**SCRIPT FINAL PRESENTATION**

1. **Opening**Selamat malam kakak2 mentor, selamat malam kakak2 fasilitator, selamat malam rekan – rekan semua. Terimakasih atas kesempatannya untuk kami bisa mempresentasikan hasil dari final project kami. Sebelum memulai izinkan kami memperkenalkan tim kami dahulu, yang pertama ada kak syifa, kedua ada saya Stevanus Ertito sebagai speaker, ketiga ada kak Vivian, keempat ada kak yudistira, kelima ada kak Sriningsih, dan terakhir ada kak Tania. Oke untuk mempersingkat waktu langsung saja ke next slide
2. **PROBLEM STATEMENT**
3. Customer Churn  
   Customer Churn adalah proses untuk mengukur tingkat pelanggan untuk melihat customer berhenti membeli produk ataupun customer yang masih beli produk.
4. Objective  
   Memprediksi customer Churn dalam 1 bulan kedepan dengan periode 6 bulan sebelumnya
5. Data Element  
   Dataset Fashion Campus  
   - Customer  
   - Product   
   - Click-Stream  
   - Transactions
6. Data Period  
   - Calibration Period : Desember 2021 – Mei 2022 (untuk bulan Juni)  
   - Holdout Period : Januari 2022 – Juni 2022 (untuk bulan Juli)
7. Labeling  
   (1) Customer yang tidak transaksi 1 bulan terakhir   
   (0) Customer yang transaksi 1 bulan terakhir
8. Relation Table  
   Pada relasi tabel ini untuk mengeksplore datanya bisa dilihat bahwa dari tabel customer dan tabel transaksi memerlukan ‘customer\_id’, transaksi ke click\_stream memerlukan ‘session\_id’, dan click\_stream ke products memerlukan ‘product\_id’
9. **FLOWCHART PROJECT**

Data preprocessing -> EDA -> Feature Engineering -> Feature Selection -> Baseline Model -> Hypertuning Parameters -> Model Evaluation

1. **EDA**

Pada pie chart customer churn & non-churn diketahui bahwa label “0” adalah customer non-churn dan label “1” adalah customer churn dimana untuk customer churn sebesar 34.8% sedangkan customer non-churn sebesar 65.2%. Kemudian adapun diagram batang untuk metode pembayaran yang berhasil atau tidak pada customer dimana di dominasi oleh metode pembayaran Credit Card. Dan yang terakhir adalah presentase device yang dipakai customer android sebesar 76.6% dan iOs sebesar 23.4%

1. **Feature Engineering & Labeling**

Terdapat 14 Feature yang telah didapatkan yaitu customer\_id, most\_recent itu pembelian terakhir dari customer, payment\_method adalah metode pembayaran, payment\_status adalah status pembayaran, promo\_amount adalah jumlah total promo, dan ada fitur RFM yaitu recency usia akun pengguna, frequency itu jumlah frekuensi transaksi, dan monetary adalah jumlah pembayaran yang dilakukan oleh customer, kemudian ada quantity jumlah produk yang dibeli, transactions\_amount adalah jumlah transaksi, is\_churn adalah churn atau tidaknya dari customer, age adalah umur customer, promo\_amount\_used adalah jumlah promo yang dipakai, terakhir adalah total\_promo\_paid adalah total promo yang dibayarkan.  
Kemudian ada pula label churn customer dari rentang 6 bulan yaitu calibration period Desember 2021 – Mei 2022 dan holdout period dari Januari 2022 – Mei 2022.

1. **Feature Selection**

Pada seleksi fitur ini kami menggunakan information value dengan rumus sebagai berikut dan ada juga rules tabel untuk information value yaitu kurang dari 0.02 kurang berguna, 0.02 – 0.1 Weak, 0.1- 0.3 Medium, 0.3 – 05 Strong, >0.5 Mencurigakan. Dengan menggunakan acuan di atas didapatkan hasil tiap fitur seperti di samping dimana untuk customer\_id dan payment\_status tidak penting dan recency lebih dari 0.5 mencurigakan dan feature lainnya bisa digunakan.

1. **Baseline Model**

Pada modelling kita menggunakan 3 algoritma yaitu logistic regression, decision tree, dan random forest dengan algoritma logistic regression akurasi sebesar 76%, decision tree sebesar 94% dan Random forest sebesar 88% dengan F1 – Score sekian.  
(Menampilkan ROC & Confusion Matrix)

1. **Hypertuning Parameters**

Kami menggunakan metode grid search dalam melakukan hypertuning, lalu didapatkan skor AUC dan score baru seperti berikut :  
Logistic Regression auc score sebesar 92% dan akurasi nya menurun menjadi 65%  
Decision Tree auc score sebesar 95% dan akurasi meningkat menjadi 96%  
Random Forest auc score sebesar 95% dan akurasi meningkat menjadi 96%

1. **Model Evaluation**

Metrics seperti accuracy, precision, recall adalah cara yang baik untuk mengevaluasi model klasifikasi untuk kumpulan data seimbang, tetapi jika data tidak seimbang dan terdapat perbedaan kelas, maka metode lain seperti ROC/AUC, koefisien Gini akan bekerja lebih baik dalam mengevaluasi performa model.

Predictive Score Logistic Regression menurun sebesar 11% dimana dari 76% menjadi 65%, sedangkan Decision Tree naik sebesar 2% dari 94% menjadi 96%, dan Random forest sebesar 10% dari 86% menjadi 96%.

1. **Feature Importance**Mengacu pada teknik feature selection yang telah dilakukan untuk mewakili "kepentingan" setiap fitur. Jadi disini kita memilih information valuenya diantara 0.1 dan 0.5 dimana ada 3 fitur penting yaitu most\_recent yaitu 0.34, promo\_amount 0.1 dan frequency 0.35. Dan untuk fitur lebih dari 0.5 akan kita drop karena akan membuat skor model mencurigakan dan jika kurang dari 0.1 kita akan gunakan saja (tidak terlalu berpengaruh)

* **Recommendation Bussiness**Fashion Campus dapat menerapkan beberapa model Machine Learning untuk memprediksi churn dari customer-customer mereka.
* Berdasarkan data-data yang ada sejauh ini, model yang memberi hasil terbaik adalah model **Random Forest**  
  Adapun rekomendasi bisnis kami terhadap perusahaan dalam mengurangi churn yaitu:
* **Royalty Programs**Nilai **frequency** yang tinggi, cenderung mengurangi probabilitas *churn,* sehingga Fashion Campus dapat menerapkan berbagai Royalty Program (promo, poin, member) untuk mengantisipasi *churn*
* **Push Notification**Angka recency (usia akun pengguna) yang tinggi, cenderung meningkatkan churn, sehingga Fashion Campus dapat fokus untuk mempertahankan customer-customer paling baru pada platform melalui push notifications.
* **Further Insight Gathering**Untuk menindaklanjuti churn prediction yang telah dihasilkan, Fashion Campus dapat melakukan eksplorasi lanjutan untuk eksplorasi insight baru seperti:  
  1) Customer Lifetime Value Prediction: kita tidak bisa meminta semua customer kita tetap bertahan pada perusahaan, namun melalui ini, Fashion Campus dapat memprediksi customer-customer yang paling patut dipertahankan karena membawa nilai tinggi kepada perusahaan  
  2) Advertising Segmentation: Fashion Campus dapat menggunakan segmentasi pasar untuk membuat ads dan bentuk-bentuk pemasaran yang tepat sasaran