
Pendekatan Data Science Untuk Menemukan Customer Churn Pada Perusahaan Fashion Dengan Metode Machine Learning

Tasyifa Nafsiah Muthmainnah¹, Apriade Voutama²

^{1,2}Sistem Informasi, Universitas Singaperbangsa Karawang

Email: ¹2010631250023@student.unsika.ac.id, ²apriade.voutama@staff.unsika.ac.id

Email Penulis Korespondensi: 2010631250023@student.unsika.ac.id

Article History:

Received May 23th, 2023

Revised Jun 21th, 2023

Accepted Jul 6th, 2023

Abstrak

Banyaknya perusahaan yang bergerak pada industri fashion, dalam hal ini membuat perusahaan fashion terus menghadirkan inovasi untuk mendorong persaingan usaha yang ketat di industri fashion tersebut. Hal ini yang menyebabkan customer churn menjadi masalah terbesar yang dihadapi oleh perusahaan, dimana pelanggan memutuskan untuk berhenti bertransaksi. Dengan banyaknya e-commerce membuat pelanggan memilih untuk berbelanja pada e-commerce dibandingkan berbelanja secara langsung. Penelitian ini bertujuan untuk menganalisa kejadian churn dengan membuat *churn prediction model* menggunakan pendekatan *data science* dengan metode *machine learning*. Setelah itu, diperlukan sebuah rekomendasi terhadap customer tersebut. Dengan adanya sebuah rekomendasi, diharapkan dapat memberikan sebuah informasi yang bermanfaat bagi perusahaan dalam pengambilan suatu keputusan. Dari berbagai uji coba pada empat model, dipilih model *RandomForestClassifier* dan menghasilkan nilai akurasi sebesar 1.00 .

Kata Kunci : *Churn, Data Science, Machine Learning, Perusahaan Fashion*

Abstract

The number of companies engaged in the fashion industry, in this case, makes fashion companies continue to present innovations to encourage intense business competition in the fashion industry. This causes customer churn to be the biggest problem faced by companies, where customers decide to stop transacting. With so many e-commerce, customers choose to shop on e-commerce rather than shopping directly. This research aims to analyze the incidence of churn by creating a churn prediction model using a data science approach with machine learning methods. After that, a recommendation is needed for the customer. With a recommendation, it is expected to provide useful information for companies in making a decision. From various trials on four models, the RandomForestClassifier model was chosen and produced an accuracy value of 1.00.

Keyword : *Churn, Data Science, Machine Learning, Fashion Company*

1. PENDAHULUAN

Setiap perusahaan mempunyai caranya sendiri untuk menawarkan layanan yang berkualitas dengan harga serendah mungkin. Hal ini diharapkan dapat menarik pelanggan sebanyak mungkin sehingga pendapatan perusahaan semakin optimal. Hal ini menimbulkan fenomena churn terjadi, dimana pelanggan memutuskan untuk beralih dari satu tempat ke tempat yang lain karena tertarik dengan harga yang lebih menarik. Fenomena churn ini tentu saja mengkhawatirkan, karena jika tidak dicegah dan ditangani akan mengakibatkan pada penurunan pendapatan perusahaan [1].

Saat ini banyak pebisnis dan masyarakat Indonesia yang melakukan transaksi bisnis secara online melalui *e-commerce*. Hal ini karena berbagai kelebihanannya, salah satunya adalah luas dan waktu yang tidak terbatas [2]. Banyak perusahaan yang bergerak pada industri fashion. Dalam hal ini membuat perusahaan fashion terus menghadirkan berbagai inovasi untuk mendorong persaingan usaha yang sangat ketat di industri fashion tersebut [3]. Ada banyak alasan yang membuat customer churn menjadi masalah terbesar yang dihadapi oleh sebuah perusahaan. Salah satunya yaitu, adanya

e-commerce membuat pelanggan memilih untuk berbelanja pada *e-commerce* dibandingkan berbelanja secara langsung [4].

Data Mining adalah teknik untuk pencarian dalam big data dengan tujuan untuk mengetahui suatu pengetahuan berdasarkan pola yang ada [5]. Teknik data mining, atau klasifikasi, diperlukan untuk membuat prediksi dari sejumlah besar data atau data berbentuk [6]. Data mining merupakan bagian dari machine learning yang dimana bertugas untuk mencari pola atau informasi yang menarik dalam data pilihan dengan menggunakan teknik atau metode tertentu [7]. Dalam aplikasi machine learning, algoritma atau langkah-langkah statistik dipelajari agar dapat mengidentifikasi pola dan fitur tertentu dalam jumlah data yang besar. Dengan tujuan untuk membuat suatu keputusan dan prediksi berdasarkan data tersebut. Semakin baik algoritmanya, maka akurasi keputusan dan prediksi sistem akan semakin bagus [8]. Klasifikasi merupakan salah satu teknik yang dapat digunakan untuk memprediksi nilai dari sekumpulan atribut dalam mendeskripsikan dan memisahkan kelas data [9].

