

Marketing Campaign

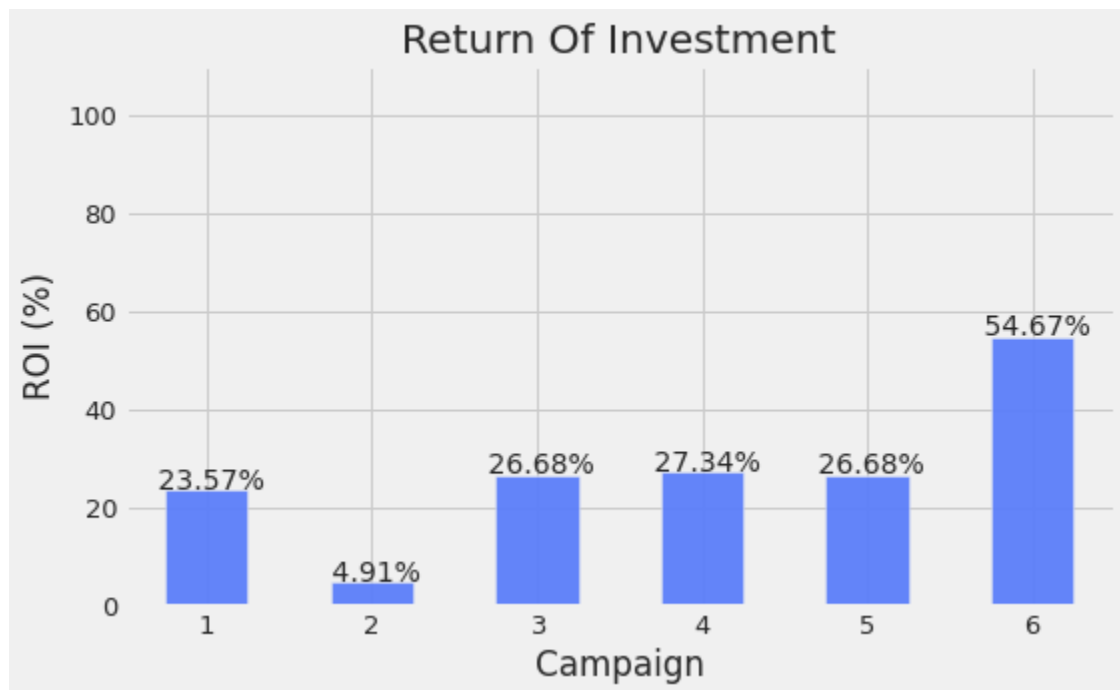
Team The Wizards:

1. Dwi Abriansya Alimuddin
2. Ni Kadek Rina Wati
3. Ivander Edo
4. Naufal Fajar Revanda
5. Almaira Nabila

1. Problem Statement

- Problem

- Company mengalami kerugian pada Campaign terakhir karena banyak customer yang tidak menerima Campaign.
- Setelah ditelusuri, kerugian ini sudah terjadi dari 6 Campaign sebelumnya.



Dari 6 Campaign sebelumnya, tidak ada yang mencapai ROI 100% yang berarti semuanya mengalami kerugian.

- Team As

Kami berperan sebagai data scientist yang akan membantu tim marketing dalam memprediksi customer yang akan menerima Campaign selanjutnya.

- **Goals**

Meningkatkan Return Of Investment (ROI) tanpa harus meningkatkan cost.

- **Objective**

Melatih model yang dapat membantu company memprediksi customer yang akan menerima Campaign selanjutnya.

