PROJECT REPORT

MACHINE LEARNING BASED FRAUD IDENTIFICATION ON E-TRANSACTION

LAUW, AGUNG WIJAYA ALAUY
16.K1.0050

Faculty of Computer Science
Soegijapranata Catholic University
2020

APPROVAL AND RATIFICATION PAGE

MACHINE LEARNING BASED FRAUD IDENTIFICATION ON E-TRANSACTION

by

LAUW, AGUNG WIJAYA ALAUY – 16.K1.0050

This project report has been approved and ratified

by the Faculty of Computer Science on January, 7, 2020

With approval,

Supervisor,

Robertus Setiawan Aji Nugroho, ST., Ph.D
NPP : 05812004264

Examiners,

1.)

Rosita Herawati, ST., MIT
NPP : 05812004263

2.)

Robertus Setiawan Aji Nugroho, ST., Ph.D
NPP : 05812004264

3.)

Y.B. Dwi Setianto, ST., M.Cs
NPP : 05872017021

Dean of Faculty of Computer Science,

Robertus Setiawan Aji Nugroho, ST., MCompIT., PhD
NPP: 058.1.2004.264

STATEMENT OF ORIGINALITY

I, the undersigned:

Name : LAUW, AGUNG WIJAYA ALAUY

ID : 16.K1.0050

Certify that this project was made by myself and not copy or plagiarize from other people, except that in writing expressed to the other article. If it is proven that this project was plagiarizes or copy the other, I am ready to accept a sanction.

Semarang, January, 7, 2020

LAUW, AGUNG WIJAYA ALAUY
16.K1.0050

ABSTRACT

Transaction fraud is fatal to any and all companies, costing them millions of moneys for the fraud cost and prevention. Machine learning to detect transaction fraud commonly uses classification method to determine the legitimacy of said transaction. It can either classify the transaction as a fraudulent or legitimate as a whole, or classify the types of fraud indication said transaction had done. For it to be acceptably fast, the training dataset should not be excessive in quantity but have high enough quality so as the model does not suffer accuracy issue instead.

Some studies include decision tree and random forest in which the random forest almost always yield higher accuracy; understandable considering random forest is and ensemble classifier consisting of many decision trees [1]. While other contenders are SVM and logistic regression, almost all of them boasts high accuracy rate for detection (above 80%) which indicates low complexity when determining a single transaction as far as testing goes. There are several types of fraud in Ecommerce, including but not limited to: Friendly Fraud where a customer (fraudster) complains and claims a refund or purchase, Clean Fraud where a fraudster uses a stolen credit card to make a purchase, Card Testing where the fraudster makes low purchase to validate stolen card information or randomly generated card number on a website with different specific notice like "Incorrect expiration date".

Some prevalent characteristic of a fraud includes but not limited to: customer is a first time customer, customer orders are bigger than average, customer is in an unusual location, customer orders same product but at high quantity, customer ships to multiple addresses, Several purchases with same IP but different card information, too many transaction in a short time span. The previous also includes potential false positives, so to get an accurate result requires a tally of points according to how suspicious or how many of the rules is broken.

Keyword: Classification, Fraud, Machine Learning

TABLE OF CONTENTS

ILLUSTRATION INDEX

INDEX OF TABLES