Laporan Final Project – Infinity

Laporan ini mencakup laporan hasil kegiatan yang dilakukan mulai dari Stage 0 sampai Stage 4. Hasil kegiatan yang dilakukan pada masing-masing stage adalah sebagai berikut.

1. Stage 0

Pada stage ini ditetapkan dataset yang akan digunakan sampai stage akhir yaitu dataset Online Shoppers Intention. Kemudian ditetapkan problem, goals, objectives, dan business metrics dengan penjelasan sebagai berikut.

1.1. Problem

Perusahaan Leslar memiliki conversion rate diangka 15% pada tahun pertama. Pada tahun kedua perusahaan menargetkan Peningkatan conversion rate sebesar 10%. Tim data scientist diminta membuat model untuk membantu mencapai target tersebut.

1.2. Goals

Meningkatkan pertumbuhan conversion rate pada tahun kedua sebesar 10%.

1.3. Objectives

a. Menganalisis korelasi antar atribut terhadap pembelian.
b. Membuat model machine learning yang dapat memprediksi apakah seorang visitor akan purchase atau tidak.

1.4. Business Metrics

Conversion Rate

2. Stage 1

Inti dari kegiatan Stage 1 adalah melakukan Exploratory Data Analysis untuk mengetahui integritas data, distribusi data, korelasi antar fitur, serta menemukan beberapa business insight. Adapun hasil kegiatan yang dilakukan pada stage ini adalah sebagai berikut.

2.1. Descriptive Statistics Analysis

Langkah analisis statistik deskriptif dilakukan untuk semua kolom pada dataset Online Shoppers Intention. Analisis statistik deskriptif terbagi menjadi dua bagian yaitu untuk kolom-kolom numerik dan kolom-kolom kategorik. Hasil analisis yang dilakukan terhadap kolom-kolom deskriptif adalah sebagai berikut.

1. Semua fitur numerik (Administrative, Administrative_Duration, Informational, Informational_Duration, ProductRelated, ProductRelated_Duration, BounceRates, ExitRates, PageValues, dan SpecialDay) memiliki nilai mean dan median yang berbeda. Kemungkinan fitur-fitur tersebut memiliki outlier atau tidak menyebar normal (skewed).

2. Semua fitur yang disebutkan pada poin 1 juga memiliki nilai maksimum yang cukup jauh dari nilai Q3 yang mengindikasikan bahwa fitur-fitur tersebut memiliki outlier.
3. Beberapa fitur (Informational, Informational_Duration, PageValues, SpecialDay) terlihat didominasi oleh nilai 0 dan memiliki nilai maksimum yang cukup besar. Indikasi kuat bahwa fitur-fitur tersebut memiliki outlier.

Kemudian hasil analisis yang dilakukan untuk kolom-kolom kategorik adalah sebagai berikut.

1. Kolom target (Revenue) memiliki 2 unique value namun sebagian besar (sekitar 85%) bernilai False. Bisa dipertimbangkan untuk penanganan imbalance.
2. Beberapa fitur kategorik yaitu Browser, Region, dan TrafficType memiliki terlalu banyak unique value. Beberapa value minor mungkin bisa dikategorikan sebagai Other.
3. Terdapat fitur-fitur yang didominasi oleh satu unique value saja seperti OperatingSystems, Browser, VisitorType, dan Weekend. Fitur-fitur tersebut mungkin bisa tidak disertakan dalam model.

Semua penemuan yang didapatkan pada analisis statisktik deskriptif akan lebih terlihat jelas pada tahap univariate analysis.

2.2. Univariate Analysis

Univariate Analysis menggunakan visualisasi untuk melihat distribusi masing-masing kolom baik fitur maupun target pada dataset Online Shoppers Intention. Dalam melakukan visualisasinya dibagi menjadi tiga bagian yaitu numeric (nums), category (cats), dan encode. Hasil analisis univariat yang dilakukan adalah sebagai berikut.

1. Univariate Analysis Num Fields (Numeric)
   a. Boxplot Num Fields

      Selain untuk melihat bentuk sebaran/kepadatan data, fungsi lain boxplot ialah untuk melihat outliers secara lebih teliti. Oleh karena itu, meskipun di kolom Informational, Informational_Duration, PageValues, SpecialDay distribusinya sangat tipis karena sebaran datanya bersifat menumpuk, namun di kolom-kolom tersebut memiliki nilai outliers yang cukup banyak.

