Avinash Bunga

Information Systems and Business Analytics, Park University

CIS625HOS2P2025 Machine Learning for Business

Professor: Abdelmonaem Jornaz

April 2, 2025

Unit 3: Discussion

Hi everyone,

When I was working on detecting fraud patterns at Uber, we built a model to find suspicious trips. We mainly used data like trip duration, distance, how the trip was paid for, and the driver's rating. At that time, we thought this was enough. However, looking back, I realize we missed out on some important information, which caused something called omitted variable bias (Uber, n.d.).

One of the big things we missed was looking at how old the driver's or rider's account was. Many fraudulent drivers often use brand-new accounts or hijack fresh accounts for their activities. Since we did not include this in our model, it could not catch this common fraud signal (Faivusovich, 2024).

Another thing we overlooked was the percentage of unpaid or "Did Not Settle" (DNS) transactions linked to an account. Fraudsters often team up with fake rider accounts that take trips and do not pay, leaving the balance unsettled. If we had included this DNS percentage in our model, we could have spotted risky accounts much better. It would have helped reduce the number of cases where fraud went unnoticed (Zelvenskiy, Harisinghani, Yu, Ng, & Wei, 2022).

Later on, after learning from our initial results, we went back and actually added these missing variables into the model. We pulled the account age and DNS data from different

tables and merged them with our main dataset. Once we did that, the model's performance noticeably improved. It started catching fraud patterns more effectively and reduced false negatives. This whole process taught me the value of stepping back, questioning my assumptions, and making sure you are really using all the relevant data available.

Looking back, this was a good learning experience for me, and it made me more careful in checking for possible omitted variables in future models I worked on.

Have any of you faced a situation where you realized after building a model that you had left out an important variable? I would love to hear how you handled it!

References

Faivusovich, A. (2024, May 31). *Times they are a-changin': Adapting fraud detection for a new era in account tenure*. Unit21 Fraud Files Blog. Retrieved April 2, 2025, from https://www.unit21.ai/blog/times-they-are-a-changin-adapting-fraud-detection-for-a-new-era-in-account-tenure

Uber. (n.d.). *Fraud activities*. Uber. Retrieved April 2, 2025, from https://www.uber.com/pt/en/drive/driver-app/fraud-activities/

Zelvenskiy, S., Harisinghani, G., Yu, T., Ng, E., & Wei, R. (2022, February 1). *Project RADAR: Intelligent early fraud detection system with humans in the loop*. Uber Blog. Retrieved April 2, 2025, from https://www.uber.com/blog/project-radar-intelligent-early-fraud-detection/