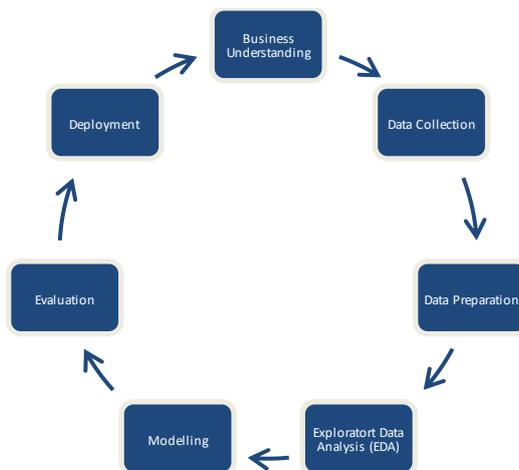
Beberapa penelitian telah dilakukan diantaranya yaitu penelitian yang dilakukan oleh Firmasnyah dan Agus Yulianto dengan berjudul “Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naïve Bayes” dimana dari pemodelan Naïve Bayes tersebut berhasil memperoleh hasil akurasi mencapai 80% dan presisi sebesar 100% [10]. Abdul Halim Hasugian, M. Fakhriza, Dinda Zukhoiriyah dalam penelitian yang berjudul “Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes” menghasilkan akurasi sebesar 99.5%, *precision* sebesar 99.49%, dan *recall* sebesar 100% [11]. Penelitian Amir Mahmud Husein, Mawaddah Harahap, Piter Fernando yang berjudul “Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning” dengan menggunakan lima model yang berbeda yaitu *Kneighbors Classifier*, *Logistic Regression*, *Linear SVC*, *Random Tree Classifier* dan *Random Forest Classifier* mendapatkan hasil dari pengujian 5 model tersebut model *Random Forest Classifier* dan *Kneighbors Classifier* yang lebih baik daripada model lainnya dengan akurasi dari kedua model tersebut sebesar 86% dan 84% [12].

Daris Hadyan Trisantri, Randy Cahya Wihandika, Sigit Adinugroho dalam penelitiannya yang berjudul “Prediksi Keputusan Pelanggan Menggunakan Extreme Learning Machine Pada Data Telco Customer Churn” dengan menggunakan *Extreme Learning Machine* untuk memprediksi customer churn mendapatkan *accuracy* sebesar 76.96%, *precision churn* 65.45%, *precision non churn* 78.65%, *recall churn* 29.38%, *recall non churn* 94.19% [13]. Ardhelia Nurfira Rahmi dalam penelitiannya berjudul “Implementasi Metode Random Forest Dan Xgboost Pada Klasifikasi Customer Churn” mendapatkan hasil dari metode *Xgboost* dengan nilai akurasi dan AUC sebesar 95.6% dan 0.876 dan dari metode *Random Forest* menghasilkan nilai akurasi dan AUC 93.5% dan 0.799 [3].

Tujuan dari penelitian ini adalah untuk menganalisa kejadian churn dengan membuat churn prediction model. Setelah itu, diperlukan sebuah rekomendasi terhadap customer tersebut. Dengan adanya sebuah rekomendasi, diharapkan dapat memberikan sebuah informasi yang bermanfaat bagi perusahaan dalam pengambilan suatu keputusan [10].

2. METODOLOGI PENELITIAN

Metodologi yang digunakan untuk membuat model prediksi terdiri dari *Business Understanding*, *Data Collection*, *Data Preparation*, *Exploratory Data Analysis*, *Modelling*, *Evaluation* dan *Deployment*. Gambar 1 dibawah ini merupakan metodologi yang diusulkan:



Gambar 1. Metodologi yang diusulkan

2.1 Business Understanding

Pada tahap ini, fokus pada mendefinisikan masalah bisnis dan memahami lingkungan penelitian. Masalah bisnis yang akan diselesaikan harus dapat didefinisikan dengan jelas dan terukur [14]. Pemahaman bisnis sangat penting bagi

perusahaan misalnya, berinvestasi dengan cara baru untuk meningkatkan penjualan produk mereka yang dapat mengakibatkan penjualan rendah di masa mendatang melalui promosi atau diskon, dan lain sebagainya. Oleh karena itu, sebagai perusahaan yang sudah mapan, penting untuk memperkirakan penjualan di masa depan secara akurat.

