

CAPSTONE PROJECT

Machine Learning

Sebuah prediksi
tentang Churn Aspect

Dipresentasikan oleh Lie Benedict Yahliel



Latar Belakang Masalah

- PT SinergiNet Indonesia adalah perusahaan yang bergerak di bidang telekomunikasi, menyediakan layanan internet, telepon, dan solusi digital bagi pelanggan individu maupun bisnis.
- Dalam beberapa tahun terakhir, perusahaan menghadapi tantangan besar: meningkatnya tingkat customer churn, yaitu pelanggan yang memutuskan berhenti berlangganan.

Latar Belakang Masalah

- kondisi ekonomi nasional yang sedang lesu sehingga daya beli masyarakat menurun.
- kompetisi di industri telekomunikasi semakin ketat.
- Fenomena ini menurunkan pendapatan bulanan dan mengancam keberlanjutan bisnis, terutama karena biaya akuisisi pelanggan baru jauh lebih tinggi dibanding mempertahankan pelanggan lama.

Diperlukan

- Model machine learning yang mampu memprediksi pelanggan yang berpotensi churn pada bulan berikutnya.
- Memanfaatkan berbagai variabel seperti durasi berlangganan (tenure), jenis kontrak, status layanan tambahan (OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport), serta pola pembayaran dan biaya bulanan.

Target dan Manfaat

- Target utama dari model ini adalah tim pemasaran dan divisi retensi pelanggan, yang akan menggunakan hasil prediksi untuk merancang strategi intervensi seperti penawaran khusus, peningkatan layanan, atau program loyalitas.
- Dengan pendekatan ini, perusahaan diharapkan dapat mengurangi tingkat churn, menjaga pendapatan, dan meningkatkan kepuasan pelanggan di tengah tekanan ekonomi.

Data yg dianalisa



1. Dependents: Status tanggungan pelanggan
2. Tenure: Lama waktu (dalam bulan) pelanggan telah berlangganan di perusahaan.
3. OnlineSecurity: Status layanan keamanan online.
4. OnlineBackup: Status layanan pencadangan data online.
5. InternetService: Status langganan layanan internet (apakah pelanggan berlangganan atau tidak).

Data yg dianalisa



- 6. DeviceProtection: Status perlindungan perangkat
- 7. TechSupport: Status dukungan teknis
- 8. Contract: Jenis kontrak berdasarkan durasi berlangganan.
- 9. PaperlessBilling: Status penagihan tanpa kertas
- 10. MonthlyCharges: Besaran biaya layanan per bulan.
- 11. Churn: Status berhenti berlangganan

Kolom yang Dianalisa dan Target

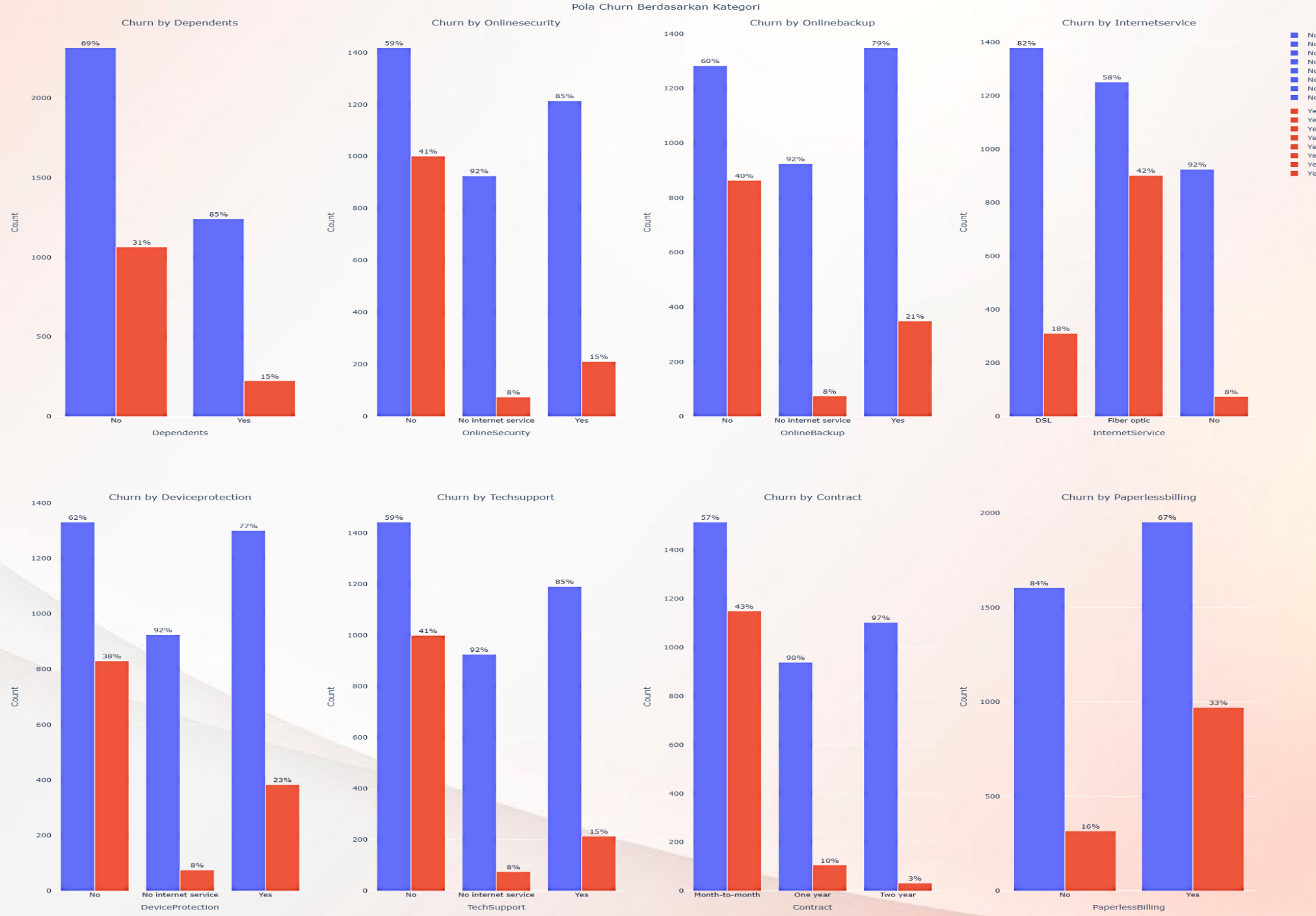
- Kolom yang dianalisa

Dependents	Device Protection
Tenure	Technology Support
Online Security	Contract
Online Back up	Paperless Billing
Internet Service	Monthly Charges