- **Business Metrics**

- Return Of Investment (ROI)

- **Available Data**

Data yang diberikan kepada kami memiliki beragam informasi, mulai dari background customer, interaksi customer dengan store, sampai respon customer terhadap beberapa Campaign sebelumnya. Dari data-data tersebut kami akan memprediksi customer seperti apa yang akan menerima Campaign selanjutnya. Rinciannya sebagai berikut:

- AcceptedCmp1 - 1 if customer accepted the offer in the 1st campaign, 0 otherwise
- AcceptedCmp2 - 1 if customer accepted the offer in the 2nd campaign, 0 otherwise
- AcceptedCmp3 - 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
- AcceptedCmp4 - 1 if customer accepted the offer in the 4th campaign, 0 otherwise
- AcceptedCmp5 - 1 if customer accepted the offer in the 5th campaign, 0 otherwise
- Response (target) - 1 if customer accepted the offer in the last campaign, 0 otherwise
- Complain - 1 if customer complained in the last 2 years
- DtCustomer - date of customer's enrolment with the company
- Education - customer's level of education
- Marital - customer's marital status
- Kidhome - number of small children in customer's household
- Teenhome - number of teenagers in customer's household
- Income - customer's yearly household income
- MntFishProducts - amount spent on fish products in the last 2 years
- MntMeatProducts - amount spent on meat products in the last 2 years
- MntFruits - amount spent on fruits products in the last 2 years
- MntSweetProducts - amount spent on sweet products in the last 2 years
- MntWines - amount spent on wine products in the last 2 years
- MntGoldProds - amount spent on gold products in the last 2 years
- NumDealsPurchases - number of purchases made with discount
- NumCatalogPurchases - number of purchases made using catalogue
- NumStorePurchases - number of purchases made directly in stores
- NumWebPurchases - number of purchases made through company's web site

- NumWebVisitsMonth - number of visits to company's web site in the last month
- Recency - number of days since the last purchase
- Z_CostContact - Cost Campaign
- Z_Revenue – Revenue Campaign

2. Pre-processing

- Data Cleansing

- Dataset memiliki data **NULL** sebanyak **24 data** pada features Income, data tersebut didrop

```
ID          0
Education   0
Marital_Status 0
Income      24
Recency     0
NumWebVisitsMonth 0
Complain    0
Response    0
Year_Enroll 0
Long_Enroll 0
Has_Child   0
Children    0
Age         0
Spending    0
Transactions 0
Acc_Cmp     0
dtype: int64
```

- Dataset tidak memiliki duplikasi
- Dataset memiliki **outlier** pada beberapa features dengan total outlier sebanyak **710 data**, agar tidak mempengaruhi pemodelan outlier akan didrop.

```
Jumlah baris sebelum memfilter outlier: 2216
Jumlah baris setelah memfilter outlier: 1506
```

Jumlah Outlier sebelum dan setelah difilter.

- Feature Engineering

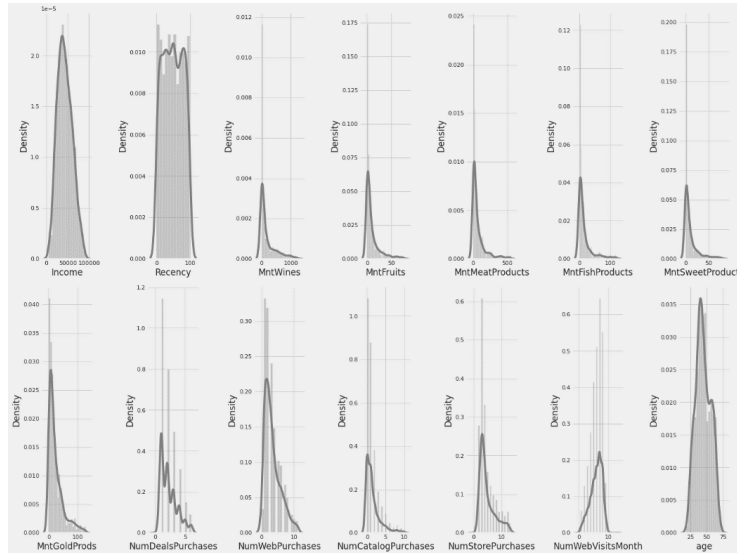
Feature Engineering dilakukan pada beberapa feature agar performa pemodelan menjadi lebih baik. Rincian sebagai berikut:

- Feature **Z_CostContact** dan **Z_Revenue** didrop.
- Feature **Marital_Status** dilakukan pengelompokan ulang menjadi 2 value saja, **In a Relationship** dan **Single**.
- Feature **Yr_Birth** kami ubah menjadi **Age**.
- Features **Children** merupakan hasil penjumlahan dari feature **Kidhome** dan **Teenhome**.
- **Dt_Customer** dikonversikan menjadi Feature **Year_Enroll**.
- **Acc_Cmp** merupakan feature yang menunjukkan seberapa sering customer menerima Campaign.

