

Penerapan Pohon Keputusan sebagai Pendeteksi Penipuan pada Transaksi *Online*

William - 13518138

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

13518138@std.stei.itb.ac.id

Abstrak - Dengan semakin berkembangnya teknologi – teknologi serta jaringan komunikasi di bidang Ilmu Informatika, kasus – kasus penipuan semakin marak terjadi. Penipuan – penipuan tersebut banyak menyebabkan kerugian di bidang finansial. Penipuan yang paling sering terjadi di masa modern ini adalah penipuan transaksi *online*. Maraknya kasus serupa menyebabkan pendeteksi tipuan (*fraud detection*) sangatlah dalam aplikasi kehidupan sehari – hari terutama dalam transaksi menggunakan transaksi *online*. Analisis terhadap kasus – kasus penipuan dapat dilakukan dengan menggunakan pohon keputusan (*decision tree*). Dalam makalah ini, penulis akan membuat suatu analisis terhadap deteksi penipuan penggunaan transaksi *online* dengan menggunakan metode pohon keputusan dan pemanfaatannya dalam perumusan strategi penggunaan dan pengembangan produk sejenis transaksi *online*.

Kata Kunci—deteksi penipuan, transaksi *online*, pohon keputusan .

I. PENDAHULUAN

Penipuan merupakan suatu perbuatan yang salah atau bersifat kriminal. Penipuan umumnya dilakukan untuk mendapatkan akses milik pihak tanpa akses legal. Untuk mencegah peristiwa penipuan dan kerugian yang ditimbulkannya adalah pencegahan penipuan (*fraud prevention*) dan sistem pendeteksi penipuan (*fraud detection system*). Sistem pendeteksi penipuan merupakan sistem yang aktif dalam mencari penipuan – penipuan yang telah melewati sistem pencegahan penipuan. Tidak ada yang tahu bagaimana seseorang dapat membobol sistem pencegahan penipuan tersebut, namun seiring dengan perkembangan teknologi yang sangat pesat, para pelaku penipuan telah menemukan berbagai cara untuk melewati sistem pencegahan penipuan. Oleh sebab itu, sistem pendeteksi penipuan sangatlah diperlukan di era modern ini. Maka dari itu, tujuan utama sistem pendeteksi penipuan adalah melakukan pengecekan terhadap seluruh transaksi yang terjadi dan mengukur besarnya kemungkinan suatu transaksi merupakan penipuan. Salah satu bentuk penipuan yang paling sering terjadi di era serba teknologi ini merupakan penipuan pada transaksi *online*, penipuan jaringan telekomunikasi, dan penipuan dalam penggunaan kartu kredit.

Dalam makalah ini, penulis ingin membahas tentang penipuan – penipuan pada penggunaan transaksi *online*. Dengan berkembangnya era industri informasi saat ini, informasi para pengguna transaksi *online* telah banyak tersebar

dan dapat diakses dengan mudah oleh siapapun. Dengan memanfaatkan berbagai teknologi yang telah ada, para pelaku penipuan dapat mengakses informasi – informasi pengguna tanpa sepengetahuan pemilik informasi. Informasi – informasi tersebut kemudian dapat digunakan untuk melakukan tindakan – tindakan penipuan. Meskipun telah banyak sistem pencegah penipuan yang telah diterapkan seperti pengecekan konfirmasi setiap kali melakukan pengguna melakukan transaksi ataupun konfirmasi PIN ketika melakukan transaksi, masih banyak kasus – kasus penipuan transaksi *online* yang terjadi karena kurangnya kemampuan sistem pendeteksi penipuan dalam mendeteksi metode – metode baru dalam tindakan penipuan.

