Development of a risk analysis framework for cryptocurrency wallets

Junhyoung Oh, Hyochang Baek, Chang Yeon Kim, Kyungho Lee

Institute of Cyber Security and Privacy, Korea University, South Korea

Introduction

- Cryptocurrency is a digital currency that operates independently without central control using blockchain technology.
- Cryptocurrency can be used as a means of crime such as money laundering.
- It is important to calculate the risk in cryptocurrency trading.

Methodology 1

- **Research Scope**
  - Upbit : Upbit is one of cryptocurrency exchanges which has a large scale of trade volume among the markets.
  - Ethereum: Ethereum is a distributed computing platform for implementing smart contract functionality based on blockchain technology
- **Types of Wallets**
  - In this study, wallets traded with Upbit's typical wallet (0x390de26d772d2e2005c6d1d24afc902bae37a4bb) were analyzed intensively.
  - We conducted a preliminary investigation by analyzing 78 wallets with 20 or more transactions with Upbit's wallet.
  - 67 of 78 wallets were classified into 4 types as follows.

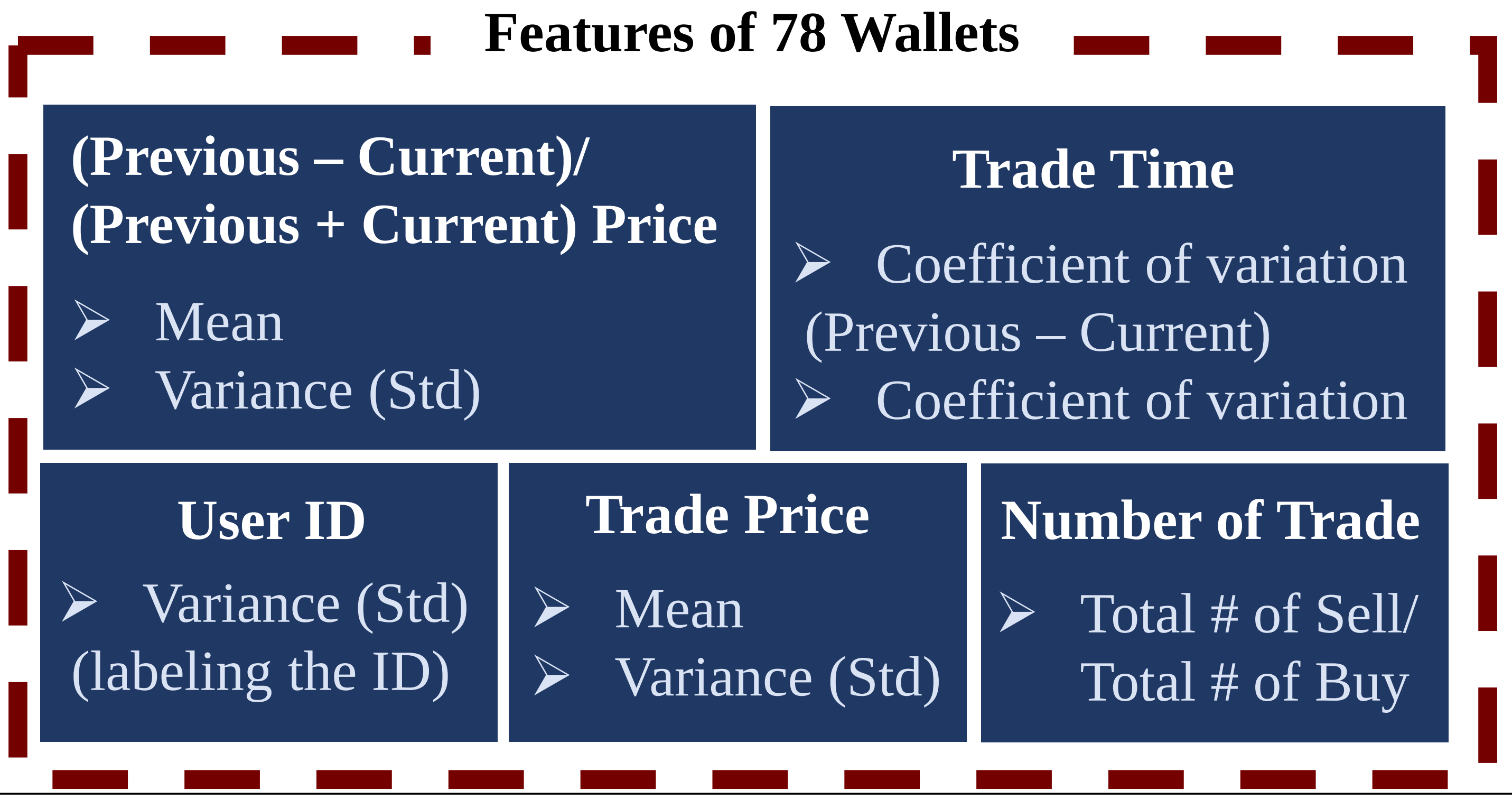
Type 1. The wallet receives from the Mining Pool and continues the transaction to the exchange for that amount.