2.2 Data Collection

Setelah memahami proses bisnis, langkah berikutnya adalah pemahaman data. Pada langkah ini, dilakukan pengumpulan seluruh data yang tersedia. Data tersebut dapat diperoleh dari berbagai sumber seperti database, data warehouse, internet, dan sumber lainnya. Kemudian, data tersebut diolah untuk dijadikan informasi yang berguna dan dapat digunakan untuk berbagai keperluan [15].

2.3 Data Preparation

Setelah proses pemahaman data, tahap selanjutnya yaitu persiapan data. Pada tahap ini, dikenal juga dengan pembersihan data atau Data Wrangling. Proses ini meliputi langkah-langkah yaitu pemilihan data yang relevan, integrasi data dengan menggabungkan kumpulan data, membersihkan data, mengatasi nilai yang hilang dengan menghapusnya atau menggantikannya dengan data yang relevan, menghapus data yang tidak sesuai, serta mendeteksi dan mengatasi *outlier*. Proses penyesuaian data sangatlah penting agar dapat sesuai dengan kebutuhan penelitian [16].

2.4 Exploratory Data Analysis (EDA)

Pada tahap ini, *Exploratory Data Analysis* (EDA) memiliki peran yang sangat penting, karena dapat membantu dalam menemukan himpunan fitur yang tepat, memilih algoritma yang cocok untuk pembuatan model, serta membangun model itu sendiri. Dan agar dapat memahami karakteristik dasar dari data yang ada sebelum menjalankan model atau analisis lebih lanjut [12].

2.5 Modelling

Pada tahap ini, pemodelan data dianggap sebagai inti dari analisis data. Yang di mana model tersebut mengambil data yang telah dipersiapkan pada tahap persiapan data sebagai *input* dan memberikan hasil yang diinginkan atau sebagai *output*. Tahap ini melibatkan pemilihan jenis model yang sesuai tergantung pada jenis masalah yang dihadapi, apakah itu masalah *classification*, *regression*, atau *clustering*. Setelah memilih model, dilakukan pemilihan dari berbagai algoritma yang tersedia. Dalam penelitian kami, kami melakukan proses pemodelan *Grid Search Cross Validation* dengan menghasilkan model yang memiliki kinerja terbaik untuk digunakan sebagai model prediksi.

2.6 Evaluation

Setelah proses pemodelan data, tahap selanjutnya yaitu mengevaluasi model yang tujuannya untuk mengukur kinerja dan akurasi dari model yang telah ada sesuai dengan target yang ingin dicapai pada tahap pertama, dalam penelitian ini menggunakan metode *train/test split* dengan melakukan proses *training* dan *testing* agar dapat memberikan hasil prediksi yang lebih akurat.

2.7 Deployment

Setelah proses evaluasi, model dapat diimplementasikan dalam bentuk dan saluran yang kita inginkan. Tahapan ini merupakan tahap terakhir dalam *data science* yang diusulkan.

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Objek pada penelitian ini adalah Perusahaan Fashion Campus. Permasalahan bisnis pada Perusahaan Fashion Campus ini adalah dengan jumlah user yang semakin meningkat, banyak user yang *diacquired* hanya untuk memenuhi jumlah user target dari perusahaan Fashion Campus. Tim Marketing diberikan tugas untuk membuat promo yang menarik untuk menarik perhatian user. Namun, hal ini berdampak dengan banyaknya user yang tidak organik, dimana mereka banyak yang tidak kembali lagi ke platform untuk melakukan transaksi. Akibatnya, churn rate pengguna meningkat. Dari masalah ini, maka tujuan dari penelitian ini adalah untuk menganalisa kejadian dengan membuat *churn prediction model* dengan menggunakan metode *machine learning*.

3.2 Data Collection

Pada tahap ini, dilakukan *import data* kedalam *google colab* dari data Perusahaan Fashion Campus. Dataset ini terdiri dari empat data yaitu data *customer*, data *product*, data *transaction*, dan data *clickstream*.

