Meta Verse Online Purchasing Anomaly Detection

Chris Whitmire’s Data 350

Presented by: Michael (Caleb) Dolan

4/24/2024

**Abstract**

For this project I decided to pick a project that aligned with my interests as a finance major and as a data analytics minor. The goal of this process is to detect anomalies within monetary transactions within the metaverse. The dataset I selected for this goal contains 78,600 rows of transactions within the year 2022. The features within this dataset are described in full below (fig. 1). In this case, anomaly will be the target variable. The reason this topic is of interest is because I feel like risk management will be massively important as AI and virtual reality develop into common everyday use. Big tech companies such as Meta have always been a career choice that I hope to have an option one day. So, to gain experience and to work with some unfamiliar data from a company that I may hopefully work for one day I chose this topic accordingly.

A screenshot of a computer

Description automatically generated

**Fig. 1 (**[Metaverse Financial Transactions Dataset (kaggle.com)](https://www.kaggle.com/datasets/faizaniftikharjanjua/metaverse-financial-transactions-dataset/data))

**Peer Reviewed Articles**

Before beginning this project, I selected three reliable articles that have performed similar projects with different data. I was not able to find any peer reviews on metaverse anomaly detection specifically so these comparisons will be slightly different. However, all these comparisons do deal with monetary transaction anomaly detection. The overall predictability scores within these articles were very high. On the lower end one peer review reported predictability scores as low as 60% with the highest scoring article achieving in the range of 90%-99%. I planned to aim high regardless and chose the metrics below to compare my results with (Fig.2), which came from a peer reviewed article sited at the end of this paper.

A screenshot of a computer

Description automatically generated

Fig. 2 (A-Review-On-Credit-Card-Fraud-Detection-Using-Machine-Learning.pdf)

**Process**

Upon cleaning, observing and feature selection of my data, I realized that my dependent variables had a very low correlation with my independent variable, which was troublesome given the fact that low correlation could mean bad results. You can see the correlation of my features in the heat map below (Fig.3). I accepted my fate and began performing hyper parameter optimization for my machine learning processes of choice. I chose to use linear regression, support vector machine, neural network, k- nearest neighbors and gradient boosting regressor for my model selection process. After finding my optimized parameters I began my folding processes to avoid bias and was shocked with my results.

A red and white squares with numbers

Description automatically generated

Fig. 3

**Results**

My results after folding and training my machine learning process were so much higher than I expected. For my decision tree I was even able to achieve a R^2 of 100% and a MSE (mean squared error) of 0. Below you will see my results in full. Since my target variable is not specifically binary, I stuck with these metrics for my calculations. However, if I had more time, I could convert my target variable so that I could produce the same scoring types as my peer reviewed papers performed.

A screenshot of a computer

Description automatically generated

Fig. 4

**Conclusion**

Throughout this project I had a lot of fun learning about anomaly detection while getting to dig into some interesting data. I would suggest using Decision Tree for anomaly detection since the results came back so well. This machine learning process is great for categorical data and can be used to make predictions in the long run. Detecting fraud is essential for any company that applies online transactions to protect the company from lawsuits and to allow its users to stay safe.

**Works Cited**

1. Mînăstireanu, E., & Meșniță, G. (2019). An analysis of the most used machine learning algorithms for online fraud detection. *Informatică Economică*, *23*(1/2019), 5–16. <https://doi.org/10.12948/issn14531305/23.1.2019.01>
2. *A REVIEW ON CREDIT CARD FRAUD DETCTION USING MACHINE LEARNING*. (2019, October). International Journal of Scientific & Technology Research. Retrieved April 24, 2024, from https://www.researchgate.net/profile/Chetan-Awati/publication/336552027\_A\_Review\_On\_Credit\_Card\_Fraud\_Detection\_Using\_Machine\_Learning/links/5da55c7745851553ff9211cd/A-Review-On-Credit-Card-Fraud-Detection-Using-Machine-Learning.pdf
3. Saputra, A., & Suharjito, S. (2019). Fraud Detection using Machine Learning in e-Commerce. *International Journal of Advanced Computer Science and Applications/International Journal of Advanced Computer Science & Applications*, *10*(9). <https://doi.org/10.14569/ijacsa.2019.0100943>

GitHub link: