

**ANALISIS PERALIHAN PELANGGAN MENGGUNAKAN
METODE *LIGHTGBM* DAN *COX PROPORTIONAL HAZARD*
DI PT KASIR PINTAR INTERNASIONAL**

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat kelulusan
di Program Studi Sains Data**



Disusun Oleh:

IBNU ZAHY' ATHA ILLAH

20083010016

**PROGRAM STUDI SAINS DATA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
SURABAYA
2024**

LEMBAR PENGESAHAN

ANALISIS PERALIHAN PELANGGAN MENGGUNAKAN METODE LIGHTGBM DAN COX PROPORTIONAL HAZARD DI PT KASIR PINTAR INTERNASIONAL

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data
pada : Selasa, 14 Mei 2024

Program Studi S-1 Sains Data
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya

Oleh:
IBNU ZAHY' ATHA ILLAH
NPM. 20083010016

Disetujui oleh Tim Penguji Skripsi:

Penguji 1

Trimano, S.Si., M.Si
NIP. 199509082022031003

Penguji 2

Eristva Maya Safitri, S.Kom., M.Kom
NIP. 199303162019032020

Pembimbing 1

Wahyu Syaifulah
Jauharis Saputra., S.Kom., M.Kom
NIP. 198608252021211003

Pembimbing 2

Aviolla Terza Damaliana, S.Si., M.Stat
NIP. 199408022022032015

Mengetahui

Fakultas Ilmu Komputer
Dekan,

Prof. Dr. Ir. Novirina Hendrasarie, MT,
NIP. 196811261994032001

Program Studi Sains Data
Fakultas Ilmu Komputer
Koordinator,

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU
NIP. 198012052005011002

Surabaya, Mei, 2024

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Ibnu Zahy' Atha Illah
NPM : 20083010016
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

“Analisis Peralihan Pelanggan Menggunakan Metode *LightGBM* dan *Cox Proportional Hazard* di PT Kasir Pintar Internasional”

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain.

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 3 Mei 2024

Hormat Saya,



Ibnu Zahy' Atha Illah

NPM. 20083010016

ABSTRAK

ANALISIS PERALIHAN PELANGGAN MENGGUNAKAN METODE *LIGHTGBM* DAN *COX PROPORTIONAL HAZARD* DI PT KASIR PINTAR INTERNASIONAL

Nama Mahasiswa / NPM : Ibnu Zahy' Atha Illah / 20083010016
Program Studi : Sains Data, FASILKOM, UPN Veteran Jatim
Dosen Pembimbing 1 : Wahyu Syaifullah Jauharis Saputra., S.Kom., M.Kom
Dosen Pembimbing 2 : Aviolla Terza Damaliana, S.Si., M.Stat

Abstrak

Persaingan yang semakin ketat dalam dunia usaha mendorong setiap sektor untuk memanfaatkan teknologi yang relevan guna menjaga pangsa pasarnya. Keberhasilan suatu perusahaan seringkali diukur dari seberapa kuat jaringan pelanggan yang mereka miliki. Peralihan pelanggan (*customer churn*) dapat menyebabkan penurunan pendapatan yang signifikan dan bahkan bisa mengancam eksistensi perusahaan itu sendiri. Penelitian ini bertujuan untuk menganalisis peralihan pelanggan sehingga dapat mencegah terjadinya fenomena tersebut. Metode yang digunakan dalam penelitian yaitu *LightGBM* untuk mengklasifikasikan pelanggan dan Regresi *Cox Proportional Hazard* untuk memproyeksikan retensi pelanggan. Dengan melakukan analisis menggunakan kedua metode ini, penulis dapat memberikan rekomendasi strategi kepada perusahaan untuk mencegah terjadinya peralihan pelanggan. Penelitian ini menghasilkan model klasifikasi *LightGBM* yang mencapai nilai *Accuracy*, *Precision*, *Recall*, *F1-score*, dan *AUC* masing-masing sebesar 0.964, 0.971, 0.990, 0.980, dan 0.965. Selain itu, model Regresi *Cox Proportional Hazard* memiliki nilai evaluasi *C-index* sebesar 0.83. Model Regresi *Cox Proportional Hazard* juga menunjukkan bahwa pelanggan yang saat ini *non-churn* memiliki rata-rata ekspektasi retensi selama 15 bulan. Rekomendasi strategi yang bisa diterapkan perusahaan untuk mencegah terjadinya peralihan pelanggan dapat didasarkan pada lima fitur paling penting, segmentasi RFM, dan ekspektasi retensi.