3.3 Data Preparation

Pada tahap selanjutnya adalah melakukan proses persiapan data yang terdiri *data integration*, *cleaning data*, *feature engineering* dan *feature selection*. Langkah pertama adalah penggabungan data atau *data integration*. Proses ini

menggabungkan data *customer*, *product*, *transaction*, dan data *clickstream*. Selanjutnya setelah tahap penggabungan data tahap berikutnya yaitu pembersihan data. Pada proses *cleansing data* ini melakukan pembersihan dan modifikasi data sehingga siap digunakan untuk analisis lebih lanjut. Pada tahap ini, mengisi nilai null atau nilai yang kosong dengan melakukan *handling missing value* untuk type data numerik dengan mean atau median sedangkan untuk type data object dengan modus. Setelah proses *cleaning data* dilanjutkan ke tahap berikutnya yaitu data integration. Kemudian membuat 15 fitur baru pada data tersebut untuk meningkatkan performa model. Pada tahap *feature selection* atau pemilihan fitur, dari total 50 fitur hanya 14 fitur saja yang relevan untuk digunakan. Dataset yang telah dilakukan *feature selection* dapat dilihat pada tabel 1.

Tabel 1. Dataset yang digunakan

	Ship Fee	Age	Duration	Rec	Freq	Mon	Total Item	Total Promo Amount	Total Promo Code Used	Max Shipping	Gender	Device Type	Payment Method	Main Location
0	10000	22	1	1463	1	199832	1	1415	1	5	F	Android	Debit Card	Pulau Jawa
1	50000	27	1412	52	43	31135863	64	123186	43	6	F	Android	Credit Card	Luar Pulau Jawa
2	10000	9	1441	23	152	2.05E+08	195	205903	152	6	M	Android	OVO	Luar Pulau Jawa
3	10000	28	0	1464	1	195705	1	0	1	6	F	Android	OVO	Pulau Jawa
4	10000	25	977	487	3	545304	3	10448	3	3	F	Android	LinkAja	Luar Pulau Jawa
...
50700	0	30	313	313	2	514508	3	0	2	6	M	Android	Debit Card	Pulau Jawa
50701	10000	17	1	234	1	359203	1	5010	1	3	F	Android	OVO	Luar Pulau Jawa
50702	5000	32	3	232	1	1158159	4	2181	1	1	F	Android	Credit Card	Pulau Jawa
50703	0	15	0	956	1	184143	1	4503	1	5	F	iOS	Credit Card	Pulau Jawa
50704	10000	19	1	955	1	365963	1	4724	1	4	F	Android	Gopay	Pulau Jawa

Kumpulan dari dataset ini berisi fitur-fitur antara lain yaitu *Shipment Fee*, *Age*, *Duration*, *Recency*, *Frequency*, *Monetary*, *Total Item*, *Total Promo Amount*, *Total Promo Code Used*, *Max Shipping*, *Gender*, *Device Type*, *Payment Method*, *Main Location*. Deskripsi dari fitur-fitur pada tabel 1 secara terperinci dapat dilihat pada tabel 2 dibawah.

Tabel 2. Deskripsi kolom

Fitur	Pengertian
Shipping fee	Biaya Pengiriman Produk
Age	Usia customer berdasarkan tanggal lahir
Duration	Durasi lamanya menjadi pengguna
Recency	Total hari sejak pembelian terakhir oleh customer
Frequency	Customer yang melakukan pembelian kembali
Monetary	Jumlah harga pembelian tiap customer
Total Item	Jumlah total barang yang ada
Total Promo Amount	sejumlah promosi produk
Total Promo Code Used	Total kode promo yang digunakan oleh customer
Max Shipping	Pengiriman maksimal barang
Gender	Jenis kelamin (kategori: F dan M)
Device Type	Tipe perangkat yang digunakan customer
Payment Method	Pembayaran utama customer
Main Location	Lokasi tempat tinggal customer

3.4 Exploratory Data Analysis (EDA)

Pada tahap ini, dilakukan sebuah proses analisis data yang menghasilkan insight dari beberapa data. Hasil EDA dapat dilihat pada gambar 2.



Gambar 2. Insight dari data

Dari insight diatas dapat terlihat bahwa Perempuan lebih banyak melakukan transaksi dibandingkan Laki-Laki, Pengguna Android lebih banyak dibandingkan iOS, Domisili pelanggan peringkat satu diraih oleh Jakarta, disusul oleh

Jawa Barat, Jawa Tengah, Jawa Timur, dan Yogyakarta ternyata peminat tertinggi berada di pulau Jawa, Metode Pembayaran paling banyak menggunakan Kartu kredit.