- Kolom Target: Churn



Temuan Data



Faktor Churn tertinggi

3 Faktor Churn tertinggi:

(1) **Contract**

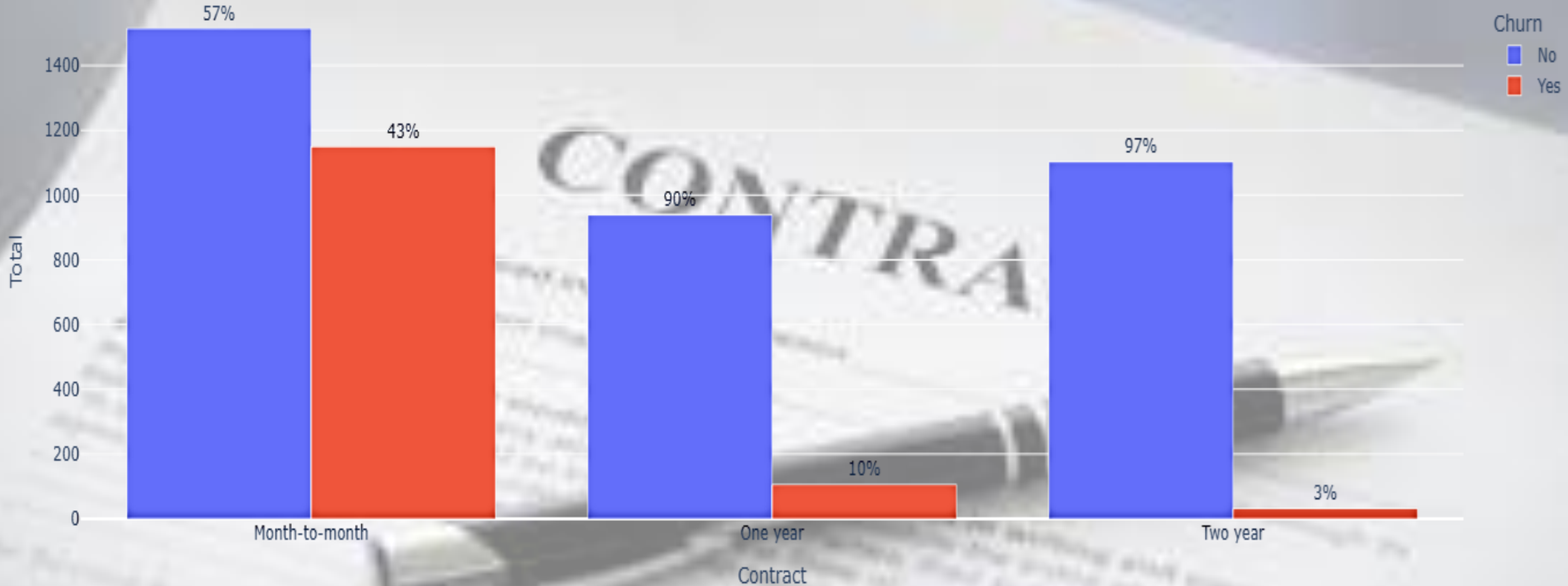
(2) **Tenure**

(3) **Internet Service**



Temuan Data: Contract

Pola Churn Menurut Kategori Contract

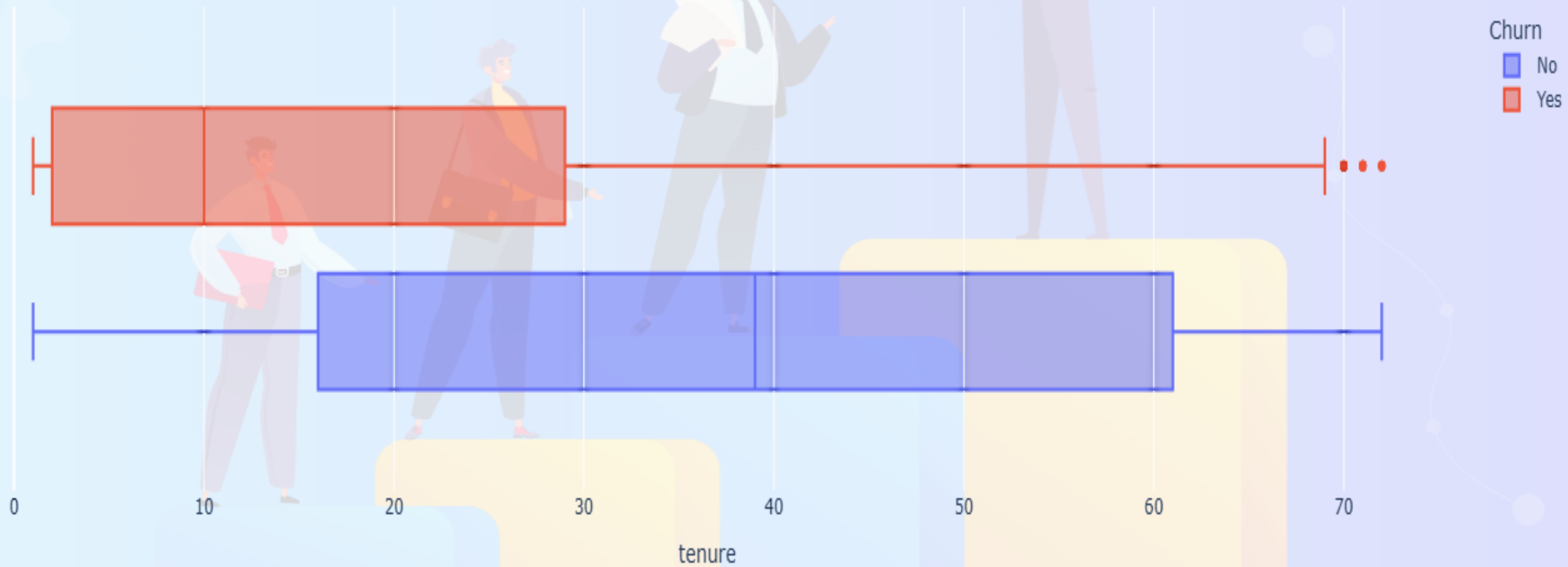


Insight & Saran: Contract

- Pelanggan dengan kontrak bulanan memiliki **risiko churn tertinggi hingga 43%**, menandakan ketidakstabilan loyalitas pelanggan terjadi pada kontrak bulanan.
- kontrak dalam dua tahun hanya 3% churn, menunjukkan loyalitas pelanggan yang jauh lebih tinggi. Semakin panjang durasi kontrak, tingkat churn menurun drastis
- Strategi retensi harus fokus pada peningkatan **konversi pelanggan bulanan ke kontrak jangka panjang** dengan **insentif** dan **komunikasi** yang jelas.

Temuan Data: Tenure

Distribusi Tenure Berdasarkan Churn

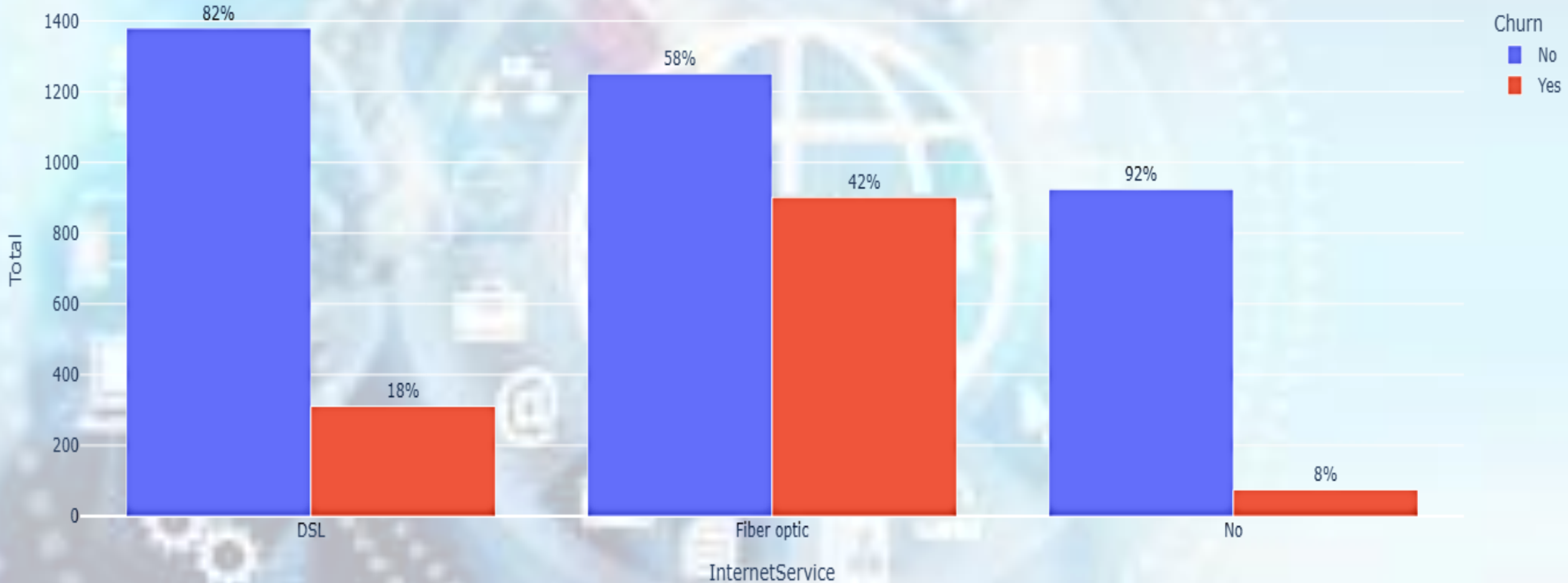


Insight & Saran: Tenure

- Pelanggan yang berada di bawah 10 bulan menempati churn rate yang tertinggi. Sementara pelanggan di atas 10 bulan cukup stabil.
- Mereka adalah orang-orang yang masih mempertimbangkan seberapa besar benefit dari berlangganan.
- Perlu memikirkan **pola insentif** dan **perluasan bentuk layanan** agar pelanggan baru merasa tertarik dan merasa diuntungkan dengan layanan yang ditawarkan.

Temuan Data: Internet Service

Pola Churn Menurut Kategori Internetservice



Insight & Saran: Internet Service

- Meskipun Fiber Optic adalah layanan paling strategis dan canggih, tetapi **churn rate hampir 42%** mengindikasikan ada masalah yang perlu diselidiki lebih lanjut.