Kata kunci: Analisis Peralihan Pelanggan, *LightGBM*, Regresi *Cox Proportional Hazard*.

ABSTRACT

ANALYSIS OF CUSTOMER TRANSITION USING THE LIGHTGBM AND COX PROPORTIONAL HAZARD METHOD AT PT KASIR PINTAR INTERNASIONAL

Student Name / NPM : Ibnu Zahy' Atha Illah / 20083010016
Study Program : Sains Data, FASILKOM, UPN Veteran Jatim
Advisor 1 : Wahyu Syaifullah Jauharis Saputra., S.Kom., M.Kom
Advisor 2 : Aviolla Terza Damaliana, S.Si., M.Stat

Abstract

Increasingly fierce competition in the business world encourages every sector to utilize relevant technology to maintain its market share. The success of a company is often measured by the strength of its customer network. Customer churn can cause a significant decrease in revenue and can even threaten the company's existence. This research aims to analyze customer churn to prevent this phenomenon from occurring. The methods used in the research are LightGBM for classifying customers and the Cox Proportional Hazard model for projecting customer retention. By conducting analysis using these two methods, the author can provide strategic recommendations to companies to prevent customer churn. This research produced a LightGBM classification model that achieved Accuracy, Precision, Recall, F1-score, and AUC values of 0.964, 0.971, 0.990, 0.980, and 0.965, respectively. Additionally, the Cox Proportional Hazard model has a C-index evaluation value of 0.83. The Cox Proportional Hazard model also indicates that current non-churn customers have an average retention expectation of 15 months. Recommended strategies that companies can implement to prevent customer churn are based on the five most important features, RFM segmentation, and retention expectations.

Keywords: *Customer Churn Analysis, LightGBM, Cox Proportional Hazard Regression*

KATA PENGANTAR

Puji syukur penulis panjatkan atas kehadiran Allah SWT berkat segala limpahan rahmat, hidayah, iman, dan juga Islamnya-Nya sehingga penulis dapat menyelesaikan buku skripsi tanpa ada kendala yang berarti. Buku ini disusun sebagai perjalanan akhir dalam pembelajaran sebagai mahasiswa S1 selama 8 semester. Buku ini juga disusun sebagai salah satu syarat yang harus dipenuhi untuk menyelesaikan perjalanan tersebut dan keperluan konversi mata kuliah skripsi di Program Studi Sains Data. Dalam kesempatan ini, penulis ingin mengucapkan terima kasih kepada pihak yang telah membantu pelaksanaan dan penyusunan buku skripsi ini, khususnya kepada:

1. Kedua orang tua tercinta yang senantiasa mendoakan penulis dan selalu mendukung serta memberikan semangat.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Prof. Dr. Ir. Novirina Hendrasarie, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Wahyu Syaifullah JS., S.Kom., M.Kom dan Aviolla Terza Damaliana, S.Si., M.Stat selaku Dosen Pembimbing 1 dan 2 serta Dosen Sains Data lainnya.
6. PT Kasir Pintar Internasional selaku penyedia data dalam penelitian.

Penulis menyadari bahwa buku skripsi ini masih memiliki kekurangan, oleh karena itu penulis sangat mengharapkan masukan dan saran dalam penulisan agar dapat dilakukan pengembangan untuk penyempurnaan dari apa yang penulis buat. Akhir kata, semoga hasil buku skripsi ini dapat memberikan manfaat dan menambah wawasan pengetahuan bagi berbagai pihak yang membacanya.