3. EDA & Insights

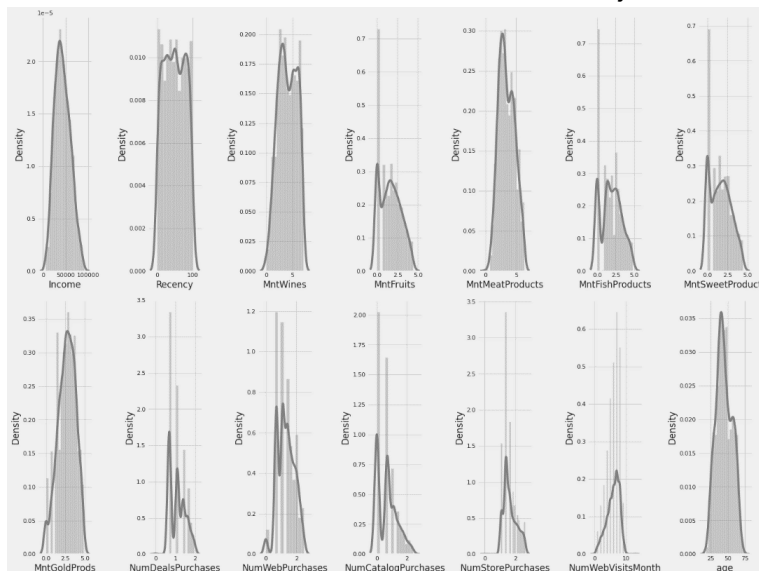
- EDA

- Distribusi features **MntWines**, **MntFruits**, **MntMeatProducts**, **MntFishProducts**, **MntSweetProducts**, **MntGoldProds**, **NumDealsPurchases**, **NumWebPurchases**, **NumCatalogPurchases**, **NumStorePurchases** positively skew sehingga butuh ditransformasi menggunakan Log Transform.



Distribusi Features sebelum dilakukan Transformasi

- Setelah dilakukan transformasi, distribusi menjadi lebih normal.