      - Administrative = kebanyakan pengguna mengunjungi page ini kurang dari 5 kali, bahkan cenderung jarang yang mengunjungi (nol kali).
      - Administrative_Duration = pengguna yang banyak menghabiskan waktu mengunjungi page sekitar di bawah 250 detik.
      - Informational = kebanyakan pengguna sangat jarang mengunjungi page ini, jumlah kunjungan sangat dominan di nol (itu sebabnya saat diplot, distribusinya terlihat tipis dan menumpuk di nol).

- Informational_Duration = durasi pengguna yang mengunjungi page ini sangat dominan di nol detik.
- ProductRelated = pengguna banyak mengunjungi web sekitar di bawah 50 page.
- ProductRelated_Duration = pengguna banyak menghabiskan waktu untuk mengunjungi page sekitar di bawah 5000 detik.
- BounceRates = persentase pengguna banyak mengunjungi web di bawah 0.025.
- ExitRates = persentase pengguna banyak mengunjungi web di antara 0.020 sampai 0.050.

b. Displot Num Fields

Semua distribusinya menghasilkan positive skewed dan terdapat beberapa lonjakan kecil seperti pada 'Informational', 'BounceRates', 'ExitRates', 'SpecialDay', dan juga terdapat long tail pada 'Administrative', 'Administrative_Duration', 'Informational', 'Informational_Duration', 'ProductRelated', 'ProductRelated_Duration', 'PageValues'. Dari plot distribusi, dapat disimpulkan di antara ketiga page di web, pengguna lebih sering mengunjungi page ProductRelated dibandingkan dua page lainnya. Hal ini bisa dilihat pada range sebaran jumlah kunjungan (x axis) pada page ProductRelated yang sebarannya mencapai ratusan.

2. Univariate Analysis Cat Fields (Category)
   - Month: dalam data ini satu tahun hanya terdapat 10 bulan dimana pada bulan January dan April itu tidak masuk. Pengguna yang mengunjungi web banyak terdapat pada bulan May dan November dan pengguna yang sepi mengunjungi web terdapat pada bulan February.
   - VisitorType: terdapat 3 tipe pengunjung dimana yang paling mendominasi yaitu pada Returning_visitor yang lebih dari 10000 pengunjung, New_visitor terdapat kurang dari 2000, dan Other yang paling minimum.
   - Weekend: pengguna yang mengunjungi dalam weekend terlihat jauh lebih sedikit yaitu sekitar 3000 dibandingkan yang tidak mengunjungi mencapai lebih dari 8000.
   - Revenue: pengguna yang Purchase jauh lebih sedikit yaitu sekitar 2000 dibandingkan dengan pengguna yang NotPurchase mencapai lebih dari 10000

3. Univariate Analysis Encoded
   - OperatingSystems = pada distribusinya terdapat 3 top OperatingSystem yaitu pada tipe 1, 2, dan 3.
   - Browser = pada distribusinya paling dominan banyak menggunakan tipe Browser 2.

- Region = pada distribustinya paling banyak menggunakan di tipe Region 1.
- TrafficType = pada distribusinya paling banyak menggunakan tipe TrafficType 2

Hal yang harus di follow-up untuk data pre-processing dikarenakan pada data ada yang right skewed dan mempunyai long tail makakami perlu menggunakan log Transformasi agar hasil distribusi mendekati distribusi normal.

2.3. Multivariate Analysis

Multivariate Analysis berfokus pada hubungan antar fitur dengan fitur serta fitur terhadap target. Visualisasi dengan heatmap merupakan salah satu cara untuk melihat hubungan serta korelasi antar fitur dan target. Berikut adalah hasil yang diperoleh menggunakan heatmap:

1. Page value memiliki korelas positif yang tinggi terhadap target. Semakin tinggi page value, semakin tinggi juga kemungkinan visitor untuk purchase.
2. Bounce rate dan exit rate memiliki korelasi negatif terhadap target yang artinya semakin tinggi bouncerate dan exitrate semakin kecil juga kemungkinan visitor untuk purchase.
3. Terdapat beberapa fitur yang berpotensi redundan, (bouncerate-exitrate) memiliki nilai korelasi 0.91 dan fitur (Administrative, Informational, dan Product related terhadap Duration) memiliki nilai korelasi dalam range 0.6 – 0.86.

***notes** : dikarenakan tipe data target adalah bool atau bukan numerical,kami belum bisa memvalidasi keterhubungan fitur-fitur dengan target dengan lebih jelas.

Melakukan analysis categorical terhadap target dengan membuat visual countplot untuk melihat lebih jelas hubungan target yang merupakan tipe categorical terhadap fitur categorical. Berikut adalah hasil yang didapatkan:

1. Segmentasi berdasarkan tipe visitor, returning_visitor merupakan tipe yang paling banyak berkunjung. Tetapi, new_visitor memiliki presentasi lebih besar dalam melakukan purchasing jika dilihat berdasarkan jumlahnya pada masing-masing tipe.
2. Ada beberapa tipe traffic yang mendominasi dari jumlah sampai kemungkinan melakukan pembelian.
3. Operating system dan browser hanya didominasi 2 sampai 3 tipe saja.
4. Jumlah visitor berdasarkan region sangat berbeda jauh dengan dimana hanya didominasi 3 daerah saja.

Untuk mendapatkan insight lebih mengenai hubungan target dan feature dibutuhkan penanganan outlier atau melakukan transformasi. Kami juga dapat mengubah kolom yang merupakan categorical menjadi numerical untuk mendapatkan insight dan hubungan yang lebih jelas.

2.4. Business Insight

Dalam menemukan business insight, hal yang perlu dilakukan yaitu menganalisis fitur-fitur dari dataset. Fitur-fitur yang digunakan dalam menganalisis dataset adalah Month, VisitorType, SpecialDay, dan Revenue. Berikut adalah beberapa hasil yang diperoleh dari analisis fitur-fitur tersebut.

1. Melihat pertumbuhan penjualan produk per bulan di tahun pertama.
2. Melihat banyaknya penjualan produk berdasarkan tipe visitor, yaitu new_visitor, returning_visitor, dan other.
3. Melihat banyaknya penjualan pada hari-hari istimewa.

Dari hasil analisis tersebut diperoleh beberapa business insight sebagai berikut.

1. Pertumbuhan penjualan produk di tahun pertama masih fluktuatif atau tidak stabil.
   - Pada bulan Mei - Jun, penjualan produk mengalami penurunan.
   - Pada bulan Jun - Nov, penjualan produk mengalami peningkatan.
   - Pada bulan Nov - Des, penjualan produk mengalami penurunan.
2. Sebagian besar penjualan produk per bulan masih berada di bawah rata-rata (rata-rata 191 transaksi per bulan).
3. Persentase returning visitor yang membeli produk sebesar 14% lebih rendah dibandingkan new visitor sebesar 25%.
4. Penjualan produk pada hari-hari istimewa masih rendah.

Berdasarkan business insight di atas, terdapat beberapa rekomendasi yang dapat dilakukan oleh perusahaan untuk meningkatkan penjualan, antara lain:

1. Untuk meningkatkan penjualan terhadap new visitor maupun returning visitor, Tim Business dan Tim Marketing dapat melakukan hal-hal berikut:
   - melakukan riset terhadap produk apa sajakah yang diinginkan oleh customer dengan menggunakan platform media sosial (ex: Instagram) atau tools riset (ex: Google Analytics).
   - membuat iklan sebagai media promosi perusahaan dengan menggunakan Google Ads atau mengkampanyekan iklan melalui media sosial.
   - menyusun strategi content marketing yang menarik perhatian customer dengan menambahkan blog ke website perusahaan.
2. Memberikan pelayanan terbaik kepada customer dan membuat program-program loyality customer secara periodik agar returning visitor banyak melakukan pembelian produk.
3. Memanfaatkan pemasaran produk di hari-hari istimewa, seperti Valentine's Day, Mother's Day, Christmas & New Year, dengan memberikan promo/diskon atau special offer kepada customer untuk meningkatkan penjualan di hari-hari istimewa.

3. Stage 2

Pada tahap Data Pre-Processing, hal yang perlu dilakukan adalah proses Data Cleansing dan Feature Engineering. Proses data cleansing dapat dilakukan dengan cara handle missing values, handle duplicated data, handle outliers, feature transformation, feature encoding, dan handle class imbalance. Adapun pada tahap feature engineering dapat dilakukan dengan mengecek Feature Selection dan Feature Extraction, serta menentukan Additional Features.