Surabaya, 3 Mei 2024

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA.....	7
2.1. Dasar Teori.....	7
2.2.1. Fenomena Peralihan Pelanggan	7
2.2.2. PT Kasir Pintar Internasional	10
2.2.3. Klasifikasi	11
2.2.4. Prediksi.....	12
2.2.5. Operasi <i>Join</i>	13
2.2.6. Korelasi	15
2.2.7. Seleksi Fitur	16
2.2.8. <i>Resampling</i>	17
2.2.9. <i>RFM</i>	20
2.2.10. Metode Klasifikasi <i>LightGBM</i>	21
2.2.11. Analisis Survival Metode Regresi <i>Cox Proportional Hazard</i>	27
2.2.12. Evaluasi Model.....	35
2.2. Penelitian Terdahulu	38
BAB III METODOLOGI PENELITIAN.....	43

3.1. Variabel Penelitian dan Sumber Data	43
3.2. Langkah Analisis	48
3.3. Diagram Alir Penelitian	52
3.4. Jadwal Penelitian	53
BAB IV HASIL DAN PEMBAHASAN	54
4.1 Analisis Peralihan Pelanggan dengan <i>LightGBM</i>	54
4.1.1. Pemahaman Data.....	55
4.1.2. Pembersihan Data.....	57
4.1.3. Integrasi Data	59
4.1.4. Imputasi Data	65
4.1.5. Rekayasa Fitur.....	66
4.1.6. Seleksi Fitur	75
4.1.7. Penanganan Proporsi Data Tidak Seimbang	77
4.1.8. Segmentasi Data.....	83
4.1.9. Pemodelan <i>LightGBM</i>	87
4.1.10. Evaluasi Model <i>LightGBM</i>	95
4.2 Analisis Peralihan Pelanggan dengan Regresi <i>Cox Proportional Hazards</i> ..	100
4.2.1. Asumsi <i>Proportional Hazard</i>	100
4.2.2. Model Regresi <i>Cox Proportional Hazard</i>	102
4.2.3. Uji Hipotesis Signifikansi Parameter secara Parsial	102
4.2.4. Uji Hipotesis Signifikansi Parameter secara Serentak	104
4.2.5. Evaluasi Model Regresi <i>Cox Proportional Hazards</i>	104
4.2.6. Proyeksi Model Regresi <i>Cox Proportional Hazards</i>	105
4.3 Analisis Preskriptif untuk Rekomendasi Perusahaan.....	107
BAB V PENUTUP.....	111
5.1 Kesimpulan	111
5.2 Saran	111
DAFTAR PUSTAKA	113
LAMPIRAN.....	120
BIODATA PENULIS	125

DAFTAR GAMBAR

Gambar 2.1 <i>Inner Join</i>	14
Gambar 2.2 <i>Left Join</i>	14
Gambar 2.3 <i>Right Join</i>	15
Gambar 2.4 <i>Full Outer Join</i>	15
Gambar 2.5 <i>Leaf wise growth</i>	23
Gambar 2.6 Ilustrasi Penghitungan <i>C-Index</i>	37
Gambar 2.7 <i>Ranking C-index</i>	38
Gambar 3.1 Flowchart Prediksi <i>Churn</i>	52
Gambar 4.1 Hasil Penanganan <i>Datetime</i>	59
Gambar 4.2 Hasil Integrasi Tabel Promo dengan Tabel Promo User.....	60
Gambar 4.3 Hasil Persiapan Tabel Laporan	61
Gambar 4.4 Hasil merge_1	62
Gambar 4.5 Hasil merge_1 dengan tabel NPS.....	63
Gambar 4.6 Hasil merge_3	63
Gambar 4.7 Hasil merge_4	64
Gambar 4.8 Hasil Kelengkapan Data dataset_final	65
Gambar 4.9 Statistika Deskriptif pada dataset_final.....	66
Gambar 4.10 Hasil Agregasi Data Transaksi	67
Gambar 4.11 Hasil Proses Agregasi Data	69
Gambar 4.12 Hasil Rekayasa Fitur Lamanya Berlangganan	72
Gambar 4.13 Hasil Rekayasa Fitur	73
Gambar 4.14 Hasil <i>Merge</i>	74
Gambar 4.15 Hasil Akhir Rekayasa Fitur	74
Gambar 4.16 Komposisi Label	75
Gambar 4.17 Hasil RFECV	76
Gambar 4.18 Hasil Proporsi	77
Gambar 4.19 Hasil Proporsi Data	78
Gambar 4.20 Hasil <i>Manual Undersampling</i>	80
Gambar 4.21 Komposisi Awal Data	81
Gambar 4.22 Hasil <i>Undersampling</i>	81