3.5 Modelling

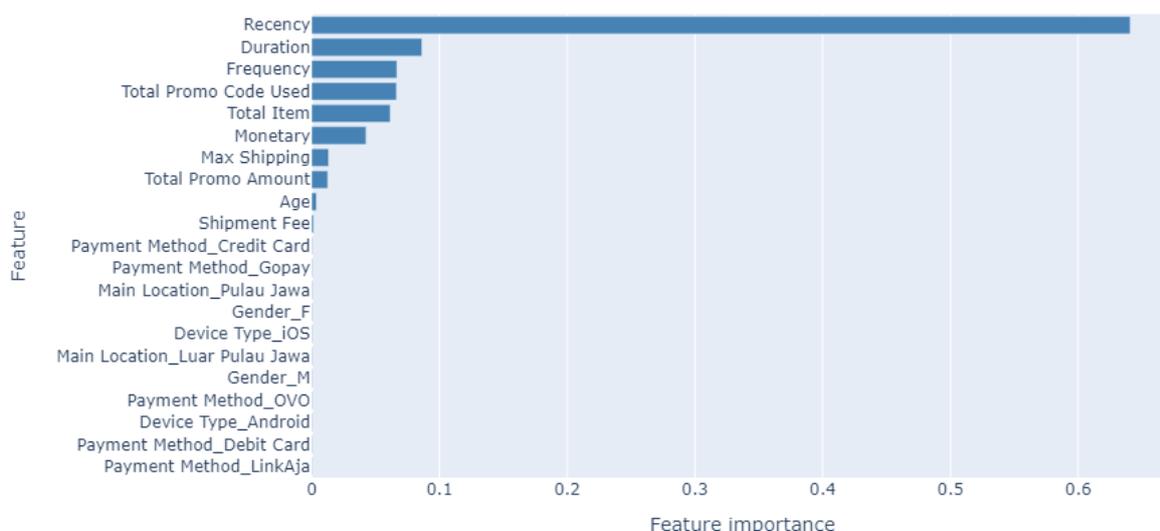
Setelah menentukan fitur penting yang akan digunakan pada tahap *data selection* dalam model machine learning, kemudian melakukan baseline models menggunakan *Cross Validation* dengan memprediksi customer yang berdasarkan data, customer yang tidak melakukan transaksi selama lebih dari 365 hari atau satu tahun akan dianggap sebagai 'churn'. Dengan kata lain, customer yang memiliki data recency di atas 365 akan dianggap sebagai 'churn'. Hasil Baseline Model yang telah dilakukan dapat dilihat pada tabel 3.

Tabel 3. Baseline Model

	Model	Accuracy training	Accuracy test	Precision	Recall	AUC	gap
0	DecisionTreeClassifier	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000
6	RandomForestClassifier	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000
7	GradientBoostingClassifier	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000
9	XGBClassifier	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000
8	ExtraTreesClassifier	1.000000	0.997890	0.998096	0.994894	0.997038	0.002110
2	KNeighborsClassifier	0.909018	0.869717	0.857640	0.680565	0.815923	0.039301
1	LogisticRegression	0.846958	0.846899	0.702661	0.883486	0.857303	0.000059
5	LinearSVC	0.826842	0.827413	0.746468	0.772885	0.811915	0.000572
4	SVC	0.776206	0.776117	0.592236	0.825162	0.790065	0.000089
3	GaussianNB	0.580958	0.580870	0.415821	0.965700	0.690313	0.000089

Pada tabel diatas Akurasi yang ditampilkan merupakan hasil dari proses Cross Validation K-Fold. Terdapat empat model yang memiliki keakuratan sempurna atau dengan akurasi 100% yaitu *DecisionTreeClassifier*, *RandomForestClassifier*, *GradientBoostingClassifier*, dan *XGBClassifier*. Top empat model tersebut merupakan asil prediksi baseline model yang paling bagus dibuktikan dengan mencoba melakukan compare model. Dari hasil baseline model, dipilih satu model terbaik yaitu *RandomForestClassifier*, karena model tersebut mampu membaca setiap feature numerik dari data dibandingkan model machine learning lainnya. Hasil Feature Importance pada model *RandomForestClassifier* dapat dilihat pada gambar 3.

