Fare Sunday

Data Science Batch 13

Tour Prediction Packages and Opportunities

Who Are We?

Khalifa
Data Engineering

Rey
Product Analyst

Nadiah
Data Analyst

Susan
Business Analyst

Dia
Data Scientist

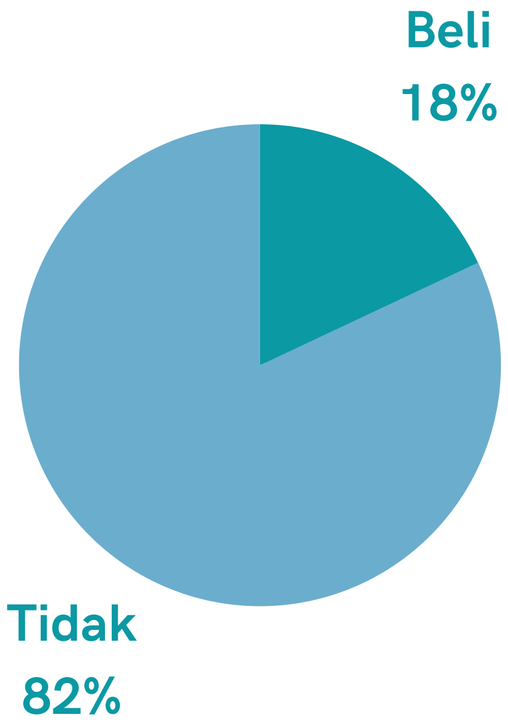
Table of Contents

Presentation Outline

Background and Problem
Insights
Contoh Kasus
Goals
Exploratory Data Analysis
Pre-processing
Data Visualisation
Model Training
Feature Importance
Business Recommendation

Background & Problem

- Perusahaan "Trips & Travel.Com" menjual produk paket tour.
- 5 jenis paket yang ditawarkan perusahaan - Basic, Standard, Deluxe, Super Deluxe, King.
- Wellness Package Release in 2021



Rendahnya daya beli disebakan oleh random marketing, menyebabkan biaya marketing yang terbuang tinggi

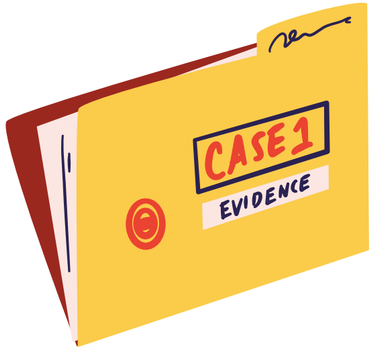
Insights

Rata-rata harga paket tour :
$2000-$4000
Total data : 4888

- 82% dari 4888 adalah 3968
- 3968 x $2000 = $7.936.000

Perusahaan kehilangan 8 juta dollar pada 2020

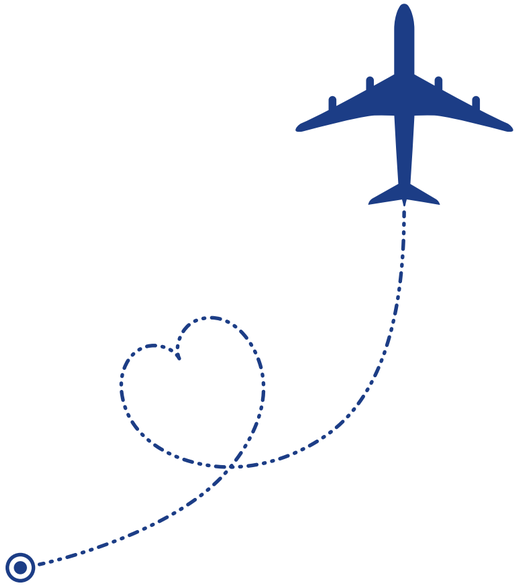
Contoh Kasus



Banyak perusahaan travel yang mengalami penurunan demand di tahun 2020

Cara apa yang mereka lakukan untuk bangkit?

Segmentasi pasar yang jelas contohnya, Traveloka mempromosi virtual tour yang dipusatkan pada segment market tertentu .



Goals

Mengidentifikasi feature yang mengaffect buying decision.

Mengidentifikasi target market paling tepat sasaran dan untuk menaikan customer buying rate menggunakan sistem rekomendasi

RAKAMIN ACADEMY

Business METRIC

- Purchase Rate: Semakin tinggi purchased rate, semakin besar juga basket size.
- Customer Satisfaction Rate: semakin tinggi rate, berarti penjualan lebih tepat sasaran

Data Preparation

CustomerID ProductTaken Age TypeofContact CityTier

DurationOfPitch Occupation Gender Designation Passport

ProductOfPitch MonthlyIncome PreferredPropertyStar MaritalStatus

OwnCar ProductOfPitch NumberOfTrips NumberOfPersonVisiting

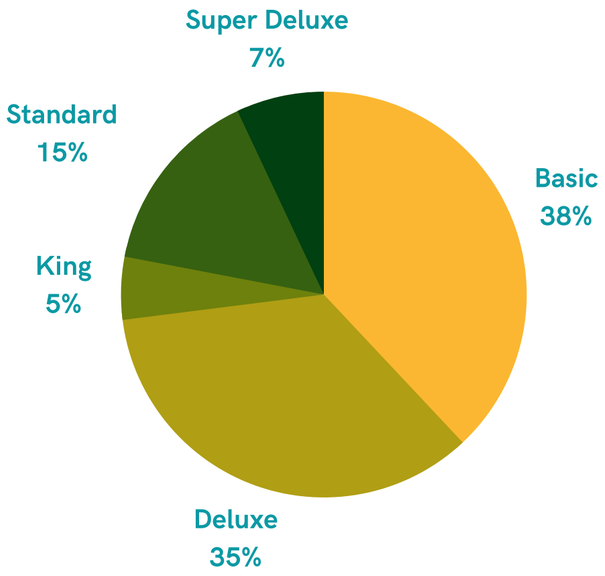
NumberOfFollowups NumberOfChildren Chance

