

**PERANCANGAN SISTEM PREDIKSI CHURN PELANGGAN
PERUSAHAAN FASHION DENGAN MEMANFAATKAN
ALGORITMA MACHINE LEARNING**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

KHIAS NURLATIF ARI SUBEKTI

19.11.2776

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**PERANCANGAN SISTEM PREDIKSI CHURN PELANGGAN
PERUSAHAAN FASHION DENGAN MEMANFAATKAN
ALGORITMA MACHINE LEARNING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

KHIAS NURLATIF ARI SUBEKTI

19.11.2776

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**PERANCANGAN SISTEM PREDIKSI CHURN PELANGGAN
PERUSAHAAN FASHION DENGAN MEMANFAATKAN
ALGORITMA MACHINE LEARNING**

yang disusun dan diajukan oleh

Khias Nurlatif Ari Subekti

19.11.2776

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Februari 2023

Dosen Pembimbing,



Ferry Wahyu Wibowo, S.Si, M.Cs.

NIK. 190302235

HALAMAN PENGESAHAN

SKRIPSI

**PERANCANGAN SISTEM PREDIKSI CHURN PELANGGAN
PERUSAHAAN FASHION DENGAN MEMANFAATKAN
ALGORITMA MACHINE LEARNING**

yang disusun dan diajukan oleh

Khias Nurlatif Ari Subekti

19.11.2776

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Februari 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Erni Seniwati, S.Kom, M.Cs
NIK. 190302231

Yuli Astuti, M.Kom
NIK. 190302146

Ferry Wahyu Wibowo, S.Si, M.Cs
NIK. 190302235



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Februari 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Khias Nurlatif Ari Subekti
NIM : 19.11.2776

Menyatakan bahwa Skripsi dengan judul berikut:

PERANCANGAN SISTEM PREDIKSI CHURN PELANGGAN PERUSAHAAN FASHION DENGAN MEMANFAATKAN ALGORITMA MACHINE LEARNING

Dosen Pembimbing : Ferry Wahyu Wibowo, S.Si, M.Cs.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Februari 2023

Yang Menyatakan,



Khias Nurlatif Ari Subekti

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Allah Subhanahu Wata'ala, yang telah memberikan nikmat ilmu, kesehatan dan kecukupan rezeki sehingga penulis mampu menyelesaikan skripsi ini dengan judul “Perancangan Sistem Prediksi Churn Pelanggan Perusahaan Fashion Dengan Memanfaatkan Algoritma Machine Learning”, sebagai salah satu syarat dalam menyelesaikan Program Sarjana (S1) Jurusan Informatika.

Penulis sadar bahwa jika tanpa bantuan dan dukungan dari berbagai pihak, penyusunan skripsi ini tidak akan berjalan lancar. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih sebesar-besarnya kepada seluruh pihak terkait yang telah memberikan dukungan dan motivasi dalam penyusunan skripsi ini.

Pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Kedua orang tua saya yang selalu memberikan dukungan baik secara material maupun doa.
2. Rektor Universitas Amikom Yogyakarta yang telah memberikan fasilitas dan kesempatan kepada penulis untuk menyelesaikan studi S1.
3. Bapak Ferry Wahyu Wibowo, S.Si, M.Cs. selaku dosen pembimbing yang telah memberikan arahan dan bimbingan dalam pengerjaan skripsi ini.
4. Serta teman-teman kelas informatika IF 03 yang saling berbagi ilmu pengetahuan selama perkuliahan berlangsung.

Selain itu penulis juga menyadari bahwa dalam skripsi ini masih banyak terdapat kekurangan. Maka dari itu, penulis meminta maaf dan mengharapkan saran dan kritik dari para pembaca agar penulis dapat membuat karya yang lebih baik di masa depan. Penulis berharap skripsi ini bisa bermanfaat bagi para pembaca pada umumnya.

Yogyakarta, 20 Februari 2023

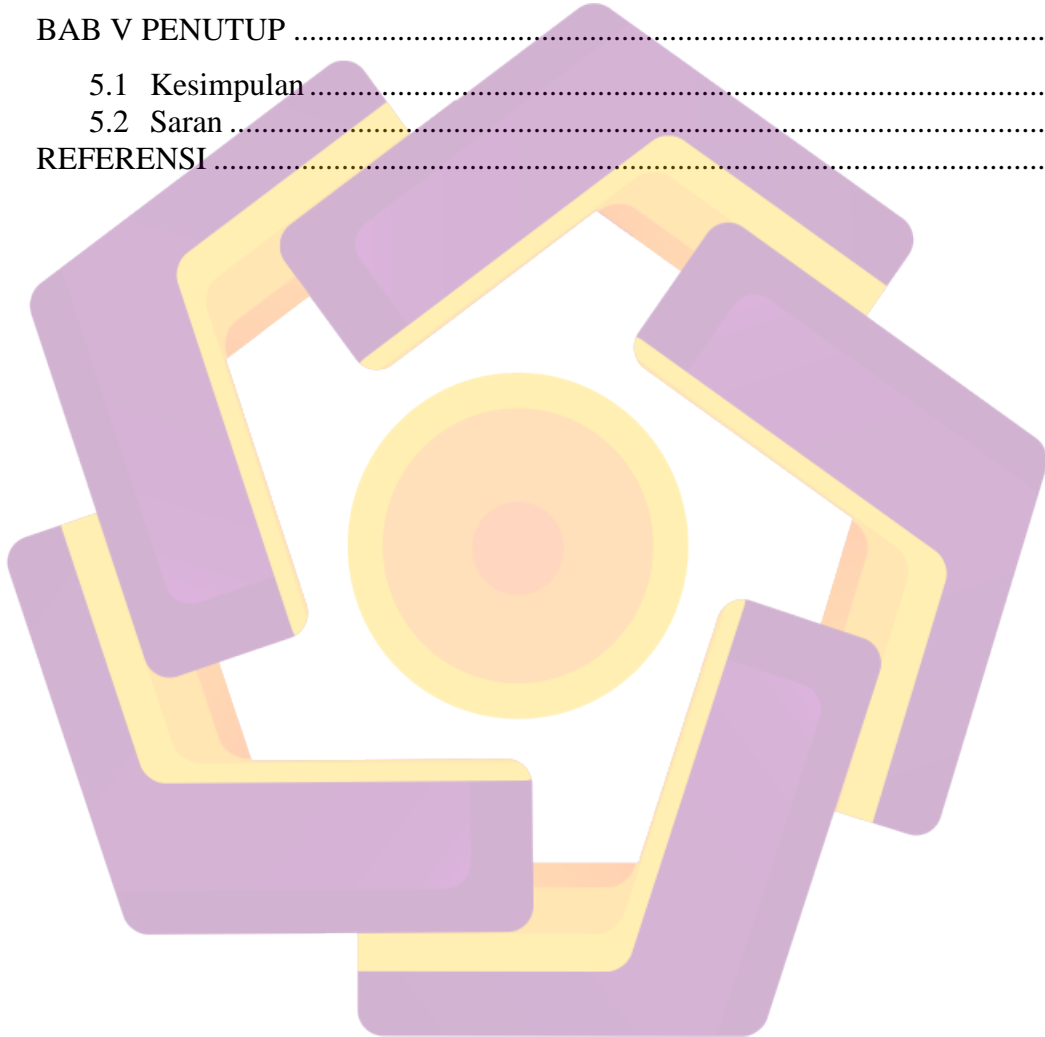
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR PERSAMAAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	4
2.1 Studi Literatur	4
2.2 Dasar Teori	12
2.2.1 CRISP-DM.....	12
2.2.2 <i>Volume, Variety, dan Velocity</i>	13
2.2.3 Data Kategori dan Data Numerik.....	14
2.2.4 <i>Supervised Learning dan Unsupervised Learning</i>	14
2.2.5 Klasifikasi	15
2.2.6 <i>Imbalance data</i>	15
2.2.7 <i>Feature Engineering</i>	15
2.2.8 <i>Binning</i>	16
2.2.9 <i>Label Encoder</i>	16

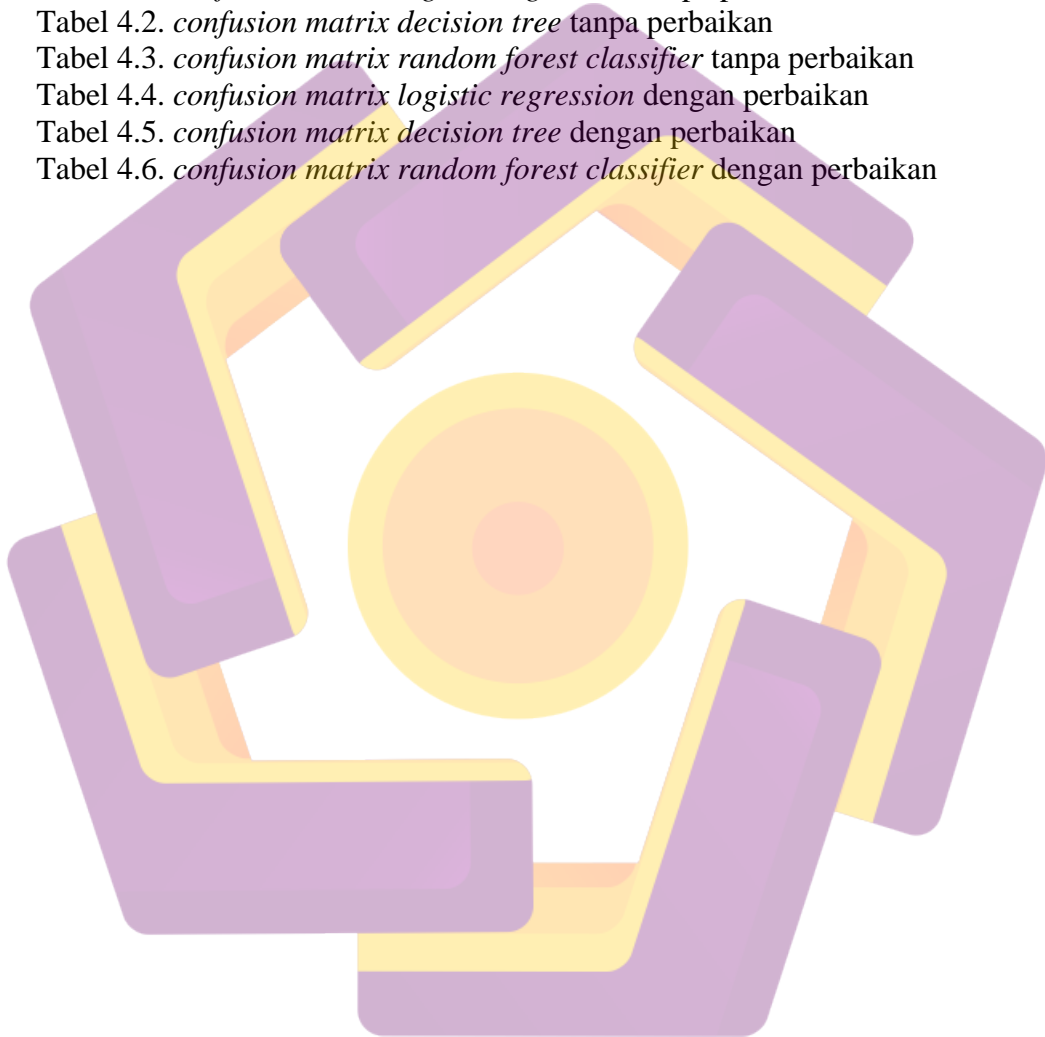
2.2.10	SMOTE	16
2.2.11	Korelasi <i>spearman</i>	17
2.2.12	<i>Training data</i> dan <i>test data</i>	18
2.2.13	<i>Logistic Regression</i>	19
2.2.14	<i>Decision tree C4.5</i>	21
2.2.15	<i>Random Forest Classifier</i>	24
2.2.16	<i>Confusion matrix</i>	25
2.2.17	<i>Accuracy</i>	26
2.2.18	<i>Overfitting</i> dan <i>Underfitting</i>	27
BAB III METODE PENELITIAN		28
3.1	Alur Penelitian	28
3.2	Alat dan Bahan	29
BAB IV HASIL DAN PEMBAHASAN		31
4.1	Akuisisi Data	31
4.2	<i>Data preprocessing</i>	34
4.3	<i>Feature Engineering</i>	42
4.3.1	<i>Payment Method</i>	42
4.3.2	<i>Total transactions</i>	44
4.3.3	<i>Shipment date limit</i>	45
4.3.4	<i>Shipment fee</i>	47
4.3.5	<i>Traffic source</i>	49
4.3.6	<i>Click item detail</i>	51
4.3.7	<i>Total amount</i>	52
4.3.8	<i>Promo amount</i>	54
4.3.9	<i>Promo used count</i>	55
4.3.10	<i>Search frequent</i>	57
4.3.11	Menggabungkan <i>feature-feature</i> menjadi satu <i>dataset</i>	58
4.3.12	SMOTE	59
4.3.13	<i>Feature selection - Korelasi Spearman</i>	60
4.4	<i>Modelling with machine learning model</i>	61
4.4.1	<i>Logistic Regression</i>	62
4.4.2	<i>Decision tree</i>	63
4.4.3	<i>Random Forest Classifier</i>	63
4.5	<i>Evaluation</i>	64
4.5.1	<i>Confusion matrix</i>	64
4.5.1.1	Tanpa SMOTE + <i>Feature Selection</i>	64

4.5.1.2	Dengan SMOTE + <i>Feature Selection</i>	65
4.5.2	<i>Accuracy</i>	65
4.5.2.1	Tanpa SMOTE + <i>Feature Selection</i>	65
4.5.2.2	Dengan SMOTE + <i>Feature Selection</i>	66
4.5.3	Penjelasan hasil <i>accuracy</i>	67
BAB V PENUTUP		69
5.1	Kesimpulan	69
5.2	Saran	70
REFERENSI		71



DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	8
Tabel 2.2. Korelasi <i>spearman</i> tidak terikat	17
Tabel 2.3. Korelasi <i>spearman</i> terikat	17
Tabel 2.4. <i>Confusion matrix</i>	25
Tabel 4.1. <i>confusion matrix logistic regression</i> tanpa perbaikan	64
Tabel 4.2. <i>confusion matrix decision tree</i> tanpa perbaikan	64
Tabel 4.3. <i>confusion matrix random forest classifier</i> tanpa perbaikan	64
Tabel 4.4. <i>confusion matrix logistic regression</i> dengan perbaikan	65
Tabel 4.5. <i>confusion matrix decision tree</i> dengan perbaikan	65
Tabel 4.6. <i>confusion matrix random forest classifier</i> dengan perbaikan	65



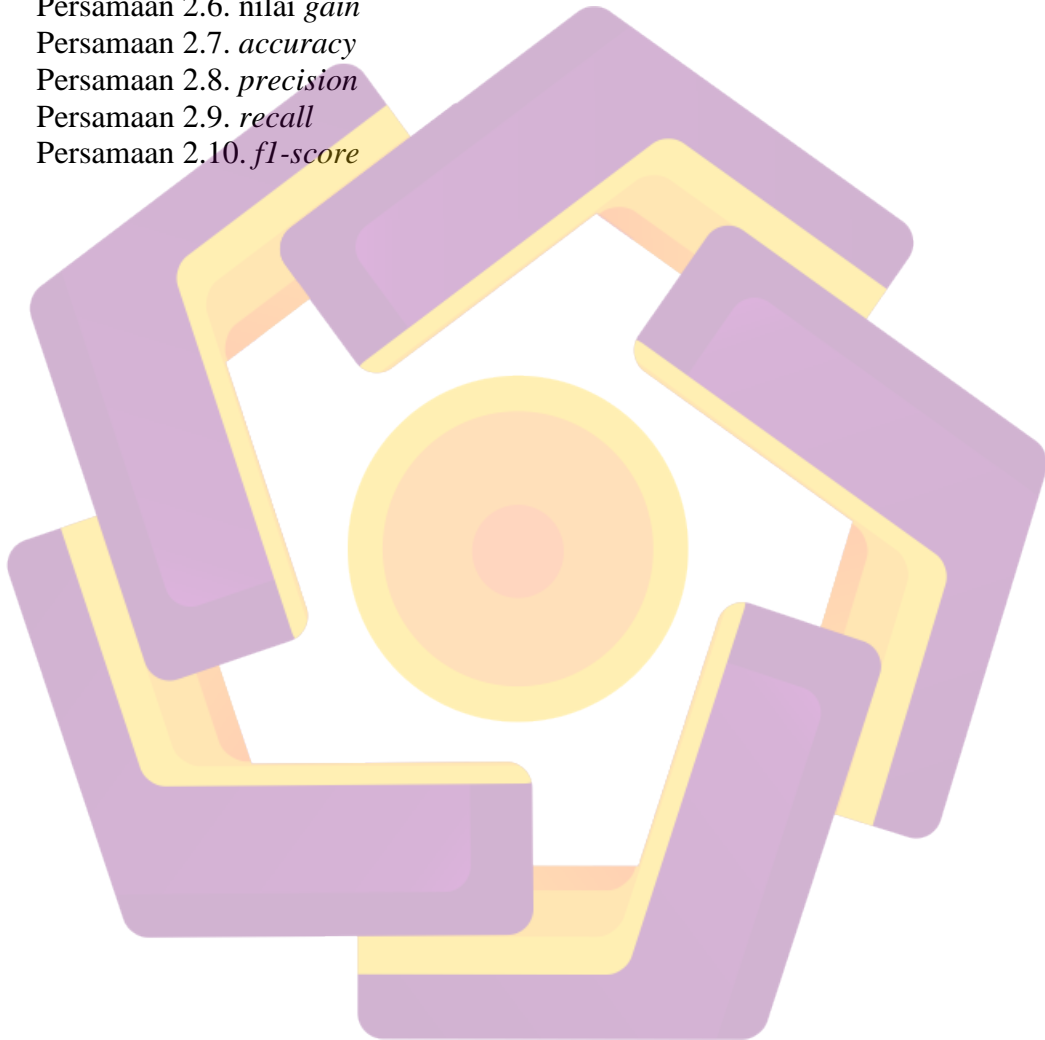
DAFTAR GAMBAR

Gambar 2.1. CRISP-DM	12
Gambar 2.2. SMOTE	17
Gambar 2.3. Diagram kartesius dengan <i>linear regression</i>	19
Gambar 2.4. <i>Logistic regression</i>	20
Gambar 2.5. <i>Decision tree</i>	21
Gambar 2.6. pembuatan <i>node</i> akar	23
Gambar 2.7. hasil akhir <i>decision tree</i> yang diharapkan	24
Gambar 2.8. <i>Random Forest classifier</i>	24
Gambar 3.1. Alur Penelitian	28
Gambar 4.1 Tabel <i>customer</i>	31
Gambar 4.2 Tabel <i>transactions</i>	32
Gambar 4.3 Tabel <i>product</i>	33
Gambar 4.4 Tabel <i>click_stream</i>	34
Gambar 4.5 <i>feature metadata</i> dari tabel <i>transactions</i>	35
Gambar 4.6 mengubah <i>str</i> menjadi <i>list</i>	35