Gambar 4.23 Hasil <i>Oversampling</i>	82
Gambar 4.24 Hasil <i>Hybrid-sampling</i>	83
Gambar 4.25 Hasil <i>Recency</i>	83
Gambar 4.26 Hasil <i>Frequency</i>	84
Gambar 4.27 Hasil <i>Monetary Value</i>	84
Gambar 4.28 Hasil Konversi <i>Recency</i>	85
Gambar 4.29 Statistika Deskriptif RFM	86
Gambar 4.30 Hasil Segmentasi 1	86
Gambar 4.31 Hasil Segmentasi 2	87
Gambar 4.32 Hasil Segmentasi 3	87
Gambar 4.33 Evaluasi AUC.....	97
Gambar 4.34 Evaluasi <i>Binary Logloss</i>	97
Gambar 4.35 Grafik ROC	98
Gambar 4.36 <i>Confusion Matrix</i> Validasi	99
Gambar 4.37 <i>Confussion Matrix</i> Uji.....	99
Gambar 4.38 Fitur Penting	100
Gambar 4.40 Hasil Estimasi Ulang Regresi <i>Cox Proportional Hazard</i>	103
Gambar 4.41 Hasil <i>C-index</i>	105
Gambar 4.42 Hasil Proyeksi Pelanggan <i>non-churn</i>	106
Gambar 4.43 Hasil Statistika Deskriptif	106

DAFTAR TABEL

Tabel 2.1. Tabel Studi Literatur yang Relevan dalam Penelitian	39
Tabel 3.1. Tabel Billing	43
Tabel 3.2. Tabel Laporan	44
Tabel 3.3. Tabel NPS	44
Tabel 3.4. Tabel Promo	45
Tabel 3.5. Tabel Promo User	46
Tabel 3.6. Tabel Toko	46
Tabel 3.7. Tabel Penjelasan Data Loyalitas Pelanggan	47
Tabel 3.8. Tabel Penjelasan Atribut Data Loyalitas Pelanggan.....	47
Tabel 3.9. Rekayasa Fitur.....	49
Tabel 3.10 Jadwal Kegiatan Penelitian <i>Churn</i>	53
Tabel 4.1 Status Keakuratan Data.....	55
Tabel 4.2 Hasil Eksplorasi Data.....	57
Tabel 4.3 Inisialisasi Model	88
Tabel 4.4 Nilai <i>Split Gain</i>	93
Tabel 4.5 Nilai Gradien.....	94
Tabel 4.6 Daftar Nilai Akurasi dan Binary Logloss	95
Tabel 4.7 Hasil Klasifikasi <i>LightGBM</i>	95
Tabel 4.8 Hasil Uji Asumsi <i>Cox Proportional Hazard</i>	101

DAFTAR LAMPIRAN

Lampiran 1. Hasil Uji Plagiasi	120
Lampiran 2. Data Penelitian.....	121
Lampiran 3. <i>Source Code</i> yang Digunakan untuk Analisis	123
Lampiran 4. Observasi Masalah dalam Penelitian.....	124