- Walau teknologi DSL lebih tua, dan churn rate hanya sekitar 18%, perkembangan teknologi akan mengarah ke penggunaan fiber optic.
- **Perlu penguatan di sektor Fiber optic** dengan memberikan **kualitas** dan **service** yang baik sehingga menarik untuk pelanggan DSL beralih ke Fiber optic.

Proses Data Cleaning

- Mencari dan menghapus duplikat pada kolom
- Menghapus tenure(jumlah bulan berlangganan) baru berlangganan kurang dari 1 bulan karena loyalitasnya masih belum terbentuk sama sekali

Asumsi False Positif (FP) dan False Negatif (FN) Berdasar kolom Churn

False Positive(FP):

- Model memprediksi pelanggan akan churn (berhenti berlangganan), tapi kenyataannya pelanggan tersebut tidak churn (masih berlangganan)
- Akibatnya: perusahaan alokasikan sumber daya, biaya promosi, atau program retensi kepada pelanggan yang sebenarnya tidak berencana berhenti, sehingga terjadi **pemborosan anggaran dan waktu** tanpa memberikan dampak signifikan terhadap penurunan churn

Asumsi False Positif (FP) dan False Negatif (FN) Berdasar kolom Churn

False Negatif (FN):

- Model memprediksi pelanggan tidak churn (akan tetap berlangganan), tapi kenyataannya pelanggan tersebut justru churn.
- Akibatnya adalah perusahaan gagal melakukan **intervensi atau strategi retensi** kepada pelanggan yang sebenarnya berisiko tinggi untuk berhenti, sehingga kehilangan pendapatan dan potensi loyalitas jangka panjang dari pelanggan tersebut.