Estimating feature importance through the RandomForestClassifier model



Gambar 3. Feature Importance model RandomForestClassifier

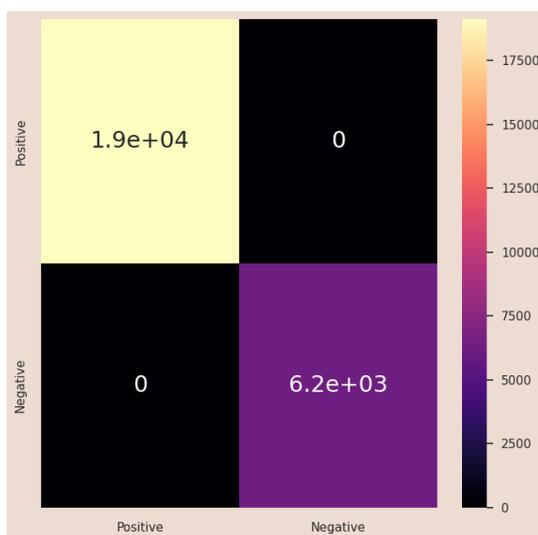
3.6 Evaluation

Hasil Modeling dari classification Report Testing pada model *RandomForestClassifier* menghasilkan nilai akurasi sebesar 1.00, precision 1.00, dan recall sebesar 1.00. Hasil dari modeling pada model tersebut dapat dilihat pada tabel 4.

Tabel 4. Classification Report Testing

	Precision	recall	f1_score	support
FALSE	1.00	1.00	1.00	7654
True	1.00	1.00	1.00	2487
accuracy			1.00	10141
macro avg	1.00	1.00	1.00	10141
weighted avg	1.00	1.00	1.00	10141

Dan hasil confusion Matrix Testing menghasilkan TP sebesar 7654, TN sebesar 2487, FP sebesar 0, dan FN sebesar 0. Hasil dari Confusion Matrix Testing pada model *RandomForestClassifier* dapat dilihat pada gambar 4.



Gambar 4. Confusion Matrix Testing

Dari 7654 diprediksi akan churn, kenyataannya memang churn. Kemudian dari 2487 diprediksi akan churn, namun kenyataannya tidak churn. Setelah melakukan baseline model dan menghasilkan sebuah akurasi, maka dilakukan Hyperparameter tuning agar model menghasilkan nilai yang lebih bagus. Hasil dari Hyperparameter Tuning dapat dilihat pada tabel 3.

Tabel 5. Hyperparameter Tuning

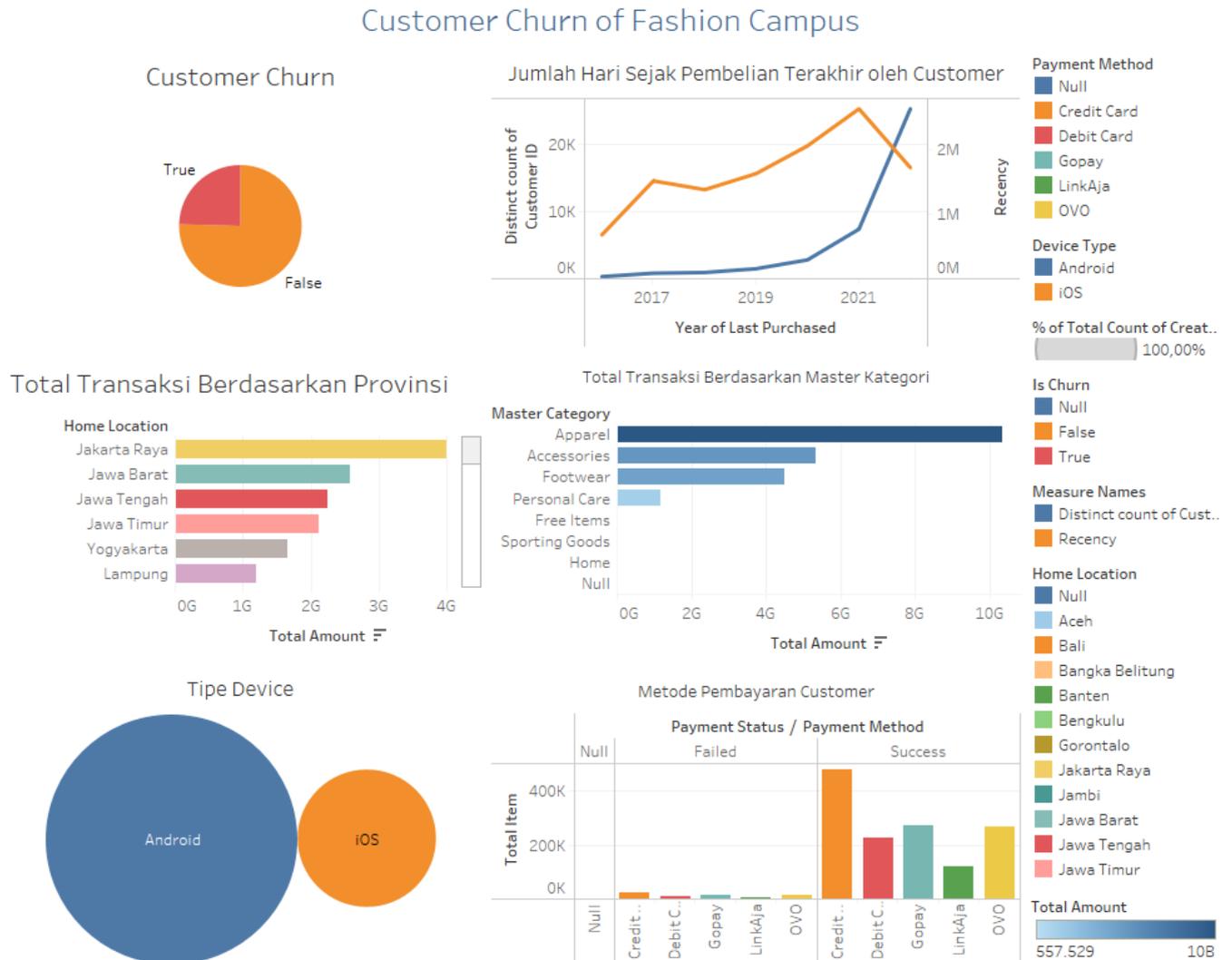
	Precision	recall	f1_score	support
FALSE	1.00	1.00	1.00	7654
True	1.00	1.00	1.00	2487
accuracy			1.00	10141
macro avg	1.00	1.00	1.00	10141
weighted avg	1.00	1.00	1.00	10141