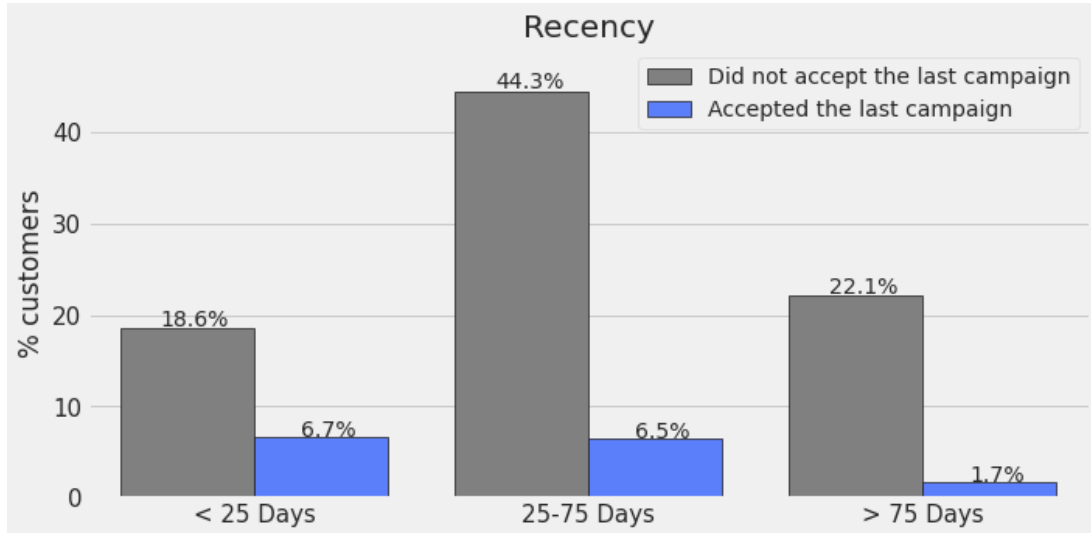


Distribusi Features setelah dilakukan Transformasi

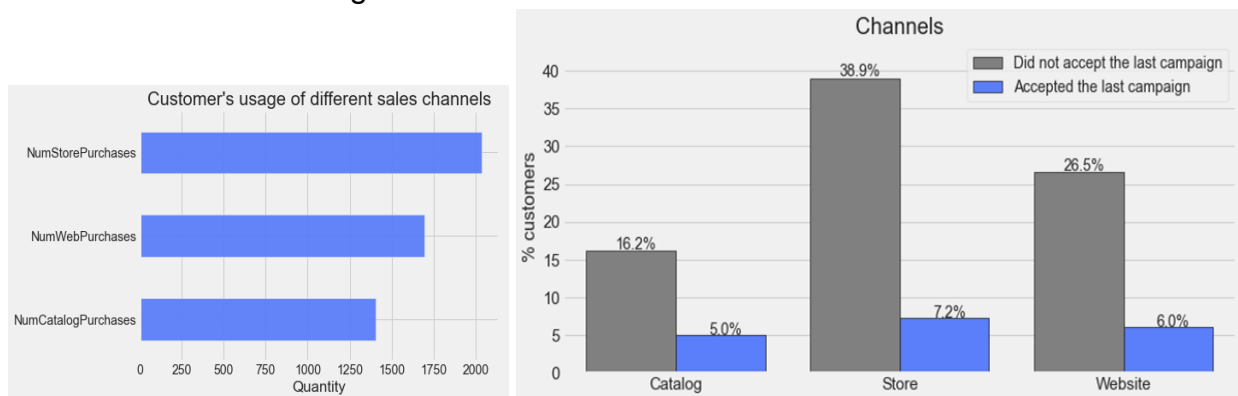
- Setelah ditransformasi, dataset kemudian distandardisasi agar seragam

- Insight

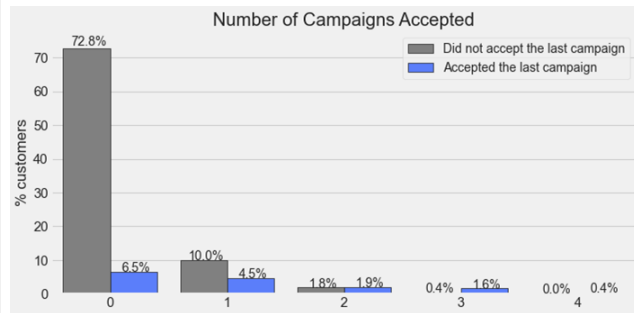
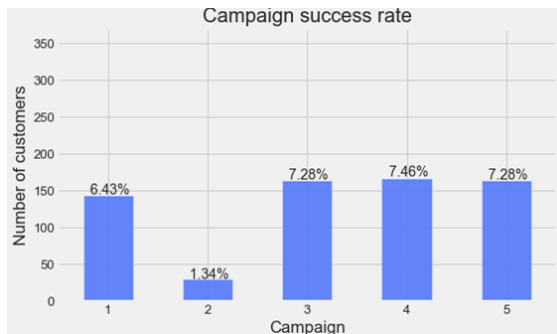
- Persentase tertinggi customer yang menerima Campaign terakhir berasal dari pelanggan yang melakukan pembelian terakhir mereka tidak lebih dari 75 hari yang lalu.