Prediksi Biaya

False Negatif (FN):

- Pendapatan yang hilang di bulan depan per orang: $(230.000 + 250.000 + 325.000 + 490.000) / 4 = \text{Rp } 323.750,00$

Prediksi Biaya

False Positive(FP):

- **Biaya pemeliharaan Jaringan internet:**
 $(500.000+2.500.000)/2=\text{Rp } 1.500.000,00$
- **Biaya pemasangan Fiber optik:**
 $(5.000.000+15.000.000)/2=\text{Rp } 10.000.000,00$
- **Biaya pelayanan service melalui whatsapp bot: Rp**
1.560.000,00
- **Total Biaya yang dikeluarkan=**
 $1.500.000+10.000.000+1.560.000=\text{Rp } 13.060.000$

Prediksi Biaya

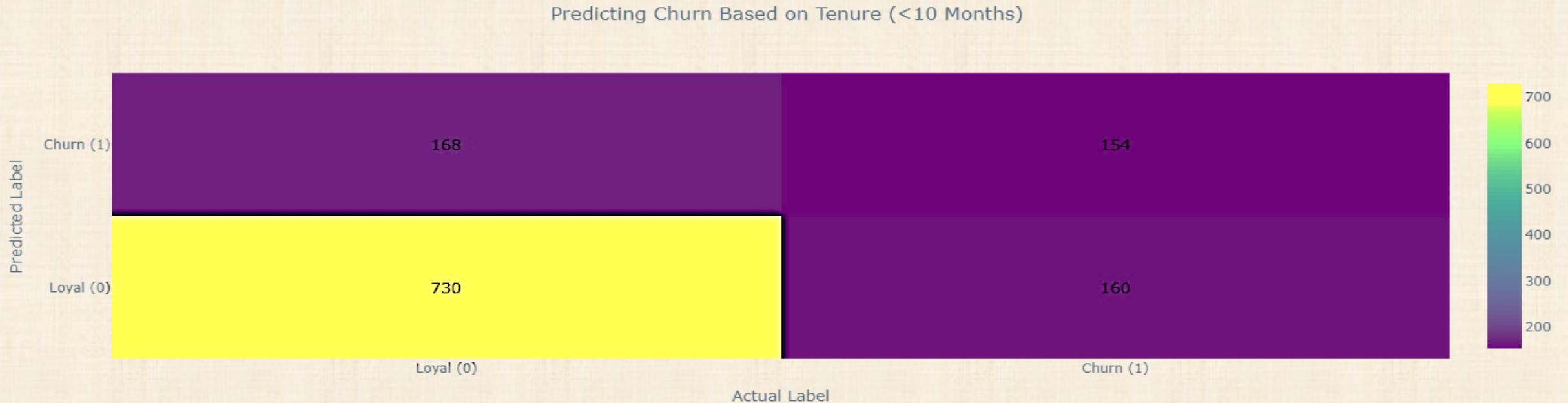
False Positive(FP):

- **Biaya retensi per orang** = $13.060.000 / TP = X / \text{orang (Pelanggan yang hasil prediksi dan actualnya loyal)}$
- Jika TP minimal 100 orang, maka biaya untuk retensi pelanggan sudah **jauh di bawah biaya False Negative**
- Data akan difokuskan untuk mendapatkan recall setinggi mungkin supaya pelanggan yang churn dapat diprediksi dengan setepat mungkin

Sumber Referensi Biaya Retensi

- <https://www.telkomsel.com/jelajah/jelajah-lifestyle/ini-harga-wifi-indihome-terbaru-tahun-2025-ih>
- <https://www.gbs-indonesia.co.id/news/biaya-instalasi-jaringan-fiber-optik-terbaru>
- <https://www.pilarsupport.com/biaya-pemeliharaan-jaringan-internet-dan-apa-saja-layanannya/>
- <https://eva.id/pricing/>
- <https://www.ruangmenyala.com/article/read/retensi-adalah>

Prediksi berdasar Tenure



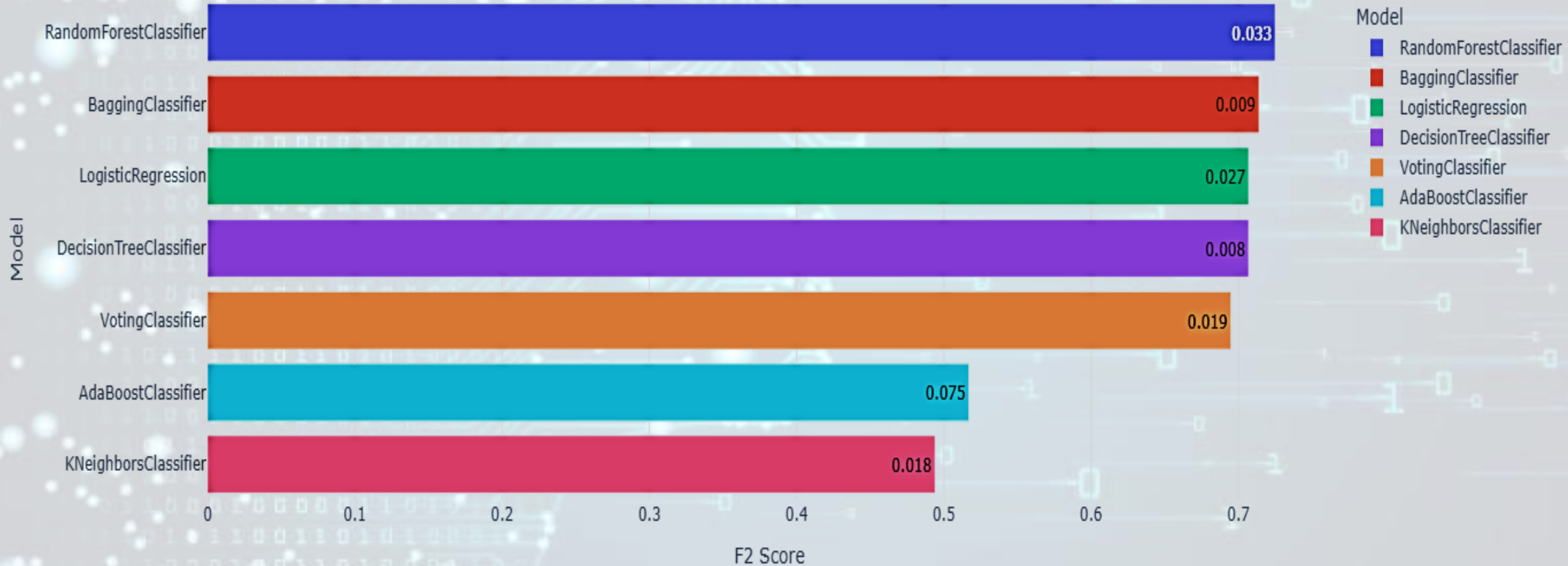
- Diketahui dari EDA Tenure bahwa pelanggan yang akan churn biasanya terdapat pada bulan 1-10
- Sebelum Machine Learning, perusahaan tidak bisa tahu siapa yang akan loyal di jangka panjang, sehingga semua customer diatas 10 bulan dianggap loyal