3.1. Data Cleansing

Dari proses data cleansing yang telah dilakukan dapat disimpulkan bahwa dataset yang digunakan tidak terdapat missing value sehingga tidak ada yang perlu dihilangkan atau diimputasi. Namun, dataset ini memiliki 125 rows yang duplikat sehingga perlu dihapus dari dataset. Kemudian kami mengecek apakah dataset tersebut terdapat outlier atau tidak dengan menggunakan IQR. Dari hasil pengecekan outliers, kami memutuskan untuk tidak menghapus outliers karena banyak data target yang bernilai 1 (Revenue True) yang hilang. Hal ini mengindikasikan bahwa outliers tersebut sebenarnya adalah pola yang wajar.

Langkah selanjutnya adalah melakukan feature encoding. Proses feature encoding ini bertujuan untuk mengubah fitur-fitur kategorik menjadi angka (numerik) agar bisa dimengerti oleh model. Fitur-fitur yang diencoding adalah Month, VisitorType, OperatingSystems, Browser, Region, Weekend, Revenue. Setelah itu, kami melakukan feature transformation dengan menggunakan Transformasi Yoe Johnson karena semua fitur numerik mengandung nilai 0 dan Transformasi Scalling. Dari hasil transformasi menyebabkan distribusi beberapa fitur menjadi lebih normal. Langkah terakhir yang dilakukan adalah mengatasi imbalance pada dataset. Dari hasil pengecekan, kami mendapatkan imbalance lebih dari 15% yang masuk ke dalam kategori moderate sehingga kami melakukan over sampling menggunakan SMOTE sebanyak 50% agar tidak terjadi overfitting pada data train.

3.2. Feature Engineering

Tahap feature engineering terdiri dari Feature Selection, Feature Extraction, dan mencari Additional Features.

1. Feature Selection

   Tahap Feature Selection bertujuan untuk menghapus fitur-fitur yang kurang relevan atau redundan. Pada tahap ini kami menghapus beberapa fitur yang merupakan fitur berkorelasi tinggi dengan fitur lainnya (redundan) dan juga fitur-fitur yang tidak variatif (fitur kategorik yang hanya didominasi satu value saja). Berikut adalah fitur-fitur yang kami hilangkan.

   - Fitur redundan: ProductRelated_Duration, BounceRates
   - Fitur tidak variatif: VisitorType

2. Feature Extraction

Untuk sementara tidak ada fitur yang diextract dari fitur-fitur yang sudah ada. Jika pada tahap modeling menemukan performa model yang kurang baik, bisa dipertimbangkan untuk extract fitur baru seperti Average Duration Per Page.

3. Additional Features

Beberapa fitur yang mungkin dapat membantu meningkatkan performa model selain fitur-fitur yang sudah ada adalah sebagai berikut.

a. Promotion Day (apakah sesi dilakukan pada saat hari-hari promosi seperti tanggal cantik atau yang lainnya)
b. Internet Speed (ini berguna untuk menentukan apakah perlu menyesuaikan tampilan website untuk visitor dengan internet low speed, karena laman yang lama dimuat cenderung membuat visitor meninggalkan laman dengan cepat)
c. ProductReview (berapa sering dan berapa lama visitor melihat laman review product)
d. DeviceType (jenis device yang digunakan)

4. Stage 3

Stage 3 final project ini terdiri dari dua langkah utama yaitu modeling dan menggali important features, insight, serta action items berdasarkan model yang dipilih (final model).

4.1. Modeling

Pada tahap modeling dicoba implementasi terhadap beberapa algoritma klasifikasi yang mungkin untuk digunakan pada dataset yang telah di-preprocessing pada stage sebelumnya. Metrics yang digunakan sebagai evaluasi model adalah ROC AUC dengan tetap mempertimbangkan skor precision dan recall. Terdapat total 8 model yang telah dicoba dengan hasil sebagai berikut.