Gambar 4.7 menggunakan fungsi <i>explode</i>	35
Gambar 4.8 <i>feature</i> baru dari <i>feature metadata</i> tabel <i>transactions</i>	36
Gambar 4.9 <i>feature metadata</i> dari tabel <i>click_stream</i>	36
Gambar 4.10 mengisi <i>null</i> dan mengubah <i>str</i> menjadi <i>dictionary</i>	37
Gambar 4.11 <i>feature</i> baru dari <i>feature metadata</i> tabel <i>click_stream</i>	37
Gambar 4.12 pelanggan yang pernah bertransaksi sukses minimal sekali	38
Gambar 4.13 pelanggan yang pernah bertransaksi sukses atau tidak	38
Gambar 4.14 kode mencari jeda waktu transaksi	39
Gambar 4.15 membuat label pelanggan	39
Gambar 4.16 hasil membuat label pelanggan	40
Gambar 4.17 kode untuk membuang label 2	40
Gambar 4.18 hasil membuang label 2	41
Gambar 4.19 rasio perbandingan label	42
Gambar 4.20 metode pembayaran pelanggan	42
Gambar 4.22 mengklasifikasi metode pembayaran	43
Gambar 4.22 melakukan fungsi <i>get_dummies</i> pada <i>feature payment method</i>	43
Gambar 4.23 <i>groupby feature payment method</i>	44
Gambar 4.24 <i>feature total transactions</i>	44
Gambar 4.25 kategori <i>feature total transactions</i>	45
Gambar 4.26 hasil <i>binning feature total transactions</i>	45
Gambar 4.27 <i>feature shipment date limit</i>	46
Gambar 4.28 <i>binning feature shipment date limit</i>	46
Gambar 4.29 hasil <i>binning feature shipment date limit</i>	47
Gambar 4.30 <i>feature shipment fee</i> setelah proses <i>groupby</i>	47
Gambar 4.31 <i>binning feature shipment fee</i>	48
Gambar 4.32 hasil <i>binning feature shipment fee</i>	48
Gambar 4.33 proses pembuatan <i>feature traffic source</i>	49
Gambar 4.34 klasifikasi <i>traffic source</i>	50