Prediksi berdasar Tenure

- Kerugian yang diterima oleh perusahaan:
 $\text{FN} * \text{Pendapatan hilang} = 160 * 323.750 = \text{Rp } 51.800.500,00$
- Perusahaan **mengalami kerugian 52 Juta** rupiah karena gagal memberikan promosi kepada customer untuk tetap berlangganan.

Pemilihan Machine Learning

Hasil Cross Validasi Machine Learning Dengan Semua Feature



Pemilihan Machine Learning

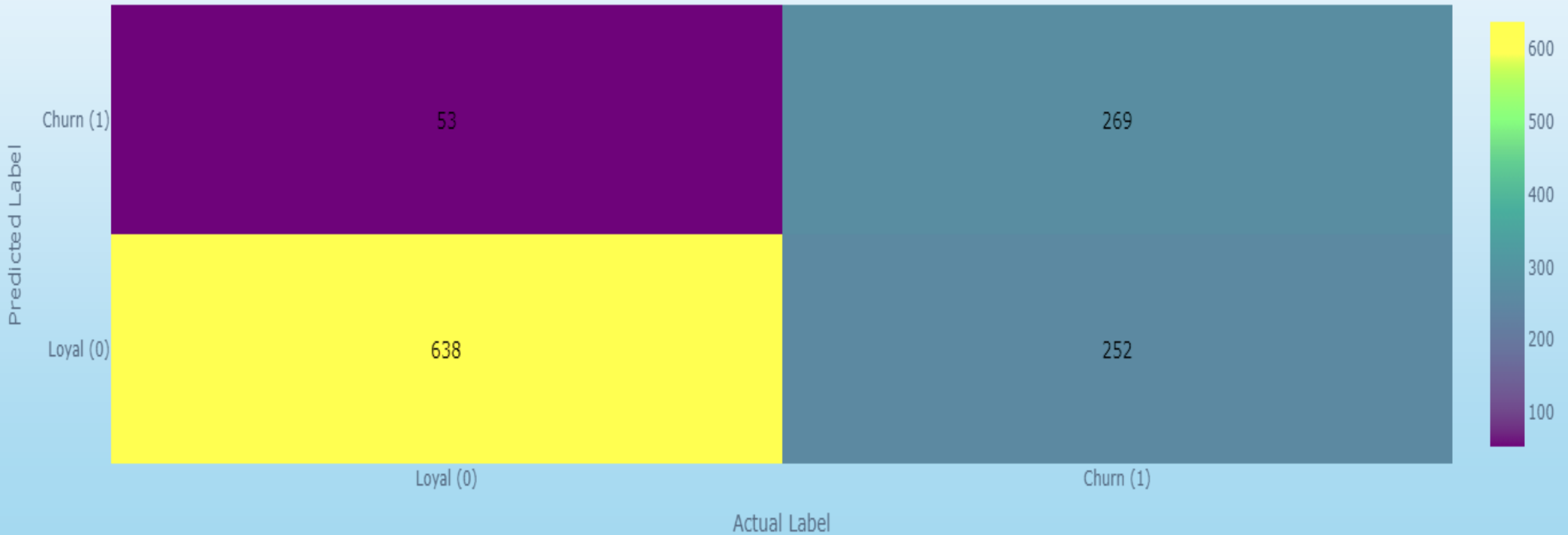
- **F2 Score** digunakan dalam machine learning untuk mengukur kemampuan model dalam memprediksi pelanggan yang akan churn, dengan fokus mengurangi *False Negative*. Semakin tinggi F2 Score, semakin baik kemampuan model dalam mengidentifikasi churn secara tepat.
- Angka pada grafik menunjukkan **mean** pada **standar deviasi (std)**. Nilai std yang lebih kecil menandakan model lebih stabil.

Pemilihan Machine Learning

- **Random Forest** memberikan performa terbaik dengan mean score tertinggi dan variasi (std) relatif rendah, mencerminkan konsistensi yang baik di berbagai fold.
- **Bagging Classifier** memiliki mean score sedikit di bawah Random Forest, namun unggul pada stabilitas dengan std score terkecil di antara seluruh model.
- Model lain, baik dari segi rata-rata skor maupun konsistensi, belum mampu menandingi performa dua model tersebut.
- Fokus tuning diarahkan pada **Random Forest** dan **Bagging Classifier** untuk memaksimalkan akurasi prediksi churn.