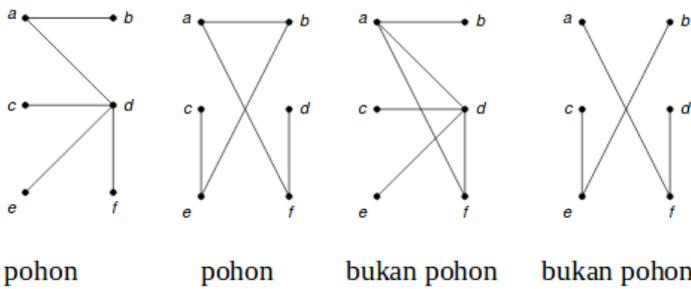
Dalam makalah ini penulis akan melakukan analisis terhadap sistem pendeteksi penipuan pada transaksi *online* dengan mengimplementasikan program yang menggunakan pohon keputusan (*decision tree*). Hasil analisis terhadap sistem pendeteksi dengan memakai *decision tree* ini nantinya dapat digunakan sebagai tolak ukur apakah metode ini dapat dimanfaatkan untuk pengimplementasian sistem pendeteksi penipuan dalam bidang penggunaan transaksi *online*.

II. LANDASAN TEORI

A. Pohon

Pohon adalah suatu graf tak – berarah terhubung yang tidak mengandung sirkuit. Pohon memiliki beberapa sifat unik. Misalkan terdapat suatu graf tak-berarah $G = (V, E)$ sederhana dengan jumlah simpul sebanyak n . Maka, semua pernyataan di bawah ini adalah ekuivalen :

1. G adalah pohon
 2. Setiap pasang simpul di dalam G terhubung dengan lintasan tunggal.
 3. G terhubung dan memiliki $m = n-1$ buah sisi.
 4. G tidak mengandung sirkuit dan memiliki $m = n - 1$ buah sisi.
 5. G tidak mengandung sirkuit dan penambahan satu sisi pada graf akan membuat hanya satu sirkuit.
 6. G terhubung dan semua sisinya adalah jembatan
- Pernyataan – pernyataan yang disebutkan di atas dapat dianggap sebagai definisi lain dari pohon.



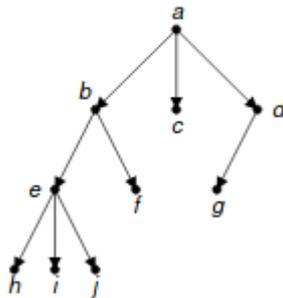
Gambar 2.1 Contoh graf yang membentuk pohon

Sumber: [http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2013-2014/Pohon%20\(2013\).pdf](http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2013-2014/Pohon%20(2013).pdf) diakses pada 29 November pukul 14:39 WIB

Pada gambar 1 di atas, G3 dan G4 bukanlah pohon. G3 mengandung sirkuit (a – d- f – a) dan G4 juga bukan pohon karena tidak terhubung. Ada banyak implementasi pohon yang berdasar pada definisi pohon di atas, seperti pohon merentang, pohon merentang minimum, dan lain lain. Namun untuk makalah ini, jenis pohon yang akan dibahas adalah pohon berakar.

B. Pohon Berakar

Pohon yang suatu simpulnya diperlakukan sebagai akar dan sisi – sisinya diberikan arah sehingga bergerak seperti menjauhi akar (*root*) dan menjadi suatu graf berarah dinamakan sebagai pohon berakar (*rooted tree*).



Gambar 2.2 Contoh pohon berakar

Sumber: [http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2013-2014/Pohon%20\(2013\).pdf](http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2013-2014/Pohon%20(2013).pdf) diakses pada 29 November 14:48 WIB

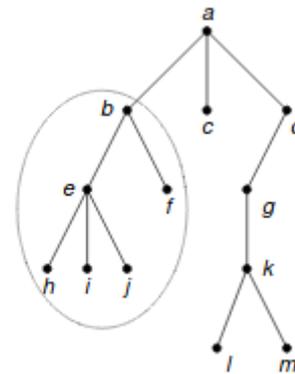
Pohon berakar memiliki beberapa jenis terminologi – terminologi penting antara lain :