Dari tabel tersebut terlihat bahwa akurasi masih sama dengan sebelum di Hyperparameter Tuning sebesar 1.00. Dengan begitu maka tidak perlu di Hyperparameter lagi karena model sudah memiliki akurasi sempurna sebesar 100%.

Ada beberapa rekomendasi bagi fashion campus untuk mengurangi churn dilihat dari hasil EDA pada gambar 2 yaitu dengan adanya kupon potongan harga, perusahaan harus meningkatkan lagi nilai dari potongan harga untuk menarik perhatian pelanggan agar melakukan pembelian. Kemudian tambahan promo, pelanggan yang churn merupakan pelanggan yang tidak kembali lagi ke platform untuk melakukan transaksi. Sehingga, pemberian promo yang menarik atau cashback secara berkala perlu diberikan agar menarik perhatian pelanggan, Lalu yang terakhir dengan Promo Kartu Kredit, pemberian promo yang lebih besar terhadap penggunaan credit card agar pelanggan tertarik dan dapat mengurangi churn karena pelanggan lebih banyak transaksi menggunakan credit card. Dengan beberapa rekomendasi dapat membantu perusahaan dalam mengambil suatu keputusan untuk mengurangi customer churn.

3.7 Deployment

Analisis visualisasi menggunakan Tableau dari data yang telah diolah dengan menggunakan python dilakukan dalam tahapan yang ada pada tahapan penelitian menghasilkan sebuah visualisasi yang dapat di tampilkan pada perusahaan fashion Campus. Visualisasi Dashboard dapat dilihat pada gambar 5.



Gambar 5. Visualisasi Dashboard

Dari Visualisasi dashboard menunjukkan bahwa customer yang churn sebesar 24.54% dan tidak churn sebesar 75.46%. Top 5 customer berasal dari provinsi Jakarta Raya, Jawa Barat, Jawa Tengah, Jawa Timur dan Jogja dengan total amonunt tertinggi 3.997.936.774.

4. KESIMPULAN

Berdasarkan hasil pengujian dengan menggunakan empat (4) model machine learning untuk memprediksi customer churn pada perusahaan fashion campus dengan memprediksi customer yang berdasarkan data, customer yang tidak melakukan transaksi selama lebih dari 365 hari atau satu tahun akan dianggap sebagai 'churn'. Dengan kata lain, customer yang memiliki data recency di atas 365 akan dianggap sebagai 'churn'. Mempunyai nilai akurasi masing-masing sebesar 100%, model Random Forest Classifier yang dipilih karena sistem mampu membaca setiap feature numerik dari data dibandingkan model lainnya. Sedangkan hasil analisis dilihat dari visualisasi dashboard menunjukkan bahwa customer yang churn sebesar 24,54% dan tidak churn sebesar 75,46%. Top 5 customer berasal dari provinsi Jakarta Raya, Jawa Barat, Jawa Tengah, Jawa Timur dan Jogja dengan total amonunt tertinggi 3.997.936.774.