1. Gaussian Naïve Bayes

Model supervised learning Gaussian Naïve Bayes dapat digunakan untuk problem klasifikasi. Pemodelan dalam final project ini dilakukan menggunakan hyperparameter default yaitu var_smoothing sebesar 10-9 dan tidak ada (none) nilai priors. Setelah model di training, kemudian model dievaluasi menggunakan test set dan menghasilkan skor AUC sebesar 86%. Nilai AUC hasil prediksi pada train set juga tidak terlalu jauh dari test set yaitu sebesar 87%. Hasil evaluasi menggunakan cross validation juga tidak berbeda signifikan antara train dan test set (87% untuk test set dan 88% untuk train set).

Perbedaan yang tidak signifikan antara AUC di test set dengan train set menunjukkan bahwa model Gaussian Naïve Bayes yang dilatih tidak overfitting sehingga tidak perlu dilakukan hyperparameter tuning.

2. Decision Tree

Decision Tree merupakan salah satu metode supervised learning classification yang dapat digunakan untuk data yang memiliki hubungan non-linear. Dengan menggunakan model tersebut pada dataset, hasil evaluasi yang diperoleh menunjukkan bahwa model tersebut mengalami overfitting dengan skor AUC train set sebesar 100% dan test set sebesar 69%.

Untuk menghindari overfitting, langkah selanjutnya adalah melakukan hyperparameter tuning. Setelah dilakukan hyperparameter tuning, hasil evaluasi yang diperoleh lebih baik dari sebelumnya dengan skor AUC train set sebesar 96% dan test set sebesar 88%. Dari hasil hyperparameter tuning, dapat diketahui juga bahwa best parameter untuk max_depth = 110, min_samples_leaf = 4, min_samples_split = 100, max_features = None, criterion = ‘entropy’, dan splitter = ‘best’.

3. Random Forest

Random Forest merupakan salah satu algoritma supervised learning yang dapat digunakan untuk klasifikasi yang dimana ini bisa dipakai pada data final project kali ini. Setelah dilakukan penerapan terhadap model tersebut ternyata hasil evaluasi AUC train 100% dan AUC test 89%.

Dari hasil tersebut mengindikasikan adanya overfitting sehingga perlu dilakukan hyperparameter tuning pada dataset. Setelah dilakukan hyperparameter tuning ada peningkatan performa menjadi AUC train 96% dan AUC test 91%. Dari hasil tersebut diperoleh bahwa best parameter untuk max_depth = 89, min_samples_leaf = 2, criterion = 'entropy’, min_samples_split =150 dan n_estimators = 104.

4. XG Boost

Algoritma XGBoost merupakan salah satu algoritma yang paling populer dan paling banyak digunakan karena algoritma ini termasuk algoritma yang powerful. Algoritma ini dapat digunakan pada kasus final project kali ini dan setelah diterapkan terhadap model hasil evaluasi cukup bagus yaitu mendapatkan Score AUC Cross Validation 95%(Train) & 94%(Test).

Dari hasil tersebut sudah menunjukkan hasil yang memuaskan, tetapi ketika dilakukan hyperparameter tuning Score model justru menurun. Jadi, model tanpa hyperparameter tuning untuk XG boost pada kasus final project kali ini lebih baik.

5. K-Nearest Neighbors

K-Nearest Neighbors dapat didefinisikan sebagai salah satu algoritma model yang digunakan pada metode classification pada supervised learning. Setelah algoritma K-Nearest Neighbors ini diterapkan pada final project kali ini, diperoleh metrics AUC Cross Validation 97% pada data train dan 84% pada data test. Namun selisih nilai train-test pada precision, recall, dan f1 yang cenderung lebar mengindikasikan bahwa ada overfitting.

Dengan terjadinya overfitting, perlu dilakukan hyperparameter tuning. Setelah dilakukan hyperparameter tuning, diperoleh selisih nilai train-test pada metrics yang lebih kecil. Hal ini menunjukkan bahwa performa model membaik setelah dilakukannya hyperparameter tuning. Diperoleh juga bahwa hyperparameter terbaiknya yakni n_neighbors (jumlah K yang optimal) berjumlah 29.

6. Ada Boost

AdaBoost merupakan salah satu algoritma dalam teknik ensemble method yang bisa digunakan untuk mencegah underfitting. Setelah algoritma ini diterapkan pada dataset, hasil evaluasi yang didapatkan cukup bagus, yaitu AUC Train sebesar 93% dan AUC Test sebesar 88%. Sedangkan skor AUC untuk Cross Validation sebesar 93% untuk train dan 92% untuk test set.