Gambar 4.35 hasil akhir <i>feature traffic source</i>	50
Gambar 4.36 proses pembuatan <i>feature click item detail</i>	51
Gambar 4.37 klasifikasi <i>feature click item detail</i>	52
Gambar 4.38 hasil akhir <i>feature click item detail</i>	52
Gambar 4.39 proses pembuatan <i>feature total amount</i>	53
Gambar 4.40 proses <i>binning feature total amount</i>	53
Gambar 4.41 hasil akhir <i>feature total amount</i>	54
Gambar 4.42 proses pembuatan <i>feature promo amount</i>	54
Gambar 4.43 klasifikasi <i>feature promo amount</i>	55
Gambar 4.44 hasil akhir <i>feature promo amount</i>	55
Gambar 4.45 proses pembuatan <i>feature promo used count</i>	56
Gambar 4.46 proses klasifikasi <i>feature promo used count</i>	56
Gambar 4.47 hasil akhir <i>feature promo used count</i>	57
Gambar 4.48 proses pembuatan <i>feature search frequent</i>	57
Gambar 4.49 proses <i>binning feature search frequent</i>	58
Gambar 4.50 hasil akhir <i>feature search frequent</i>	58
Gambar 4.51 kode <i>merge</i>	59
Gambar 4.52 <i>dataset</i> akhir	59
Gambar 4.53 SMOTE	59
Gambar 4.54 rasio target variable sebelum SMOTE	60
Gambar 4.55 jumlah <i>target variable</i> setelah SMOTE	60
Gambar 4.56 Korelasi <i>spearman</i>	60
Gambar 4.57 membagi <i>dataset</i>	62
Gambar 4.58 rasio <i>feature label</i>	62
Gambar 4.59 <i>Logistic Regression</i>	63
Gambar 4.60 <i>Decision tree</i>	63
Gambar 4.61 <i>Random Forest Classifier</i>	63
Gambar 4.62 <i>accuracy logistic regression</i> tanpa perbaikan	66
Gambar 4.63 <i>accuracy decision tree</i> tanpa perbaikan	66
Gambar 4.64 <i>accuracy random forest classifier</i> tanpa perbaikan	66
Gambar 4.65 <i>accuracy logistic regression</i> dengan perbaikan	67
Gambar 4.66 <i>accuracy decision tree</i> dengan perbaikan	67
Gambar 4.67 <i>accuracy random forest classifier</i> dengan perbaikan	67

DAFTAR PERSAMAAN

Persamaan 2.1. korelasi <i>Spearman</i> tidak terikat	18
Persamaan 2.2. korelasi <i>Spearman</i> terikat	18
Persamaan 2.3. <i>linear regression</i>	19
Persamaan 2.4. <i>logistic regression</i>	20
Persamaan 2.5. <i>entropy</i>	22
Persamaan 2.6. nilai <i>gain</i>	22
Persamaan 2.7. <i>accuracy</i>	26
Persamaan 2.8. <i>precision</i>	26
Persamaan 2.9. <i>recall</i>	26
Persamaan 2.10. <i>f1-score</i>	27




DAFTAR LAMBANG DAN SINGKATAN

$[]$	kurung siku (<i>list</i> pada <i>Python</i>)
$\{\}$	kurung kurawal (<i>set</i> pada <i>Python</i>)
e	nilai Euler / basis logaritma natural ($e \approx 2,718$)
Σ	operator penjumlahan
C4.5	<i>Decision tree</i>
SVC	<i>Support Vector Machines Classifier</i>
TP	<i>True Positive</i>
TF	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>



DAFTAR ISTILAH

<i>Assessment</i>	penilaian terhadap sesuatu hal, dalam konteks laporan ini adalah penilaian terhadap tingkat akurasi dari hasil prediksi model
<i>Bar plot</i>	visualisasi data menggunakan diagram batang
<i>Baseline model</i>	pengolahan <i>dataset</i> dengan model machine learning yang sederhana.
<i>Boxplot</i>	visualisasi data menggunakan suatu kotak yang terdiri dari nilai maksimum, nilai minimum dan nilai kuartil
<i>Churned</i>	kata sifat yang mendefinisikan pelanggan tidak kembali bertransaksi di perusahaan yang bersangkutan
<i>Confusion matrix</i>	penilaian hasil prediksi model dengan mencocokkan hasil prediksi dengan hasil sebenarnya
<i>Dataset</i>	data-data dalam bentuk tabel yang berisi informasi tertentu
Data latih	data yang digunakan untuk melatih model (<i>training data</i>)
Data uji	data yang digunakan model setelah dilatih (<i>test data</i>). Hal ini digunakan untuk menguji model apakah dapat digunakan untuk data yang akan datang
<i>Dictionary</i>	format pada data di Python yang berisi <i>keys</i> dan <i>value</i> serta ditutup dengan kurung kurawal ({})
<i>Feature</i>	istilah lain dari kolom pada tabel
<i>Feature Engineering</i>	penambahan feature baru/mengubah feature yang sudah ada menjadi feature yang lebih baik untuk analisis
Label	<i>feature</i> yang menjadi target dalam analisis
<i>List</i>	format data pada Python yang ditutup dengan kurung siku ([])
<i>Machine learning</i>	pembelajaran mesin menggunakan algoritma tertentu untuk memprediksi suatu hal yang diinginkan oleh analis
<i>Model</i>	kode komputer yang berisi fungsi untuk mengolah data dengan algoritma tertentu