1. Anak (*child*) dan Orangtua (*parent*)
Jika terdapat simpul a dan b, dimana arah pohon dari simpul a menuju simpul b, maka a disebut orang tua dari simpul b dan sebaliknya simpul b disebut anak dari simpul a.
2. Lintasan (*path*)
Jalan adalah barisan berhingga dari titik dan sisi dimulai dan diakhiri dengan titik sedemikian sehingga setiap sisi menempel dengan titik sebelum dan sesudahnya. Lintasan adalah jalan yang memiliki dan melewati titik yang berbeda.
3. Saudara kandung

Dua buah simpul atau lebih dikatakan sebagai saudara kandung jika simpul – simpul tersebut memiliki orang tua yang sama.

4. Upapohon

Upapohon dapat juga disebut sebagai subpohon dari suatu pohon. Jumlah subpohon / upapohon yang dapat terbentuk dari suatu pohon dihitung berdasarkan jumlah anak pohon tersebut.



Gambar 2.3 Contoh Upapohon

Sumber: [http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2013-2014/Pohon%20\(2013\).pdf](http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2013-2014/Pohon%20(2013).pdf) diakses pada 29 November 2019 pukul 15:05 WIB

5. Derajat

Derajat pada pohon memiliki definisi yang sedikit berbeda dengan simpul pada graf. Derajat suatu simpul pada pohon berakar adalah jumlah upapohon (atau jumlah anak) pada simpul tersebut.

6. Daun (*leaf*)

Simpul pada suatu pohon berakar yang memiliki derajat 0 (nol) disebut sebagai daun dari suatu pohon berakar.

7. Simpul dalam (*internal nodes*)

Simpul – simpul dalam suatu pohon berakar yang memiliki anak disebut sebagai simpul dalam.

8. Aras (*level*) atau Tingkat

Aras dari akar adalah 0. Untuk simpul – simpul selain akar arasnya adalah 1 + panjang lintasan dari akar ke simpul tersebut.

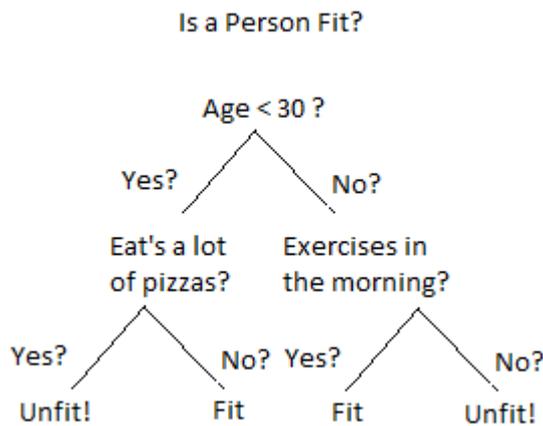
9. Tinggi atau Kedalaman

Merupakan aras maksimum dari suatu pohon.

C. Pohon keputusan

Salah satu bentuk implementasi dari pohon berakar adalah pohon keputusan. Pohon keputusan sendiri merupakan suatu representasi sederhana dari metode – metode pengklasifikasian terhadap suatu hal. Setiap simpul (*node*) dari decision tree merepresentasikan bentuk pertanyaan untuk menentukan apakah suatu fakta merupakan bagian dari simpul anak sebelah kiri (*left node*) atau simpul anak sebelah kanan (*right node*). Pertanyaan – pertanyaan yang terdapat di simpul pohon keputusan bergantung terhadap hal apa yang ingin dibandingkan saat data berada di simpul tersebut. Metode pohon keputusan dalam mengklasifikasikan suatu hal sering dikaitkan dengan paradigma “Divide and Conquer”, dimana

pohon keputusan akan berusaha terus membagi permasalahan kedalam 2 jawaban (Yes / No).