Type 2. The wallet is received from a specific wallet and sent directly to the exchange. The specific wallet deals with a large amount of money from any one wallet and sends it to a small number of wallets.

Type 3. The wallet is traded on several occasions, but the total amount sent and received is almost the same in one session.

Type 4. The wallet exchanges almost the same amount as the exchange.

- **Feature extraction based on types of wallets**
  - Based on these types, we derived five major features in the cryptocurrency transaction.

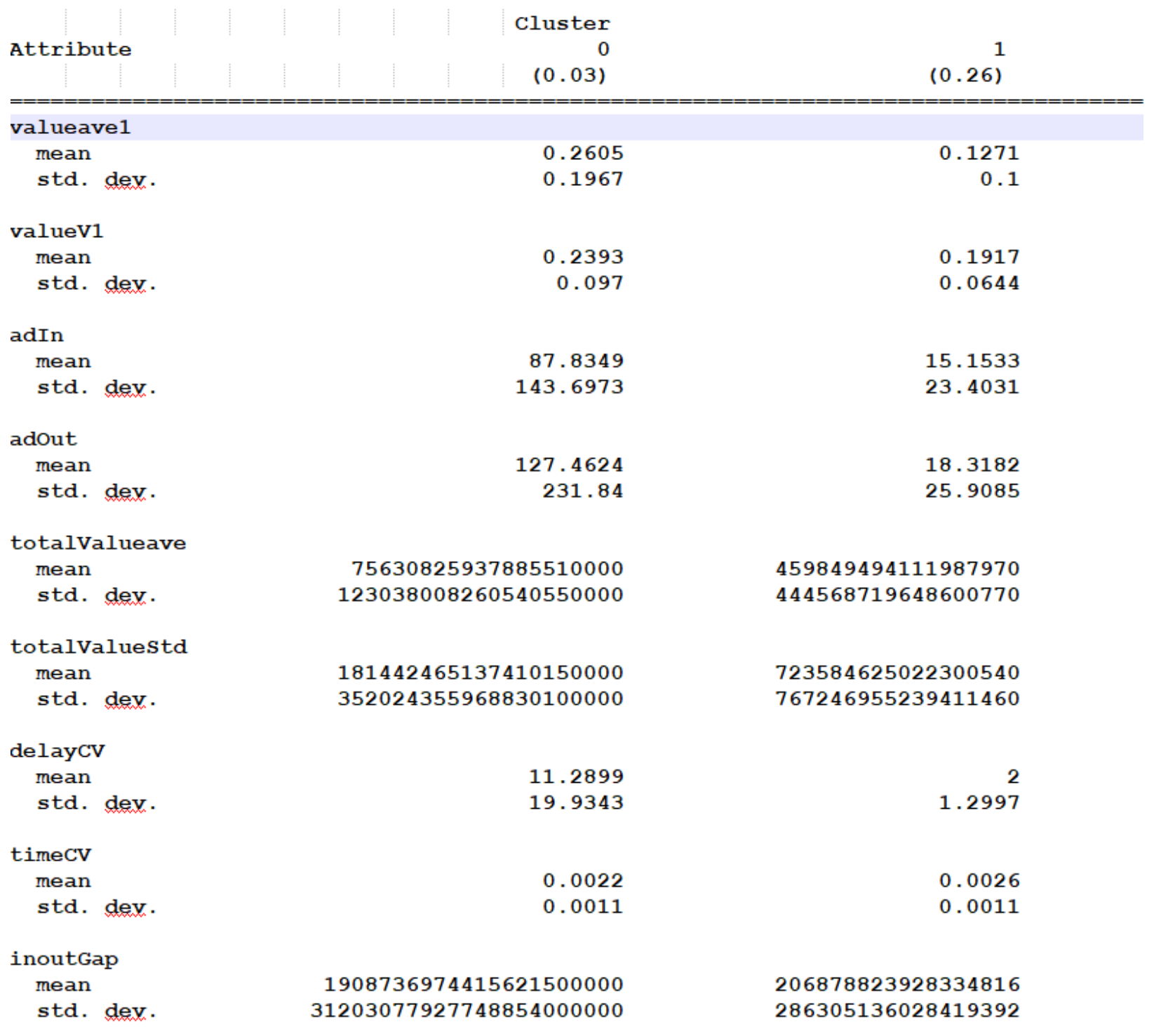


First and second feature features were constructed considering type 3.

Fifth feature features were constructed considering type 4.

Methodology 2

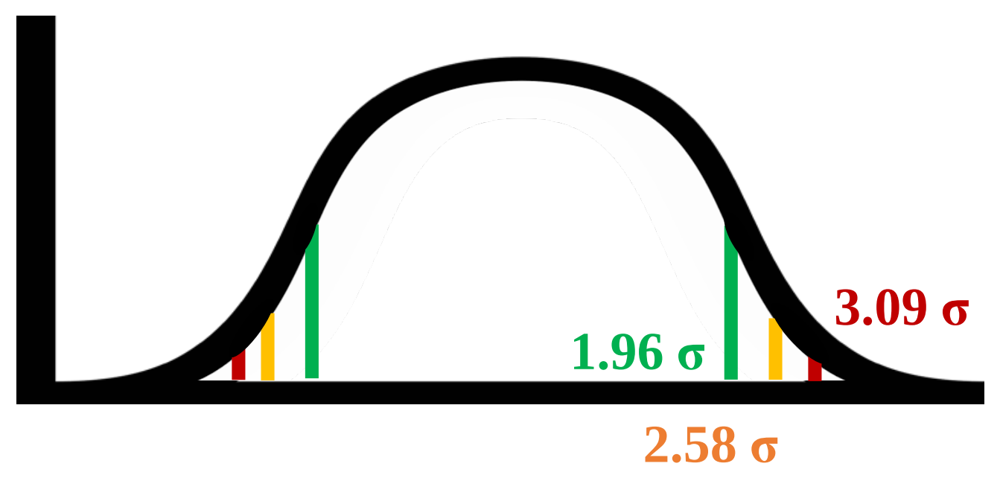
- **Applying Machine Learning Algorithm**



```
                                Cluster
Attribute                             0                          1
                                 (0.03)                     (0.26)
==================================================================
valueave1
  mean                           0.2605                     0.1271
  std. dev.                      0.1967                        0.1

valueV1
  mean                           0.2393                     0.1917
  std. dev.                       0.097                     0.0644

adIn
  mean                          87.8349                    15.1533
  std. dev.                    143.6973                    23.4031

adOut
  mean                         127.4624                    18.3182
  std. dev.                      231.84                    25.9085

totalValueave
  mean             75630825937885510000         459849494111987970
  std. dev.       123038008260540550000         444568719648600770

totalValueStd
  mean            181442465137410150000         723584625022300540
  std. dev.       352024355968830100000         767246955239411460

delayCV
  mean                          11.2899                          2
  std. dev.                     19.9343                     1.2997

timeCV
  mean                           0.0022                     0.0026
  std. dev.                      0.0011                     0.0011

inoutGap
  mean           1908736974415621500000         206878823928334816
  std. dev.     3120307792774885400000 0        286305136028419392
```

- Feature values were obtained based on all transactions of 5790 wallets that have traded more than 20 times with Upbit's wallets
- EM algorithm, which is typically used for unsupervised learning, was applied using all feature values of each wallet.
- The first cluster accounted for as little as 3%, but the other clusters were distributed almost evenly.