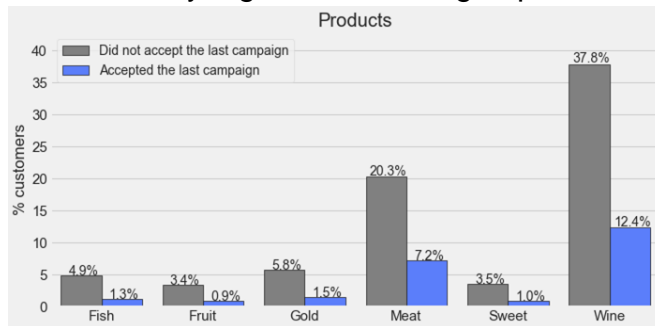
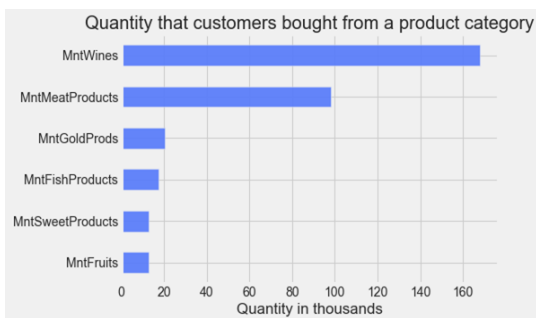
- Persentase pembelian customer yang menerima Campaign terakhir tidak terlalu berbeda di ketiga channel penjualan. Namun, sebagian besar pembelian dilakukan melalui store (> 2000 pembelian), diikuti oleh website dan katalog.



- Tiga Campaign terakhir adalah Campaign paling “berhasil” dari kelima Campaign yang diberikan kepada Customer. Semakin sering Campaign diterima oleh Customer, semakin besar kemungkinan mereka akan menerima Campaign berikutnya.



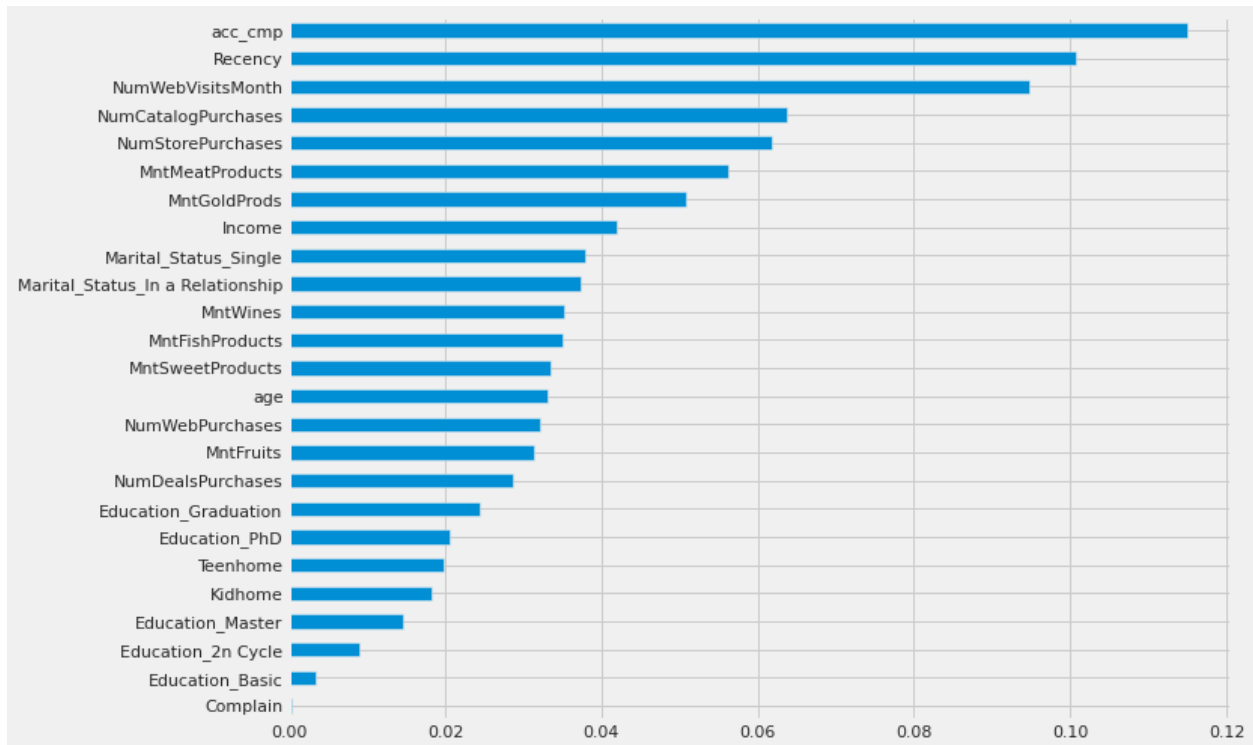
- Produk yang paling banyak dibeli oleh customer adalah wine dan meat. Customer yang membeli kedua produk ini juga cenderung menerima Campaign terakhir daripada customer yang membeli kategori produk lain.