Gambar 2.4 Contoh bentuk pohon keputusan

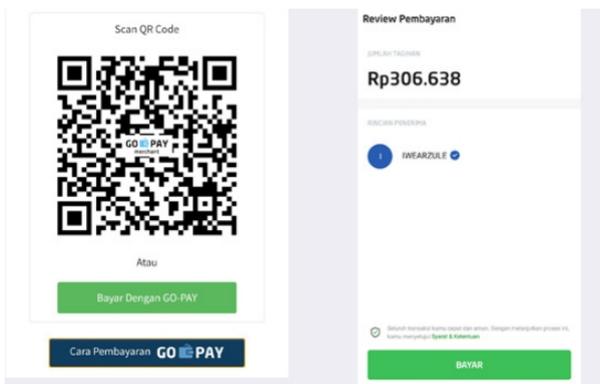
Sumber : <https://www.xoriant.com/> diakses pada 1 Desember 2019 pukul 10:39 WIB

D. Transaksi Online

Transaksi *online* adalah suatu alat pembayaran menggunakan internet yang banyak digunakan di dunia pada era modern ini. Pengguna internet dapat membeli barang di toko – toko *online* berlisensi, seperti Tokopedia, Bukalapak, eBay dengan metode transaksi secara *online*. Salah satu metode transaksi *online* yang telah mendunia adalah PayPal. PayPal mengatasi kekurangan dana dan pengiriman uang tradisional seperti cek yang prosesnya dapat memakan waktu yang cukup lama.

Di Indonesia sendiri, sudah muncul berbagai metode untuk melakukan transaksi secara *online*. Terdapat Gopay, Ovo, Dana dan masih banyak lagi sistem – sistem pembayaran online yang telah berkembang di Indonesia. Sistem pembayaran ini memberikan kemudahan bagi penggunanya, seperti tidak perlu membawa uang cash kemana – mana, serta mempermudah transaksi seperti ketika di pintu tol.

Dari banyak keunggulan – keunggulan yang diberikan dengan memakai transaksi *online*, muncul permasalahan – permasalahan baru. Dikarenakan jumlah pemakai yang bertambah banyak seiring dengan perkembangan teknologi, kasus – kasus penipuan yang melibatkan transaksi secara *online* juga meningkat. Mulai dari pencurian data – data pribadi hingga transaksi illegal banyak terjadi di era teknologi ini.



Gambar 2.5 Contoh Transaksi Online dengan Gopay

III. PENERAPAN POHON KEPUTUSAN PADA ANALISIS FINANSIAL UNTUK MENDETEKSI PENIPUAN TRANSAKSI ONLINE

A. Deskripsi Data

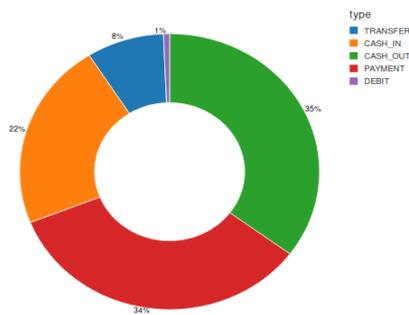
Proses pertama yang diperlukan untuk melakukan analisis finansial guna mendeteksi penipuan adalah dengan mengumpulkan data – data yang penting dan dapat dijadikan referensi untuk menguji pohon keputusan. Dalam kasus ini, data yang diperoleh penulis sudah dalam bentuk yang bersifat *structured*. Oleh sebab itu, data dapat langsung digunakan untuk menganalisa penipuan – penipuan pada transaksi *online*.

Data yang diperoleh merupakan hasil simulasi dari sistem PaySim yang digunakan untuk mensimulasikan transaksi *online* berdasarkan sampel – sampel data transaksi *online* asli. Data yang digunakan untuk melatih model berjumlah ¼ kali data sesungguhnya untuk mempermudah proses komputasi serta menyederhanakan permasalahan. Data sesungguhnya berjumlah 6353307 data *entry*. Data sesungguhnya dapat diakses melalui *platform* Kaggle.com : <https://www.kaggle.com/ntnu-testimon/paysim1>.