<i>Overfitting</i>	hasil prediksi pada model menggunakan training data lebih baik daripada menggunakan test data dengan selisih yang cukup besar
<i>Parameter</i>	ketentuan pada fungsi-fungsi tertentu pada bahasa pemrograman
<i>Set</i>	format data pada Python yang ditutup dengan kurung kurawal ({})
<i>Threshold</i>	suatu nilai pada <i>logistic regression</i> untuk menentukan klasifikasi pada <i>output</i> tertentu
Variabel dependen	variabel yang tidak dapat berdiri sendiri sehingga dipengaruhi oleh variabel lain
Variabel independen	variabel yang berdiri sendiri dan tidak terpengaruh dengan variabel lain
<i>Windowing</i>	pembuatan <i>dataset</i> yang berisi informasi dan perilaku pelanggan pada kurun waktu tertentu yang menjadi indikasi apakah pelanggan yang bersangkutan kembali lagi bertransaksi di perusahaan atau tidak

INTISARI

Customer churn analysis adalah metode yang digunakan oleh perusahaan untuk mengidentifikasi dan memprediksi pelanggan yang berpotensi untuk tidak kembali lagi bertransaksi di perusahaan. Hal tersebut merupakan masalah yang sangat penting karena pelanggan yang tidak loyal dapat mengurangi potensi pendapatan perusahaan dan meningkatkan biaya untuk mendapat pelanggan baru.

Untuk melakukan *churn customer analysis*, perusahaan biasanya mengumpulkan data tentang perilaku pelanggan dalam beberapa bulan ke belakang untuk mencari informasi tentang penyebab pelanggan kembali atau tidak kembali bertransaksi di perusahaan. Data-data tersebut kemudian dianalisis menggunakan teknik algoritma seperti analisis regresi atau analisis klasifikasi untuk mencari tahu faktor-faktor yang memprediksi kemungkinan *churn customer*. *Customer churn analysis* dapat membantu perusahaan untuk mengetahui alasan pelanggan meninggalkan perusahaan dan mengambil tindakan preventif yang sesuai. Misalnya, perusahaan dapat meningkatkan layanan pelanggan atau menawarkan promosi khusus kepada pelanggan yang berpotensi akan meninggalkan perusahaan.

Dengan demikian, perusahaan dapat menjaga pelanggan loyal dan meningkatkan pendapatan secara keseluruhan. Selain itu, perusahaan juga dapat menggunakan hasil *customer churn analysis* untuk mengembangkan strategi pemasaran yang lebih baik dan menargetkan pelanggan potensial yang lebih efektif.

Kata kunci: *churn*, *customer*, perusahaan, laba, rugi.

ABSTRACT

Customer churn analysis is a method used by companies to identify and predict customers who have the potential to not return to transact at the company. This is a very important issue because disloyal customers can reduce a company's revenue potential and increase the cost of acquiring new customers.

To carry out customer churn analysis, companies usually collect data on customer behavior in the past few months to find information about why customers return or do not return to transact with the company. These data are then analyzed using algorithmic techniques such as regression analysis or classification analysis to find out the factors that predict the possibility of customer churn. Customer churn analysis can help companies to find out why customers leave the company and take appropriate preventive action. For example, a company could improve customer service or offer special promotions to customers who are likely to leave the company.

Thus, companies can maintain loyal customers and increase overall revenue. In addition, companies can also use the results of customer churn analysis to develop better marketing strategies and target potential customers more effectively.

Keyword: churn, customer, company, profit, loss