Tidak ada perbedaan yang signifikan antara AUC train set dan test set menjadi indikasi bahwa model tidak mengalami overfitting, sehingga tidak diperlukan hyperparameter tuning.

7. Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi seperti Support Vector Classification (SVC).

Pada final project kali ini dapat diterapkan algoritma Support Vector Classification (SVC) dan diperoleh metrics AUC Cross Validation (train) 94% dan AUC Cross Validation (test) 92%.

Dari hasil tersebut perbedaannya tidak signifikan antara AUC Cross Validation(train) dan AUC Cross Validation(test) yang menunjukkan tidak overfitting sehingga tidak perlu dilakukan hyperparameter tuning.

8. Logistic Regression

Regresi logistik adalah teknik analisis data yang menggunakan matematika untuk menemukan hubungan antara dua faktor data. Kemudian menggunakan hubungan ini untuk memprediksi nilai dari salah satu faktor tersebut berdasarkan faktor yang lain. Prediksi biasanya memiliki jumlah hasil yang terbatas, seperti ya atau tidak.

Pada final project ini, diperoleh Metrics AUC train sebesar 0.89 dan AUC test Sebesar 0.86. Sedangkan pada Cross Validation AUC ditemukan CV AUC train dan CV AUC test sebesar 0.89. Dari hasil di atas, perbedaannya tidak terlalu signifikan sehingga tidak menunjukan overfitting. Oleh karena itu, tidak perlu dilakukan hyperparameter tuning

4.2. Final Model

Setelah mencoba beberapa kandidat model, akhirnya terpilih model XG Boost. Model ini dipilih karena memiliki skor AUC tertinggi (91% untuk test set) dan juga memiliki skor precision recall yang baik. Model XG Boost yang dilatih menggunakan hyperparameter default tanpa dilakukan tuning karena tidak overfitting.

1. Important Feature

Berdasarkan model XG Boost yang dipilih, didapatkan 4 fitur terpenting yaitu PageValues, ExitRates, Administrative, dan ProductRelated.

2. Business Insight

Dari fitur-fitur penting di atas, didapatkan beberapa business insight sebagai berikut.

a. Jika nilai Page Values semakin tinggi, maka kemungkinan pengunjung untuk purchase juga semakin tinggi.
b. Semakin tinggi exit rate, maka kemungkinan untuk purchase semakin rendah.
c. Semakin sedikit page Administrative yang dikunjungi, maka semakin tinggi kemungkinan pengunjung untuk purchase.
d. Semakin banyak page Product Related yang dikunjungi, maka kemungkinan pengunjung untuk purchase semakin rendah.

3. Action Items

Dari insight yang kami temukan, dirumuskan beberapa contoh action items yang mungkin dapat membantu meningkatkan bisnis kedepannya. Beberapa action items yang kami rumuskan adalah sebagai berikut.

a. Untuk sesi dengan page value tinggi namun tidak purchase (abandoned cart), bisa diberikan email reminder untuk melanjutkan transaksi.
b. Lakukan optimasi file gambar dan visual lainnya untuk mempercepat page load time (mengurangi exit rates).
c. Minimalisir transaction error issue dan permudah proses transaksi agar visitor tidak terlalu sering membuka page administrative.
d. Optimasi sistem product recommendation agar sesuai dengan tiap-tiap visitor.

5. Stage 4

Pada stage 4, hal yang dilakukan adalah menyusun materi presentasi berdasarkan langkah-langkah pengerjaan pada stage 0 sampai stage 3. Pada materi presentasi tersebut, kami akan menjelaskan background yang kami usung dalam penyelesaian permasalahan ini, seperti problem statement, goals, dan objectives. Selain itu, kami memberikan insight yang diperoleh dari proses EDA dan memberikan penjelasan dari pre-processing yang telah kami lakukan dalam mengatasi missing value, duplicated rows, outliers, dan lain-lain.

Pada materi presentasi, kami juga akan menjelaskan alasan pemilihan XG Boost sebagai model yang paling baik dibandingkan model yang lain. Dari model tersebut, kami akan melakukan simulasi model untuk melihat conversion rate yang diperoleh sebelum dan sesudah menggunakan model. Kemudian kami memberikan beberapa business recommendation untuk mengatasi permasalahan yang ada agar goals yang diinginkan dapat tercapai.