Beberapa definisi penting terhadap fitur - fitur data hasil sintesa PaySim sebagai berikut :

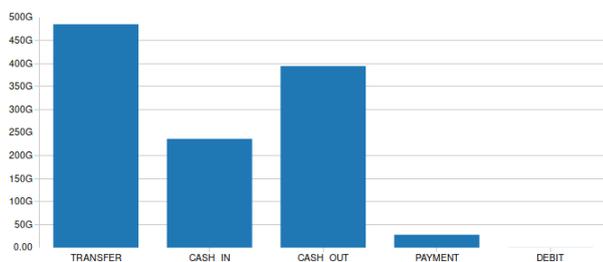
Nama fitur	Penjelasan
Step	Mengasumsikan waktu sesungguhnya yang ada di dunia nyata
Type	Terdapat 4 jenis tipe transaksi pada data ini : 1. CASH – IN 2. CASH – OUT 3. DEBIT 4. PAYMENT 5. TRANSFER
Amount	Jumlah uang dalam suatu transaksi
NameOrig	Nama customer yang menginisiasikan suatu transaksi
OldbalanceOrg	Jumlah uang sebelum transaksi
NewBalanceOrg	Jumlah uang setelah transaksi
NameDest	Nama customer yang menjadi penerima dari hasil transaksi tersebut
OldbalanceDest	Jumlah uang resipien sebelum inisiasi transaksi dilakukan
NewbalanceDest	Jumlah uang resipien setelah transaksi selesai dilakukan

Perbandingan jenis – jenis transaksi dalam data :



Gambar 3.1 Jenis – jenis transaksi online pada dataset

Perbandingan jumlah uang yang terlibat dalam masing – masing transaksi :



Gambar 3.2 Jumlah uang dalam transaksi pada masing – masing jenis transaksi (dalam satuan dollar)

B. Decision Tree Classifier

Setelah data training siap untuk dianalisa, maka kita dapat menggunakan pohon keputusan untuk mengklasifikasikan jenis – jenis transaksi ke dalam 2 (dua) macam kategori, yakni kategori transaksi legal dan transaksi illegal (penipuan) dengan menggunakan pohon keputusan.

Pertama – tama, penulis membagi data yang dimiliki kedalam 2 bagian, yakni data untuk *training* dan data untuk *testing*. Pembagian data menjadi 2 jenis ini dimaksudkan agar hasil pemodelan pohon keputusan yang diperoleh dari hasil training dapat diuji kebenarannya terhadap data yang diperoleh dari sumber yang sama, namun tidak pernah dilihat sebelumnya.

Dikarenakan permasalahan komputasi, penulis hanya akan menggunakan ¼ dari total data yang tersedia. Hal ini penulis lakukan agar waktu pelatihan model tidak terlalu lama dan tidak membebani kemampuan komputasi yang dimiliki penulis.

Kemudian, data hasil training diteliti lebih lanjut. Dikarenakan di tahap sebelumnya proses ini telah dilakukan, maka penulis akan melewati tahap ini.

Selanjutnya, penulis melatih model dengan data training yang telah dipisahkan di tahap sebelumnya. Dengan memakai data training tersebut, penulis memperoleh model decision tree sebagai berikut. Ilustrasi model yang diberikan hanya