UCAPAN TERIMA KASIH

Terima kasih kepada bapak Apriade Voutama, M.Kom selaku penulis kedua yang telah membimbing dan memberikan arahan dalam penyelesaian artikel ini dan memberikan arahan dalam penyelesaian artikel ini dan terima kasih juga kepada semua pihak yang terkait dan atas kerjasamanya.

DAFTAR PUSTAKA

- [1] M. Herawati, I. L. Wibowo, and I. Mukhlash, "Prediksi Customer Churn Menggunakan Algoritma Fuzzy Iterative Dichotomiser 3," *Limits J. Math. Its Appl.*, vol. 13, no. 1, p. 23, 2016, doi: 10.12962/j1829605x.v13i1.1913.
- [2] N. N. F. R, D. S. Anggraeni, and U. Enri, "Pengelompokan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient," *Temat. J. Teknol. Inf. Komun.*, vol. 9, no. 1, pp. 29–35, 2022, [Online]. Available: <http://jurnal.plb.ac.id/index.php/tematik/article/view/901>.
- [3] A. N. Rachmi, "Implementasi Metode Random Forest Dan Xgboost Pada Klasifikasi Customer Churn," 2020.
- [4] R. Novendri, R. Andreswari, and ..., "Implementasi Data Mining Untuk Memprediksi Customer Churn Menggunakan Algoritma Naive Bayes," *eProceedings ...*, vol. 8, no. 2, pp. 2762–2773, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/download/14678/14455>.
- [5] Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, "Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.
- [6] F. M. Almufqi, A. Voutama, S. Informasi, I. Komputer, U. S. Karawang, and T. Timur, "PERBANDINGAN METODE DATA MINING UNTUK MEMPREDIKSI PRESTASI," vol. 15, no. 1, pp. 61–66, 2023, doi: 10.30736/jt.v15i1.929.
- [7] M. R. Raharjo and A. P. Windarto, "Penerapan Machine Learning dengan Konsep Data Mining Rough Set (Prediksi Tingkat Pemahaman Mahasiswa terhadap Matakuliah)," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 317, 2021, doi: 10.30865/mib.v5i1.2745.
- [8] M. Mahendra, R. Chandra Telaumbanua, A. Wanto, and A. Perdana Windarto, "Akurasi Prediksi Ekspor Tanaman Obat, Aromatik dan Rempah-Rempah Menggunakan Machine Learning," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 6, pp. 207–215, 2022, doi: 10.30865/klik.v2i6.402.
- [9] S. Informasi, F. I. Komputer, U. Buana, and P. Karawang, "Prediction of Rice Field Planted Area with CRISP-DM Using Classification and Regression Tree (Cart) Algorithms," vol. 5, no. 1, pp. 578–587, 2023.
- [10] A. Yulianto and F. Firmansyah, "Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naive Bayes," *Remik*, vol. 6, no. 1, pp. 41–47, 2021, doi: 10.33395/remik.v6i1.11196.
- [11] A. H. Hasugian, M. Fakhriya, and D. Zukhoiriyah, "Volume 6 ; Nomor 1," *Januari*, vol. 6, pp. 98–107, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/index>.
- [12] A. M. Husein and M. Harahap, "Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning," *Data Sci. Indones.*, vol. 1, no. 1, pp. 8–13, 2021, doi: 10.47709/dsi.v1i1.1169.
- [13] D. Hadyan Tisantri, R. Cahya Wihandika, and S. Adinugroho, "Prediksi Keputusan Pelanggan Menggunakan Extreme Learning Machine Pada Data Telco Customer Churn," *J. Pengemb. Teknol. Inf. dan Ilmu Komputer*, vol. 3, no. 11, pp. 10516–10523, 2019.
- [14] A. P. Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 164, 2023, doi: 10.53513/jsk.v6i1.7351.
- [15] T. Tommy and A. M. Husein, "Model Prediksi Prestasi Mahasiswa Berdasarkan Evaluasi Pembelajaran Menggunakan Pendekatan Data Science," *Data Sci. Indones.*, vol. 1, no. 1, pp. 14–20, 2021, doi: 10.47709/dsi.v1i1.1168.
- [16] K. Ciptady, M. Harahap, J. Jonvin, Y. Ndruru, and I. Ibadurrahman, "Prediksi Kualitas Kopi Dengan Algoritma Random Forest Melalui Pendekatan Data Science," *Data Sci. Indones.*, vol. 2, no. 1, 2022, doi: 10.47709/dsi.v2i1.1708.