All Feature: Random Forest

Confusion Matrix — Random Forest(All Features)



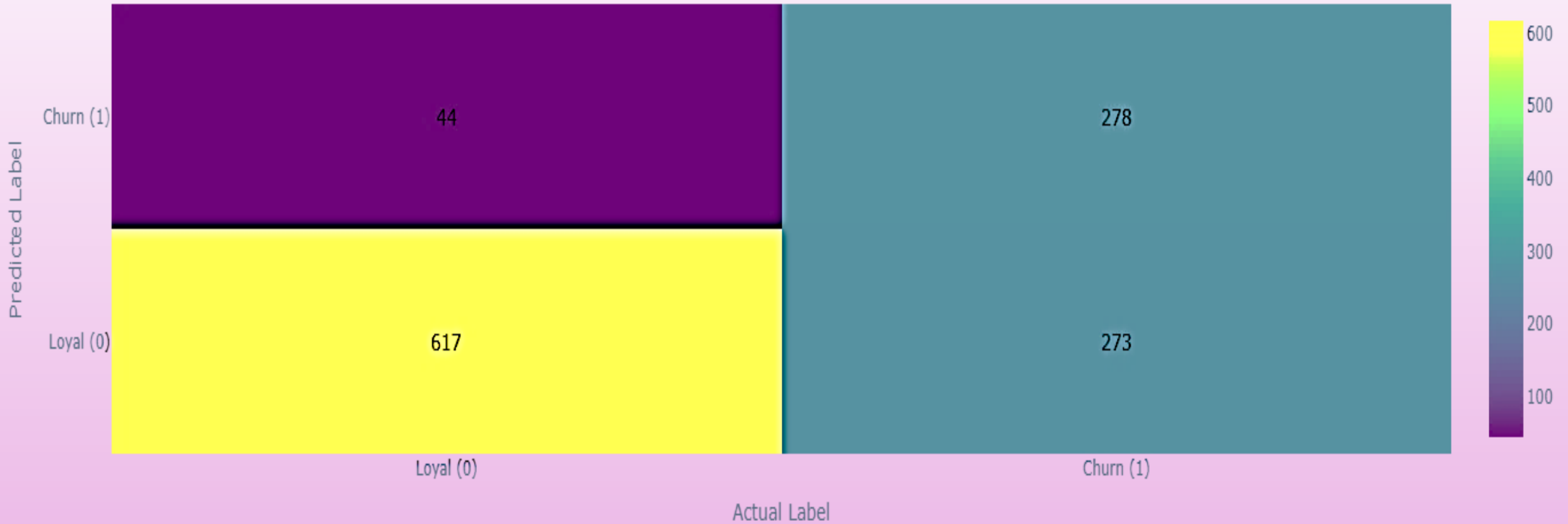
All Feature: Random Forest

Biaya promosi yang tidak perlu dikeluarkan:

- $FP * \text{Biaya Promosi} / TP = 53 * 13.060.000 / 608 = \text{Rp } 1.084.922,00$
- Kerugian karena kehilangan customer:
- $FN * \text{Pendapatan hilang} = 252 * 323.750 = \text{Rp } 81.585.000,00$
- Total kerugian yang ditanggung perusahaan:
- $1.084.922 + 81.585.000 = \text{Rp } 82.669.922,00-$

All Feature: Bagging Classification

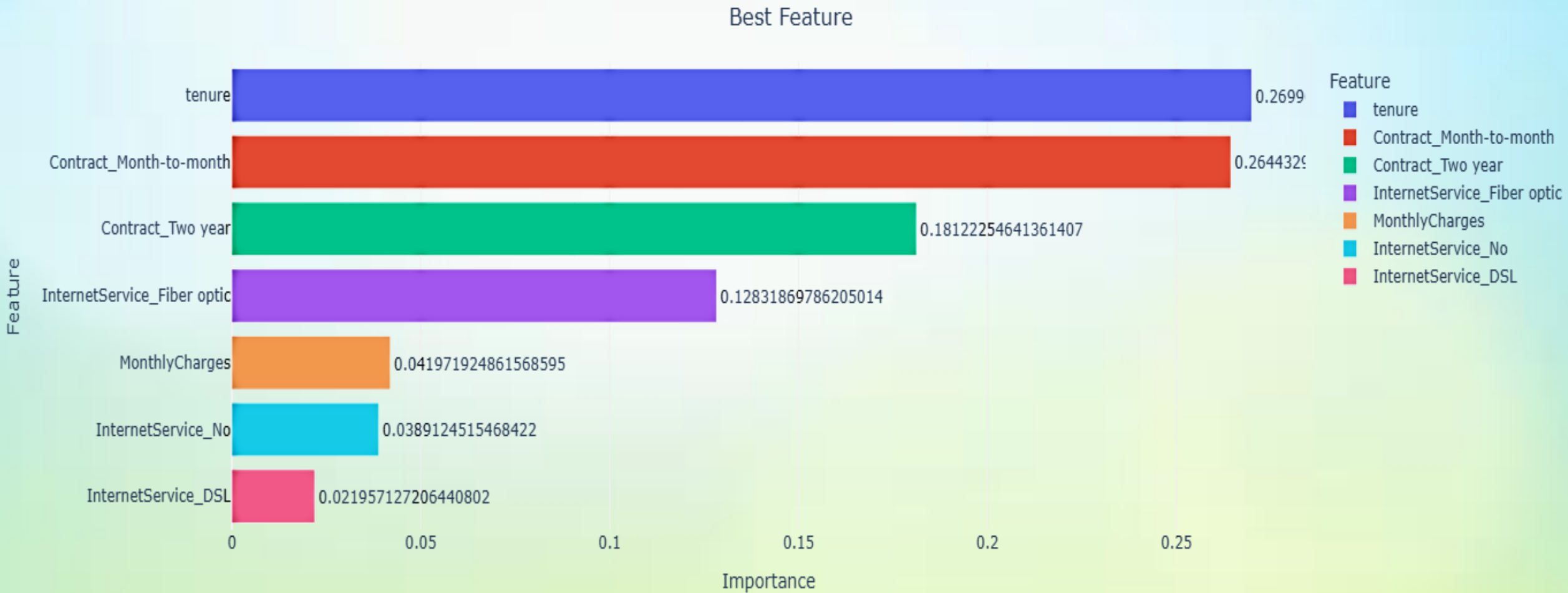
Confusion Matrix — Bagging Classifier(All Features)



All Feature: Bagging Classification

- Kerugian yang ditanggung Dengan machine learning Bagging Classification(All Feature):
- Biaya promosi yang tidak perlu dikeluarkan:
- $FP * \text{Biaya Promosi} / TP = 44 * 13.060.000 / 617 = \text{Rp } 931.345,00$
- Kerugian karena kehilangan customer:
- $FN * \text{Pendapatan hilang} = 273 * 323.750 = \text{Rp } 88.383.500,00$
- Total kerugian yang ditanggung perusahaan:
 $931.345 + 88.383.500 = \text{Rp } 89.314.845,00$

