CS 846 Project Proposal

Detecting Potential Fraudulent Trading Patterns in Decentraland: A Data-Driven Analysis of the Ethereum Network and OpenSea Marketplace

Hesam Sarkhosh Sarkendi
Master Student at University of Waterloo
Student ID: 21072434
Instructor: Professor Paolo Alencar
Date: Tuesday, October 8, 2024

Overview and Objective

With the rise of decentralized ecosystems in the Web3 space, virtual economies such as **Decentraland**, a decentralized virtual reality platform built on the **Ethereum blockchain**, are gaining traction. In Decentraland, users can buy, sell, and trade virtual parcels of land as **non-fungible tokens (NFTs)**, with these transactions facilitated on platforms like **OpenSea**, a leading NFT marketplace. While these platforms provide opportunities for growth and creativity, they also introduce new risks for fraudulent activities, such as wash trading and price manipulation.

This project aims to detect potential fraudulent trading patterns in Decentraland's virtual land market by analyzing data from both the **Ethereum blockchain** and the **OpenSea marketplace**. In particular, the project focuses on detecting specific trading anomalies, including wash trading and other potential fraudulent behaviours.

The project will leverage the **IITP-VDLand dataset**, a comprehensive collection of Decentraland parcel data, which includes attributes such as parcel characteristics, trading history from OpenSea, Ethereum transaction activities, and social media interactions. A unique feature of this dataset is the "Rarity score," which measures the uniqueness of each parcel, offering a novel dimension for analyzing the market. Access to this dataset was granted by the authors and is not publicly available.

Approach

The approach will follow a structured data science pipeline to uncover potential fraudulent trading patterns, potentially using traditional data analysis or machine learning methods. The key steps are outlined as follows:

1. **Data Exploration and Preprocessing**:

   The dataset will be thoroughly explored to understand the structure and content of each segment (Characteristics Data, OpenSea Trading History, Ethereum Activity Transactions, and Social Media Data). Special attention will be given to identifying missing values, outliers, and trends in the "Rarity score" and other key metrics. Data will be cleaned and prepared for analysis, including normalizing and transforming features where necessary.

2. **Feature Engineering**:

   Based on the dataset's characteristics, relevant features for detecting trading anomalies will be derived. This could include metrics such as transaction frequency, price volatility, and patterns in parcel ownership. Social media interactions will be integrated to see if hype-driven trading correlates with suspicious transactions. The "Rarity score" will be used as a unique feature to assess its potential relationship with abnormal trading behaviours.

3. **Model Development**:

   As part of this project, the goal is to develop a model that can detect potential fraudulent trading patterns in Decentraland's market data. At this stage, the specific type of model or machine learning method has not been finalized. The choice of model will depend on the insights gathered during data exploration and feature engineering, which could point toward either a supervised or unsupervised approach.

4. **Model Evaluation and Validation**:

   Once the models are developed, they will be validated using statistical tests and evaluated based on their precision, recall, and ability to detect anomalies. Cross-validation and different evaluation metrics will be employed to ensure robustness.

5. **Result Interpretation and Visualization**:

   The outcomes of the models will be interpreted, and potential fraudulent patterns will be visualized using graphs and heat-maps.

Milestones

Milestone 1: Data Exploration and Preprocessing (Weeks 1–2)

- Explore the IITP-VDLand dataset (Characteristics, Trading History, Ethereum Activity, Social Media Data).
- Clean and preprocess the data.
- Perform exploratory data analysis (EDA) and visualize initial patterns.

Milestone 2: Feature Engineering and Model Development (Weeks 3–5)

- Engineer features based on parcel trading history, rarity score, social media activity, and Ethereum transaction data.
- Build machine learning models for anomaly detection (e.g., clustering, anomaly detection algorithms).
- Identify and detect specific fraudulent trading patterns through model analysis.

Milestone 3: Model Evaluation and Validation (Weeks 6–7)

- Evaluate models using cross-validation and suitable metrics.
- Refine models based on performance and fine-tune the parameters.

Milestone 4: Final Analysis and Report (Week 8)

- Generate results and interpret detected patterns.
- Create visualizations.
- Write the final report summarizing the project, findings, and implications for fraud detection in virtual economies.