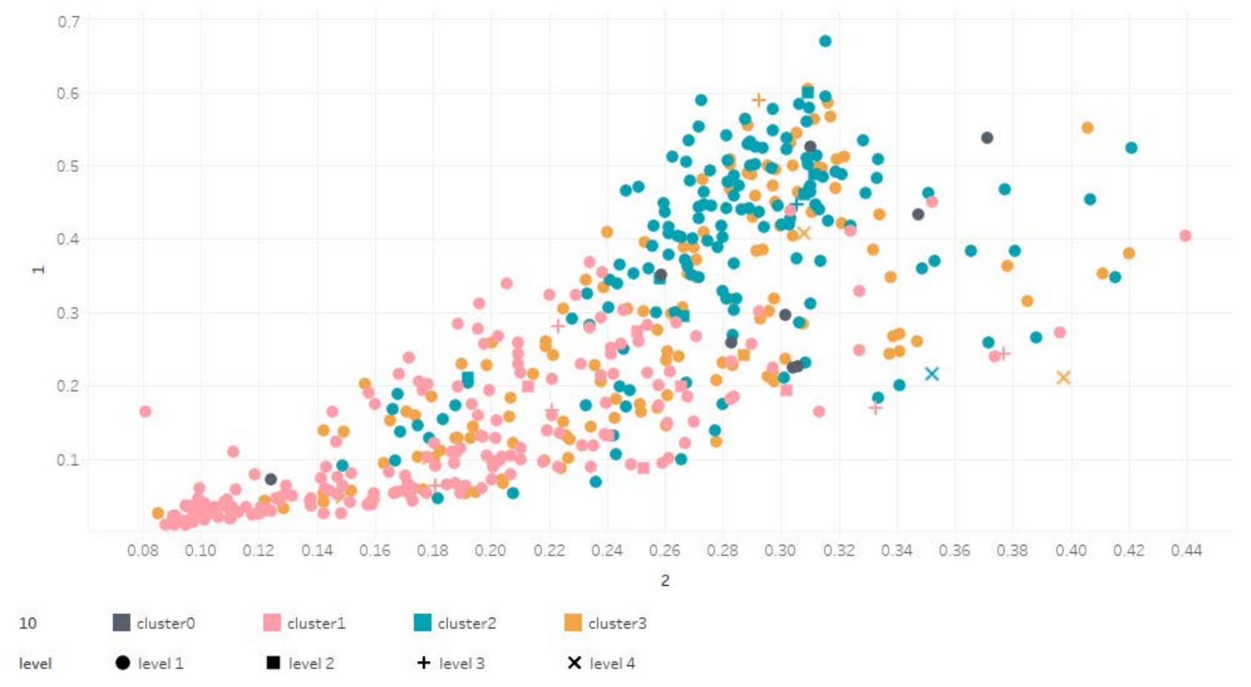
- **Risk Management of Machine Learning Results**



Risk Level	z-score	Cumulative Probability of Cluster Value
1	|z| < 1.96	0% ~ 95%
2	1.96 ≤ |z| < 2.58	95% ~ 99%
3	2.58 ≤ |z| < 3.09	99% ~ 99.8%
4	3.09 ≤ |z|	99.8% ~ 100%

- The average value of each cluster was calculated, and the distance between the feature value and the average value of each data was obtained.
- Based on the distances of each data, the standard deviation was calculated to derive the z-score.
- The risk level was classified according to z-score of each wallet.