menggambarkan bagaimana suatu model menentukan suatu transaksi bersifat *fraud* (penipuan) atau tidak.

```
from pyspark.ml import Pipeline
from pyspark.ml.feature import StringIndexer
from pyspark.ml.feature import OneHotEncoderEstimator
from pyspark.ml.feature import VectorAssembler
from pyspark.ml.classification import DecisionTreeClassifier

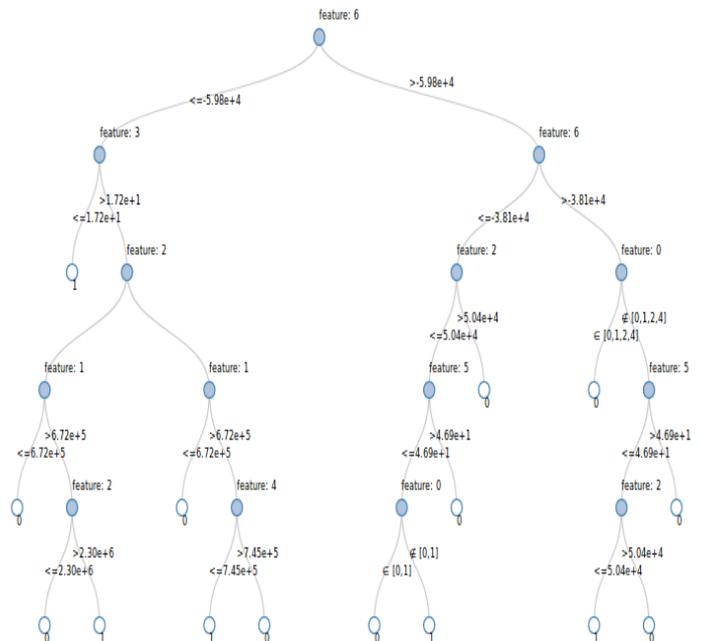
# Encodes a string column of labels to a column of label indices
indexer = StringIndexer(inputCol = "type", outputCol = "typeIndexed")

# VectorAssembler is a transformer that combines a given list of columns into a single vector column
va = VectorAssembler(inputCols = ["typeIndexed", "amount", "oldbalanceOrg", "newbalanceOrig", "oldbalanceDest", "newbalanceDest", "orgDiff", "destDiff"], outputCol = "features")

# Using the DecisionTree classifier model
dt = DecisionTreeClassifier(labelCol = "label", featuresCol = "features", seed = 54321, maxDepth = 5)

# Create our pipeline stages
pipeline = Pipeline(stages=[indexer, va, dt])
```

Gambar 3.3 Potongan kode untuk melatih model dengan PySpark



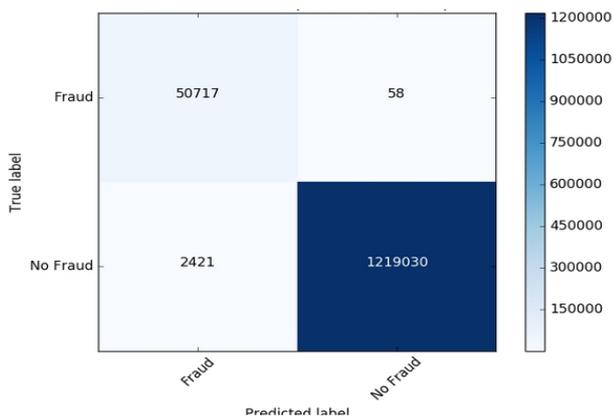
Gambar 3.4 Hasil Pohon Keputusan menggunakan modul Pyspark pada Python.

Dengan menggunakan model *decision tree* ini, kita dapat mengetahui apakah suatu transaksi bersifat *fraud* atau tidak. Dengan memberikan model data – data yang diperlukan (seperti Type, Amount, OldbalanceOrg, NewBalanceOrg, OldbalanceDest, NewbalanceDest), maka model dapat memprediksi apakah transaksi tersebut *fraud* atau tidak

