted training data

method of reducing bias for the example data set on credit limits.

References

Nedlund, E. (2019). *Apple Card is accused of gender bias. Here's how that can happen | CNN Business — cnn.com.* `https://www.cnn.com/2019/11/12/business/apple-card-gender-bias/`. ([Accessed 09-07-2025])