# *UNDERSAMPLING MAJORITY CLASS* PADA KASUS *IMBALANCED DATASET* DAN APLIKASINYA PADA DETEKSI ANOMALI TRANSAKSI KARTU KREDIT

**SKRIPSI**

**Diajukan untuk memenuhi persyaratan mendapatkan gelar Strata Satu Program Studi Informatika**

**Disusun Oleh:**

**MUHSIN AHADI**

**M0515026**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS SEBELAS MARET**

**2019**

## HALAMAN PERSETUJUAN

**SKRIPSI**

**_UNDERSAMPLING MAJORITY CLASS_ PADA KASUS _IMBALANCED DATASET_ DAN APLIKASINYA PADA DETEKSI ANOMALI TRANSAKSI KARTU KREDIT**

**Disusun Oleh:**

**MUHSIN AHADI**

**M0515026**

Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji pada tanggal ..20..September..2019

| Pembimbing I | Pembimbing II |
| --- | --- |
| Heri Prasetyo S.Kom, M.Sc.Eng., Ph.D. | Dr. Anto Satriyo Nugroho, M.Eng. |
| NIP. 198303022016l001 | NIP. 197010211989111001 |

# HALAMAN PENGESAHAN

**SKRIPSI**

***UNDERSAMPLING MAJORITY CLASS* PADA KASUS *IMBALANCED DATASET* DAN APLIKASINYA PADA DETEKSI ANOMALI TRANSAKSI KARTU KREDIT**

**Disusun Oleh:**

**MUHSIN AHADI**

**M0515026**

Telah dipertahankan di hadapan Dewan Penguji

pada tanggal 20 September 2019

**Susunan Dewan Penguji**

1. Heri Prasetyo S.Kom, M.Sc.Eng., Ph.D. ( )
   NIP. 1983030220161001
2. Dr. Anto Satriyo Nugroho, M.Eng. ( )
   NIP. 197010211989111001
3. Dr. Wisnu Widiarto S.Si., M.T. ( )
   NIP. 197006012008011009
4. Dr. Umi Salamah S.Si., M.Kom. ( )
   NIP. 197002171997022001

**Disahkan Oleh**

Kepala Program Studi Informatika

Dr. Wiharto, S.T., M.Kom.

NIP. 197502102008011005

# MOTTO

"Usaha yang kita tanam pada hari kemarin dan sekarang, adalah buah yang akan dipetik kemudian hari" - *anonymous*

# PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

Ibu, Bapak, Bang Waji, Bang Uti, Mpo Dada, Bang Ibnu, dan semua orang yang telah memberikan dukungan moril dan materiil.

# *UNDER-SAMPLING MAJORITY CLASS* PADA KASUS *IMBALANCED DATASET* DAN APLIKASINYA PADA DETEKSI ANOMALI TRANSAKSI KARTU KREDIT

**MUHSIN AHADI**

**Program Studi Informatika**

**Fakultas Matematika dan Ilmu Pengetahuan Alam**

**Universitas Sebelas Maret**

## ABSTRAK

Kartu kredit adalah salah satu sistem pembayaran yang banyak digunakan di seluruh dunia. Tetapi sistem keamanan kartu kredit masih rentan terjadi *fraud*. Salah satu cara untuk mengatasi *fraud* adalah menggunakan *machine learning* dengan metode *anomaly detection*. *Anomaly detection* adalah metode pencarian pola yang janggal pada suatu keadaan. Metode tersebut dapat dilakukan dengan menggunakan algoritma klasifikasi. Masalah yang sering terjadi dalam klasifikasi adalah terjadinya *imbalanced dataset*, yaitu keadaan *class* tujuan yang diklasifikasi tidak seimbang rasionya. Masalah ini dapat menyebabkan hasil klasifikasi menjadi bias karena *classifier* lebih sering mendeteksi kelas mayoritas dibanding kelas minoritas. *Imbalanced dataset* dapat diatasi dengan menggunakan metode *undersampling*. Penelitian ini bertujuan untuk menangani *imbalanced dataset* dengan metode *undersampling K Nearest Neighbors-Undersampling* pada deteksi anomali transaksi kartu kredit dengan *classifier Support Vector Machine* dan *K-Nearest Neighbors* agar menghasilkan performa prediksi yang optimal. *K Nearest Neighbors-Undersampling* mengurangi kelas mayoritas dari 199013 menjadi 367, 661, 949, 1244, dan 1574 dengan variasi nilai K 4, 8, 12, 16, dan 20 secara berturut-turut. Metode *K Nearest Neighbors* mampu meningkatkan nilai *true positive* dan mengurangi nilai *false negative*. Hasil penelitian menunjukkan performa terbaik dihasilkan 20NN-Und + *Support Vector Machine* parameter C = 1 bekerja paling optimal dengan nilai *precision* 0.95 dan *recall* 0.75 dalam waktu 4.51 detik.