C. Hasil Eksperimen

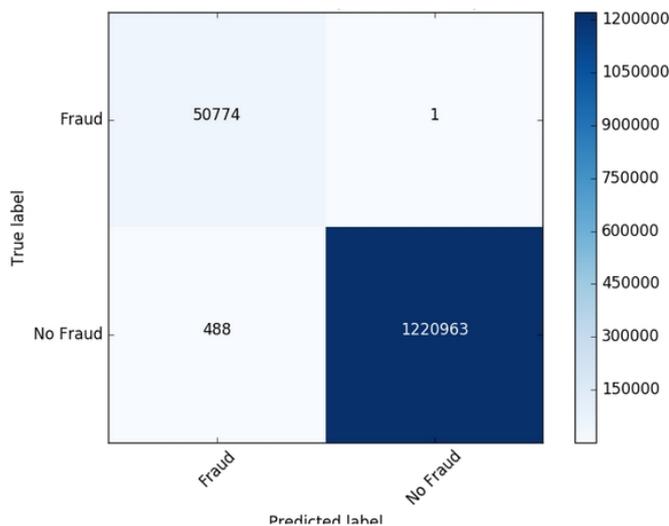
Dari model decision tree yang dihasilkan pada tahap sebelumnya, kita dapat melakukan pengujian terhadap data *sample* yang telah disisihkan sebelumnya oleh penulis. Masing – masing dataset berisikan 1272226 data entry

Hasil pengujian pada dataset sampel 1:



Gambar 3.5 Gambar Confussion matrix untuk dataset uji 1

Hasil pengujian pada dataset sampel 2:



Gambar 3.6 Gambar Confussion matrix untuk dataset uji 2

Selain memeriksa hasil dari confusion matrix, penulis juga menguji model *machine learning* yang dihasilkan pada tahap sebelumnya dengan metrik PR (Precision – Recall Curve) dan AUC (Area Under ROC Curve).

Metric	Data Test 1	Data Test 2
PR	0.9539170535377599	0.9904709171789063
AUC	0.9984378183482442	0.9997903902204509

D. Analisa Hasil

Pada uji kasus pertama dengan data sebanyak 1272226 data *entry*, diperoleh jumlah kesalahan prediksi sebanyak 2479 data *entry*. Hal ini menunjukkan tingkat kesalahan sebesar 0.1948% atau akurasi model untuk data uji 1 sebesar 99.805%.

Pada uji kasus kedua dengan data sebanyak 1272226 data

entry, diperoleh jumlah kesalahan prediksi sebesar 489 data *entry*. Hal ini menunjukkan tingkat kesalahan sebesar 0.0384% atau akurasi model untuk data uji 2 sebesar 99.96%.

Berdasarkan hasil yang diperoleh, kita dapat menyimpulkan bahwa nilai akurasi yang didapat dari model pohon keputusan (*decision tree*) cukup baik untuk dalam menentukan apakah suatu transaksi *online* bersifat *fraud* (penipuan) atau tidak. Namun, perlu diperhatikan juga bahwa terdapat perbedaan tingkat akurasi yang cukup jauh antara hasil *testing* pada data uji pertama dan hasil *testing* pada data uji kedua. Kedua hal ini dapat menjadi indikasi dari kemungkinan *overfitting* yang dialami suatu model hasil *machine learning*. Tetapi, masalah – masalah seperti *overfitting* dapat terselesaikan apabila model *decision tree* diberikan lebih banyak data selama proses pelatihannya. Melalui indikator – indikator seperti metrik PR dan AUC, kita dapat menentukan apabila suatu model, dalam hal ini *decision tree* yang kita pakai *overfitting* atau tidak.

Hasil pemodelan dan analisis ini kedepannya dapat dijadikan sebagai indikator apakah suatu transaksi yang dilakukan secara *online* merupakan suatu *fraud* / penipuan atau tidak, yang jelas akan mempermudah pekerjaan bagi kedua belah pihak baik penginisiasi transaksi maupun resipien dari transaksi tersebut. Para ahli pendeteksi penipuan transaksi *online* tidak perlu mengecek seluruh transaksi dan hanya perlu mengecek transaksi – transaksi yang memiliki kemungkinan adalah suatu transaksi penipuan (*fraud*). Hal ini akan meningkatkan efisiensi dari proses penindakanlanjutan transaksi penipuan serta mempercepat proses penanganan kerugian yang terjadi akibat transaksi penipuan tersebut.

