BLG561E Deep Learning Project Proposal: Predicting Malice in Blockchain Tokens

1st İsmail ÇİFTÇİ
Computer Engineering
Computer and Informatics
ciftcii22@itu.edu.tr
150220033

2nd İbrahim ÖKSÜZ
Computer Engineering
Computer and Informatics
oksuz22@itu.edu.tr
150220011

I. Project Description

New tokens are frequently launched on the cryptocurrency market, particularly on the Ethereum blockchain [1]. In launching these new tokens, a smart contract is deployed that defines the token's economics. While many of these tokens are legitimate projects, there is also a significant amount of fraudulent schemes exercised in the ecosystem, known as "pump and dump" and "rug pulls". This project aims to utilize deep learning techniques to analyze blockchain data and identify malicious patterns.

II. Problem Definition

The problem we want to produce a solution for in our project is the early detection of those fraudulent schemes. Our objective is to build a predictive model using early-stage token data. We will investigate two primary modeling approaches: predictive survival analysis to estimate the time until a token price effectively reaches zero, and binary classification to determine if a token is destined to fail, for example, within three months. The final model type will be selected based on further research of what our model could become capable of with our compiled dataset, and what we know how to implement.

III. Dataset

We will generate a custom dataset by scraping data from the Ethereum blockchain, using tools compatible with block explorers such as Etherscan. The data collection will focus on the launch window of various tokens.

The key features to be extracted might include:

- **Token Ownership Distribution:** The initial allocation of tokens defined in the smart contract.
- **Transaction Flow:** The movement of tokens between wallets within the first few hours of trading.
- **Holder Analysis:** The number of unique holders and the concentration of wealth among the top addresses.

IV. Methodology

We will employ a Neural Network (NN) architecture to process the collected blockchain data.

A. Input and Output

The input to the neural network will be the vector of features derived from the token's early-stage data. The output will accord to the chosen objective: either a continuous variable for predicted survival time or a binary variable for fraud classification.

B. Evaluation Metric

To evaluate the success of the model, we might use two different approaches based on the output type.

- **Survival Time Prediction:** Metrics will focus on measuring the accuracy of the time prediction like Mean Absolute Error (MAE) and metrics suitable for survival analysis.
- **Fraud Classification:** Standard classification metrics will be used, including Precision, Recall, and F1-Score.

References

[1] "Cryptocurrency 'pump and dump' scams," Yahoo Finance. https://uk.finance.yahoo.com/news/cryptocurrency-pump-and-dump-scams-university-c html