Results



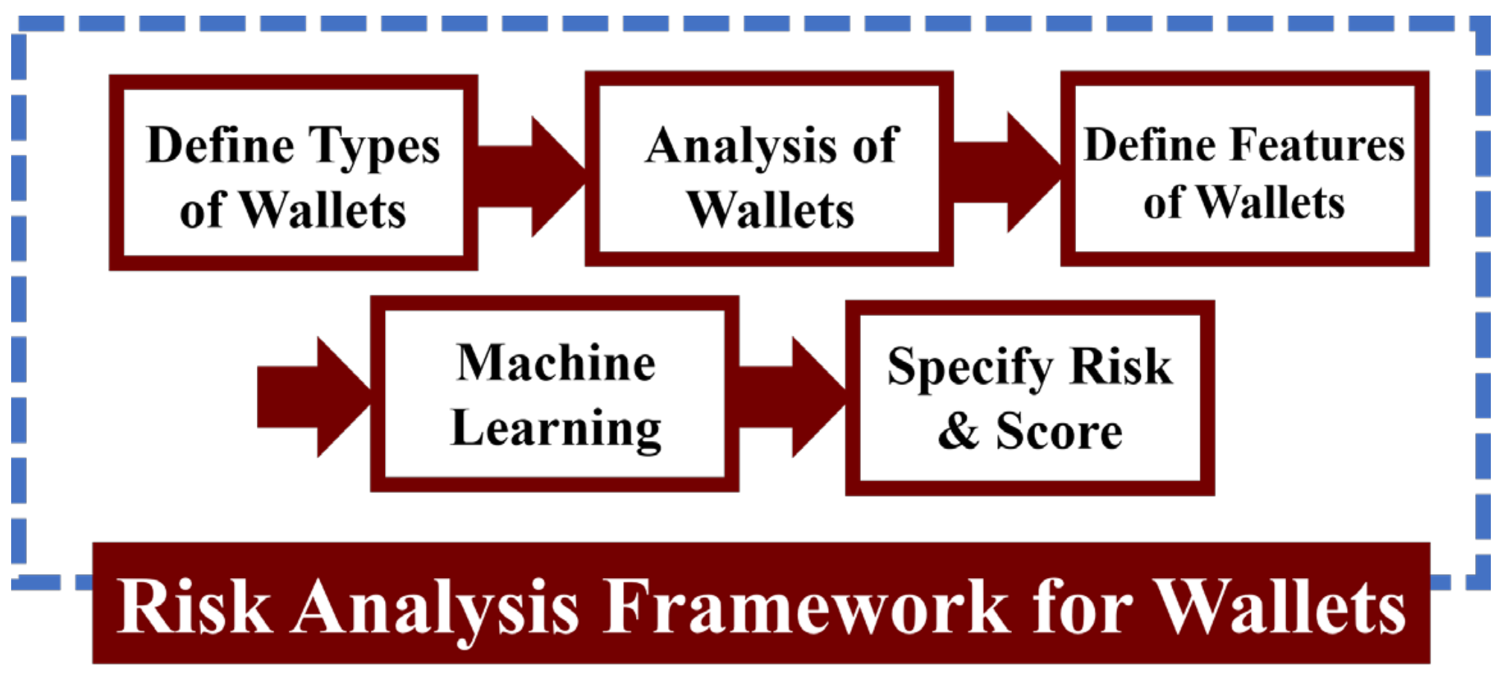
Risk Level	C1 (Type 4)	C2 (Type 1)	C3 (Type 3)	C4 (Type 2)	Total
1	162	1,422	2,625	1,232	5,441
2	1	39	90	21	151
3	0	19	55	5	79
4	1	21	42	55	119
Total	164	1,501	2,812	1,313	**5,790**

- For each cluster, 10 wallets were randomly selected among the wallets belonging to the risk level 1.
- When analyzing the transaction type of this wallet, more than 7 out of 10 coincided with 4 types classified through the preliminary survey.
- The number of total wallets by type and the number by risk level are described in in the table on the left.
- The left graph is an example of clustering two arbitrary features and dividing them by risk level.