Feature Selection

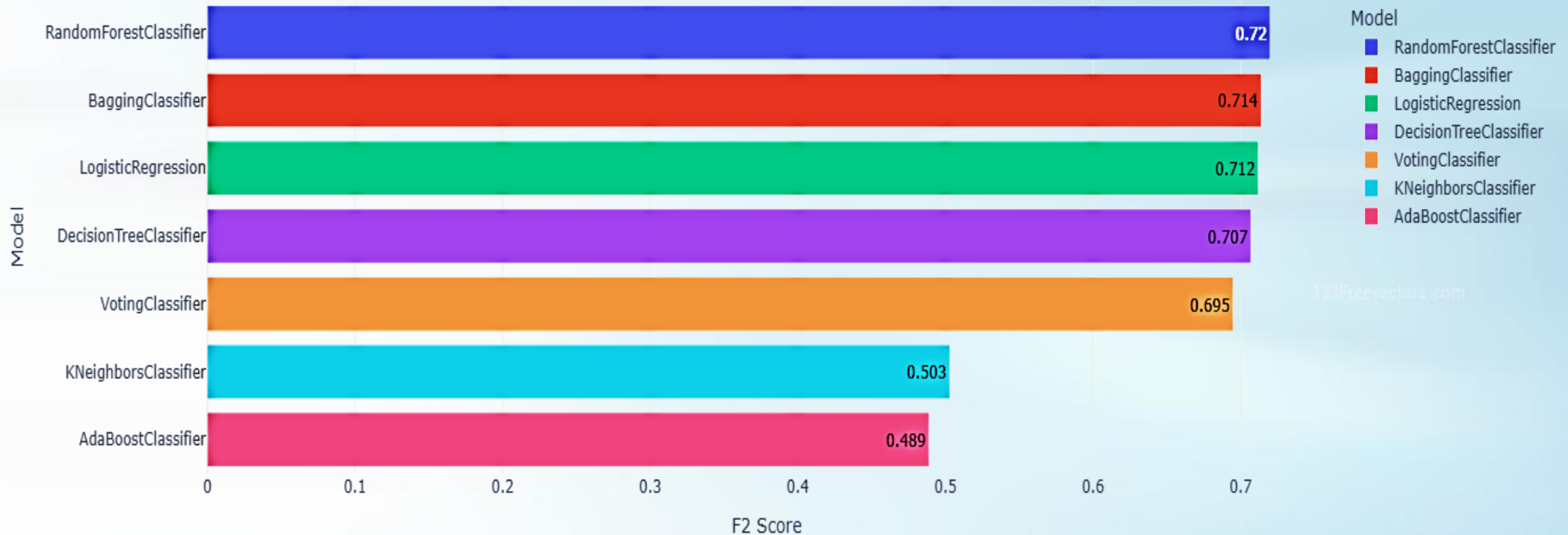


Feature Selection

- Feature yang akan diuji selanjutnya dipilih berdasarkan nilai importance di atas 0.01 karena fitur-fitur tersebut memberikan kontribusi signifikan terhadap prediksi model.
- Terdapat 4 dari 11 fitur yang dianggap paling relevan, yaitu: **tenure, contract, internet service, monthly charges**
- Fokus pada fitur-fitur ini mungkin bisa meningkatkan efisiensi model dengan mengurangi kompleksitas tanpa mengorbankan performa prediksi

Cross-Validation Feature Selection

Hasil Cross Validasi Machine Learning Dengan Feature Selection

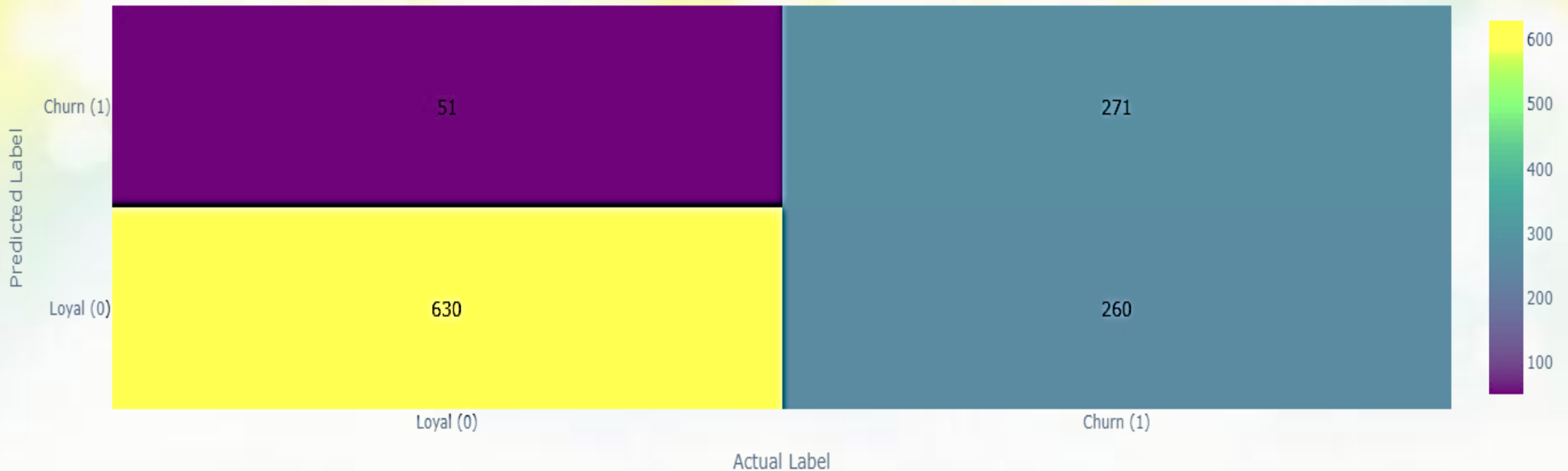


Cross-Validation Feature Selection

- **Random Forest** memberikan performa terbaik dengan mean score tertinggi dan variasi (std) relatif rendah, mencerminkan konsistensi yang baik di berbagai fold.
- **Bagging Classifier** memiliki mean score sedikit di bawah Random Forest, namun unggul pada stabilitas dengan std score terkecil di antara seluruh model.
- Model lain, baik dari segi rata-rata skor maupun konsistensi, belum mampu menandingi performa dua model tersebut.
- Fokus tuning diarahkan pada **Random Forest** dan **Bagging Classifier** untuk memaksimalkan akurasi prediksi churn.

