

A major issue inherent to my model is the data on which it has to be trained; credit card transactions are an understandably private topic, and certain personal details such as what was purchased and from whom are necessary for a comprehensive model. In fact, the only dataset I could obtain for my project was entirely synthetic, as any real datasets had to be processed and encoded such that their labels could not be determined. The data I used included columns for names, numbers and CVVs for every card which I immediately dropped—any commercial model should not be exposed to such private information, however that raises the question of where training data will be sourced from and who will do such cleaning. It should be a given that no transaction will be able to be used for model training without the user's consent, however the mechanisms behind such an opt-out/opt-in feature could be either annoying to a user or deceptively obscured. One solution could be that, when a transaction is identified as fraud, the owner of the card may be asked if the transaction information can be reported and used to prevent fraud in the future; even still, records of authentic transactions will be needed as well. Ultimately, of course, credit card companies will need to be transparent about what data they are collecting, how it is collected and what it is being used for.

As mentioned previously, my model was trained on synthetic data and thus may or may not be applicable to real world scenarios. It was also difficult to visualize any clear relationships and draw any specific conclusions given the high dimension count of my data after processing, though I believe with more time I could have come up with more methods for graphing the trends. Given all this, it would be interesting to see how the model performs on real transactions if I were able to source some.