Future Work

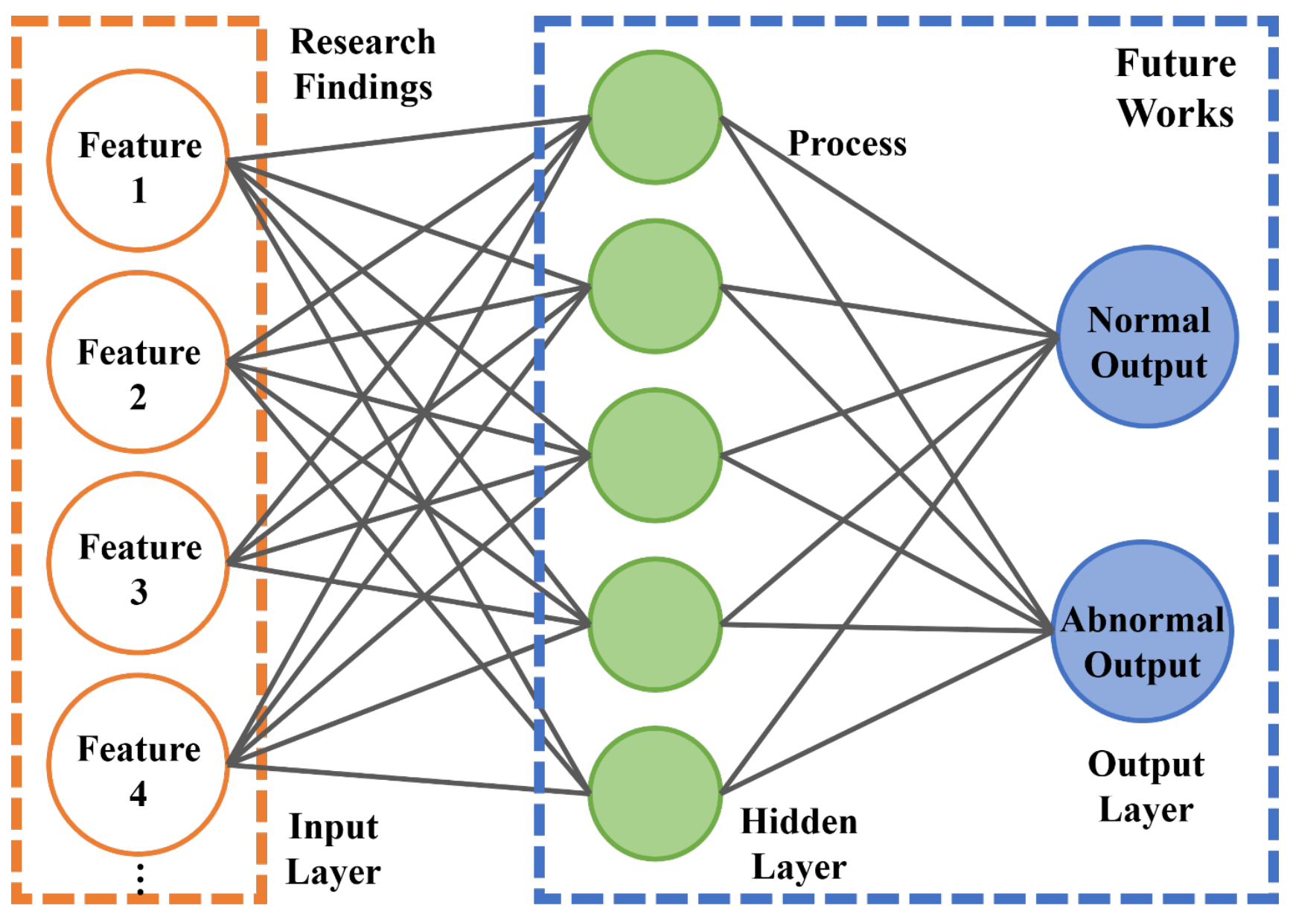
- **Applying Risk Management Process of This Study**

- More accurate results should be obtained by studying more wallets in various exchanges.
- In addition, this methodology can be applied to calculate the risk per cryptocurrency.



- **Applying Deep Learning Algorithm to contents of this study**

- There is a limitation in analyzing a transaction for one cryptocurrency for one exchange.
- We distinguished types and extracted features, so the number of types is insufficient and the number of effective features is limited.
- Deep learning can be used to analyze much more wallets and extract features automatically.



Acknowledgement

This research was supported by the MSIT(Ministry of Science, ICT),Korea, under the ITRC(Information Technology Research Center) support program (IITP-2018-2015-0-00403)supervised by the IITP(Institute for Information &communications Technology Promotion)