IV. PEMANFAATAN FRAUD DETECTION DALAM TRANSAKSI DARING

A. Menjaga keamanan data pribadi konsumen

Ketika transaksi *online* mulai diperkenalkan, salah satu hal yang menjadi perhatian dari calon pengguna adalah keamanan data pribadi. Jika transaksi *online* dinyatakan aman melalui penerapan metode ini, maka konsumen dapat bernapas lega sebab artinya data – data pribadinya telah terjamin keamanannya.

B. Aktivitas Pembayaran Tunai Menurun, Non Tunai Meningkat

Bila penggunaan uang tunai sebagai alat pembayaran dapat ditekan, maka produksinya yang membebaskan anggaran tinggi pada negara juga dapat ditekan. Dengan penerapan metode *fraud detection* yang diterapkan di makalah ini, diharapkan keamanan dalam bertransaksi *online* dapat terus meningkat sehingga secara otomatis masyarakat akan beralih pada transaksi *online* dan mampu meningkatkan literasi keuangan negara.

C. Kemudahan dalam Bertransaksi

Semakin meningkatnya kepercayaan masyarakat bahwa transaksi *online* itu aman, otomatis juga berimbas pada jumlah pembelanjaan *online*. Manfaat ini tentu sangat dirasakan bagi konsumen di mana mereka tidak perlu lagi kebingungan ketika

menginginkan sebuah item dari luar maupun internasional.

V. KESIMPULAN

Konsep Graf dan Pohon dalam ilmu Matematika Diskrit ternyata sangat aplikatif. Salah satu jenis pohon yang kegunaannya cukup meningkat pesat dan semakin meluas adalah pohon keputusan. Pohon keputusan dapat digunakan untuk mendeteksi penipuan – penipuan yang mungkin terjadi saat melakukan transaksi secara *online*. Kasus – kasus penipuan ini sangatlah penting untuk diperhatikan sebagai indikator untuk menjaga keamanan dalam bertransaksi secara online. Strategi – strategi penanganan terhadap penipuan transaksi *online* juga dapat dikembangkan melalui analisis terhadap kasus – kasus penipuan yang terjadi transaksi *online*.

VI. PENUTUP

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Kuasa atas berkat dan rahmat-Nya penulis dapat menyelesaikan makalah ini dengan baik dan pada waktu yang tepat. Penulis mengucapkan terima kasih kepada kedua orang tua serta teman teman yang terus memberikan dukungan baik secara moral maupun doa. Ucapan terima kasih turut penulis sampaikan kepada para dosen yang telah membimbing penulisan makalah ini. Akhir kata, penulis memohon maaf apabila terdapat kekurangan dan kesalahan kata dalam makalah ini. Penulis berharap makalah ini dapat digunakan sebaik – baiknya dan dikembangkan lagi sehingga dapat menghasilkan manfaat untuk masyarakat luas.

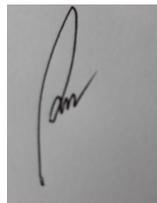
DAFTAR PUSTAKA

- [1] Rinaldi Munir, Diktat Kuliah IF2120 : Matematika Diskrit. Bandung: Program Studi Teknik Informatika Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung, 2006.
- [2] <https://heartbeat.fritz.ai/introduction-to-pytorch-for-deep-learning-5b437cea90ac> – Tutorial PyTorch dan PySpark
- [3] <https://heartbeat.fritz.ai/evaluation-metrics-for-machine-learning-models-d42138496366> – Metode Evaluasi untuk Model Machine Learning
- [4] <https://finance.detik.com/berita-ekonomi-bisnis/d-3711580/5-risiko-berbelanja-online-yang-perlu-anda-waspadai> – Bahaya Transaksi Online

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 4 Desember 2019



William / 13518138