**Kata kunci :** *anomaly detection, undersampling, imbalanced dataset, fraud*

# *UNDER-SAMPLING THE MAJORITY CLASS ON IMBALANCED DATASET AND ITS APPLICATION ON ANOMALY DETECTION CREDIT CARD TRANSACTION*

**MUHSIN AHADI**

**Department of Informatics**

**Faculty of Mathematics and Natural Sciences**

**Sebelas Maret University**

## ABSTRACT

*Credit card is one of the most widely used payment system all over the world. But credit card's security system still vulnerable against fraud. One way to handle fraud is using machine learning with anomaly detection method. Anomaly detection is a method to find strange pattern in some circumstances. It can be done using classification algorithm. The problem that often occurs in classification is imbalanced dataset, which is state of the objective class is in unbalanced ratio.This problem could lead the classification result biased because the classifier detects majority class more often than the minority class. Imbalanced dataset could be handled by using undersampling method. This research's goal is to handle imbalanced dataset using K Nearest Neighbors-Undersampling in anomaly detection credit card transaction using Support Vector Machine and K-Nearest Neighbors classifier in order to produce optimal predictive performance. K Nearest Neighbors-Undersampling reduced majority class from 199013 to 367, 661, 949, 1244, and 1574 with variance of K 4, 8, 12, 16, and 20 consecutively. K Nearest Neighbors could improved true positive value and reduced false negative value. The result showed best performance was produced by 20NN-Und + Support Vector Machine parameter C = 1, with precision 0.95 and recall 0.75 in 4.51 seconds.*

**Keyword:** *anomaly detection, undersampling, imbalanced dataset, fraud*

# KATA PENGANTAR

Puji syukur kepada Allah SWT atas segala limpahan rahmat dan karunia-Nya, penulis dapat menyelesaikan penulisan skripsi ini. Sholawat dan salam senantiasa penulis haturkan kepada Rasulullah SAW sebagai pembimbing seluruh umat manusia

Skripsi ini tidak akan selesai tanpa adanya bantuan dari banyak pihak, karena itu penulis menyampaikan terima kasih kepada:

1. Bapak Dr. Wiharto, S.T., M.Kom. selaku Kepala Program Studi Informatika yang telah memberikan dukungan selama proses perkuliahan dan penyusunan Tugas Akhir.
2. Bapak Heri Prasetyo, S.Kom., M.Sc., Ph.D selaku dosen pembimbing I atas ilmu, waktu, dan masukan yang diberikan, serta kesediaan untuk membimbing penulis selama pelaksanaan Tugas Akhir.
3. Bapak Dr. Anto Satriyo Nugroho, M.Eng. selaku dosen pembimbing II atas ilmu, waktu, dan masukan yang diberikan, serta kesediaan membimbing penulis selama penulisan Tugas Akhir.
4. Orang Tua dan keluarga penulis yang senantiasa memberikan dukungan dan doa kepada penulis untuk menyelesaikan Tugas Akhir.
5. Teman-teman Informatika 2015 atas segala kebersamaan dan dukungannya selama perkuliahan di Program Studi Informatika.
6. Semua pihak yang tidak dapat penulis sebutkan satu persatu.

Penyusunan skripsi ini tentunnya masih ada beberapa kekurangan. Oleh sebab itu, kritik dan saran pembaca sangat diperlukan. Semoga skripsi ini bisa bermanfaat bagi pembaca maupun penulis sendiri.

Surakarta, Agustus 2019

Penulis

# DAFTAR ISI

# DAFTAR TABEL

## DAFTAR LAMPIRAN



# DAFTAR GAMBAR