4. Modelling Experiment

- Modelling

- Pemodelan yang akan dilakukan adalah pemodelan supervised learning classification untuk memprediksi customer yang akan menerima Campaign berikutnya.
- Pemodelan menggunakan data train dan test dengan ratio 70:30, dengan target adalah Response
- Dataset Imbalance sehingga dilakukan Oversampling SMOTE pada data train
- Algoritma yang digunakan adalah Logistic Regression, KNN, Decision Tree, Random Forest, XGBoost, dan Adaboost
- Metrics Evaluation yang digunakan adalah Precision dan F0.5_score. Precision akan menunjukkan customer yang benar-benar akan menerima Campaign. F0.5_score berguna untuk memberikan weight pada Precision dan Recall. Weight yang digunakan memiliki rasio Precision 2 : Recall 1, Hal ini dilakukan agar target market pada Campaign selanjutnya tidak terlalu kecil.
- Algoritma Random Forest memiliki hasil pemodelan terbaik.
- Dari Algoritma Random Forest diperoleh Feature Importance, kita akan mengeliminasi feature yang kurang mempengaruhi model.



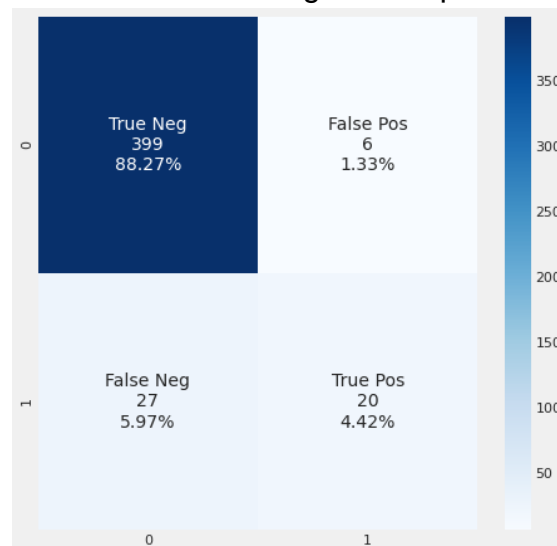
Feature Importance dari Algoritma Random Forest. Feature Complain akan dieliminasi lalu dilakukan fitting dan predict ulang.

- Setelah eliminasi feature, hasil pemodelan Random Forest adalah:

```

↳ Accuracy (Test Set): 0.93
Precision (Test Set): 0.77
Recall (Test Set): 0.43
F1-Score (Test Set): 0.55
FBeta-Score (Test Set): 0.66
AUC: 0.71
  
```

- Confusion Matrix data test dengan hasil prediksi adalah:



5. Conclusion & Recommendation

- Conclusion

- Model yang dibuat telah berhasil memprediksi customer yang akan menerima Campaign selanjutnya dengan Hit-Rate sebesar 77%, ini merupakan peningkatan cukup tinggi dari Campaign sebelumnya yang hanya bernilai 14.9% saja.
- Namun dengan Hit-Rate 77% ini memiliki Capture-Rate yang cukup rendah hanya 43%, hal ini berarti Campaign selanjutnya hanya memiliki sedikit target.

- Recommendation

- Untuk menghasilkan model yang lebih baik, dibutuhkan data yang lebih banyak dan tidak imbalance.
- Untuk Campaign selanjutnya, dapat membuat loyalty card agar mendorong customer dalam melakukan transaksi.
- Untuk Campaign selanjutnya dapat fokus pada product favorit yaitu Wines dan Meats.