Future Selection: Random Forest

Confusion Matrix — Random Forest (Feature Selection)



Future Selection: Random Forest

- Dengan machine learning Random Forest (Feature Selection):
- Biaya promosi yang tidak perlu dikeluarkan:
- $FP * \text{Biaya Promosi} / TP = 51 * 13.060.000 / 630 = \text{Rp } 1.057.239,00$
- Kerugian karena kehilangan customer:
- $FN * \text{Pendapatan hilang} = 260 * 323.750 = \text{Rp } 84.175.500,00$
- Total kerugian yang ditanggung perusahaan:
 $1.057.239 + 84.175.500 = \text{Rp } 85.232.739,00$

Future Selection: Bagging Classification

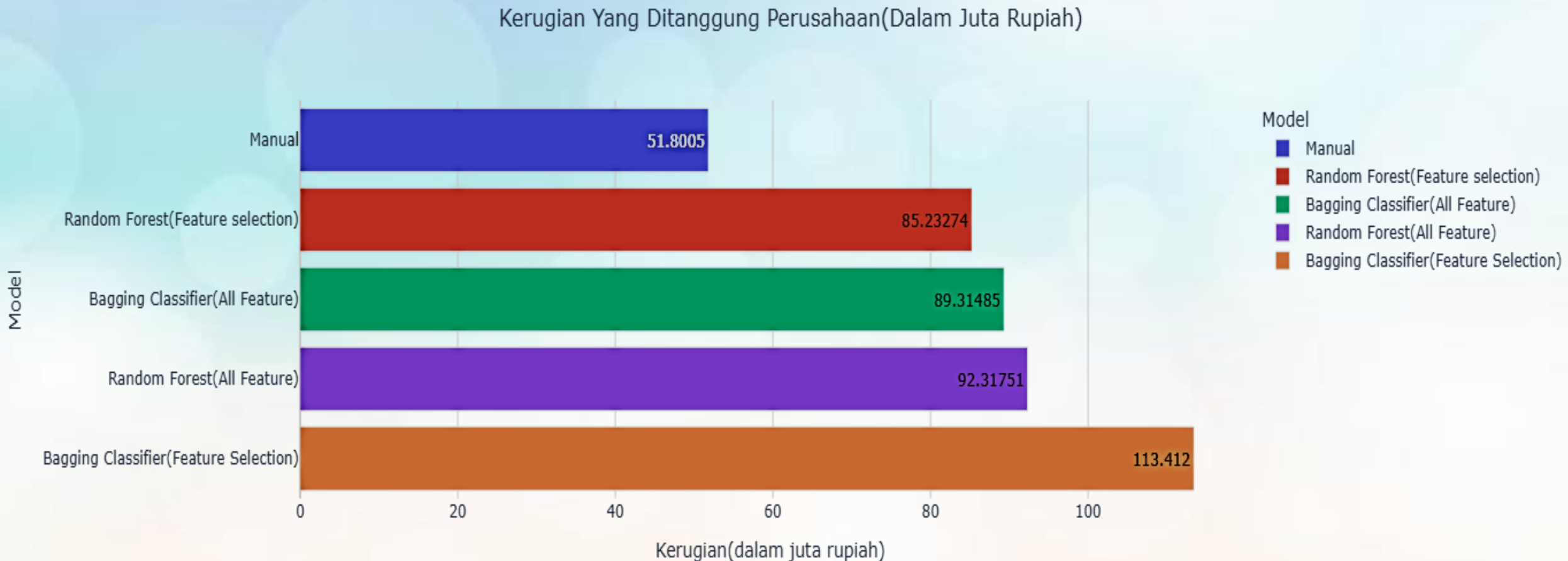
Confusion Matrix — Bagging Classifier (Feature Selection)



Future Selection: Bagging Classifier

- Kerugian yang ditanggung Dengan machine learning Bagging Classification (Feature Selection):
- Biaya promosi yang tidak perlu dikeluarkan:
- $FP * \text{Biaya Promosi} / TP = 31 * 13.060.000 / 542 = \text{Rp } 746.974,00$
- Kerugian karena kehilangan customer:
- $FN * \text{Pendapatan hilang} = 348 * 323.750 = \text{Rp } 112.665.000,00$
- Total kerugian yang ditanggung perusahaan:
 $746.974 + 112.665.000 = \text{Rp } 113.411.974,00$

Kerugian yang Ditanggung



Kerugian yang Ditanggung

- Kerugian terendah tercatat pada metode manual, yaitu sebesar 52 juta rupiah
- Random Forest dengan semua feature dengan kerugian sebesar 82 juta rupiah, menjadi metode terbaik di antara model machine learning yang diuji
- Random Forest dengan feature selection menempati posisi urutan kedua dengan kerugian sebesar 85 juta rupiah
- Bagging Classifier dengan semua fitur menempati posisi ketiga dengan kerugian 89 juta rupiah
- **Bagging Classifier dengan feature selection** justru memberikan kerugian paling tinggi (113 juta rupiah)

Kerugian yang Ditanggung

Rekomendasi optional:

Metode manual merupakan opsi yang terbaik saat ini karena mencatat kerugian yang paling kecil untuk penanganan retensi pelanggan

Meskipun metode manual saat ini mencatat kerugian paling kecil, machine learning **tetap perlu dipertimbangkan** karena dapat memberikan scalability dan konsistensi prediksi dalam volume data besar

Dari opsi machine learning yang diuji, **Bagging Classifier (All Feature) dipilih untuk diadopsi**. Alasannya, meskipun lebih besar 7 juta rupiah dari Random Forest dengan semua feature, model ini menunjukkan **stabilitas prediksi** yang lebih konsisten sehingga berpotensi menekan biaya after care dalam jangka panjang.