Exploratory Data Analysis

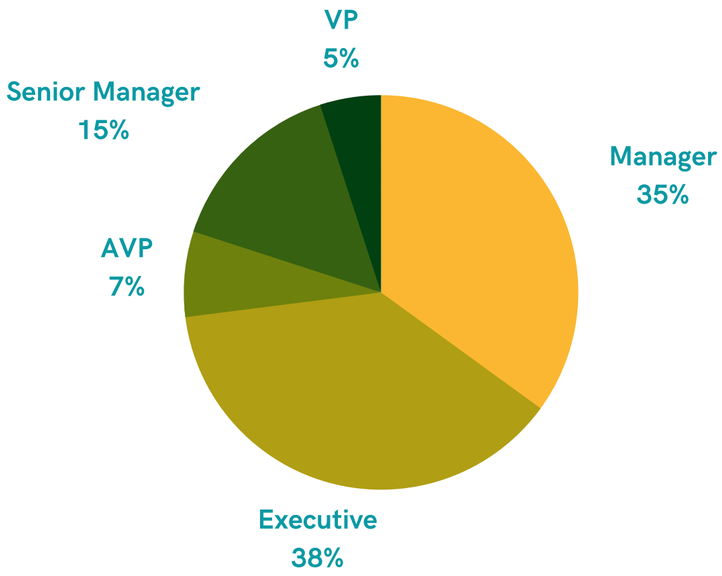
Categorical = TypeofContact, Occupation, Gender, MaritalStatus, Designation

Numerical = CustomerID, ProdTaken, Age, CityTier, DurationPitched, NumberOfPersonVisiting, NumberOfFollowups, NumberOfTrips, Passport, PitchSatisfactionScore, OwnCar, NumberOfChildrenVisiting, MonthlyIncome

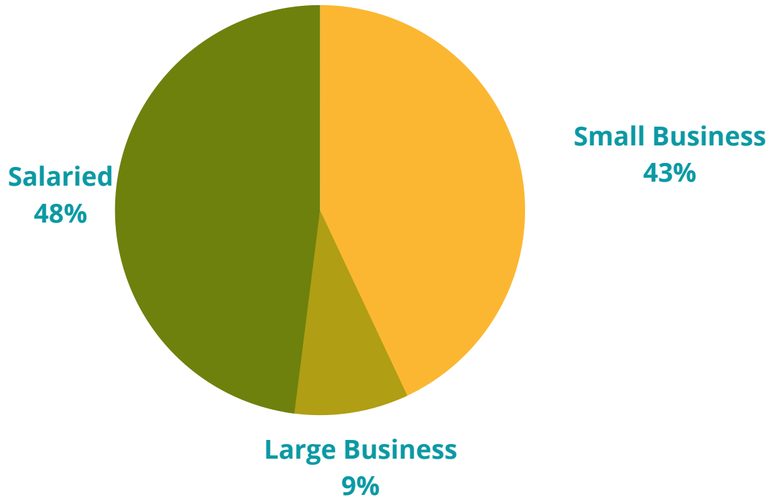
ProductPitch

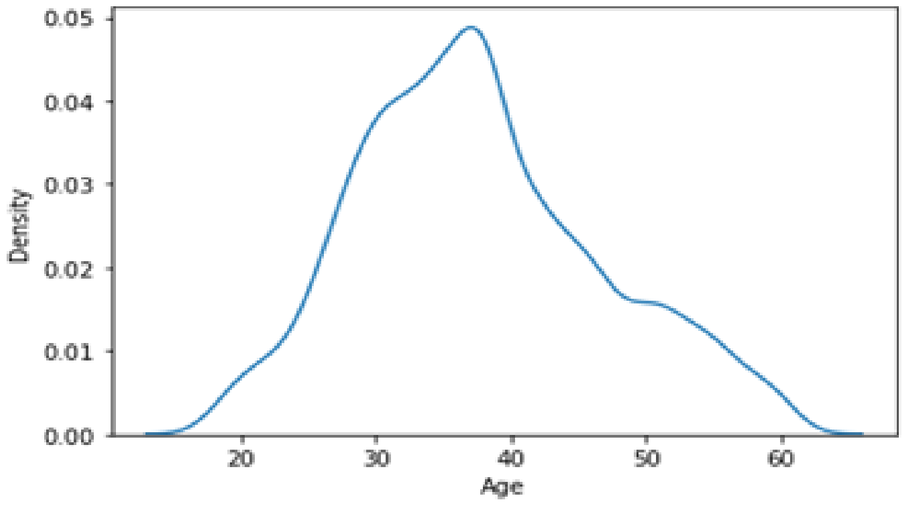


Designation

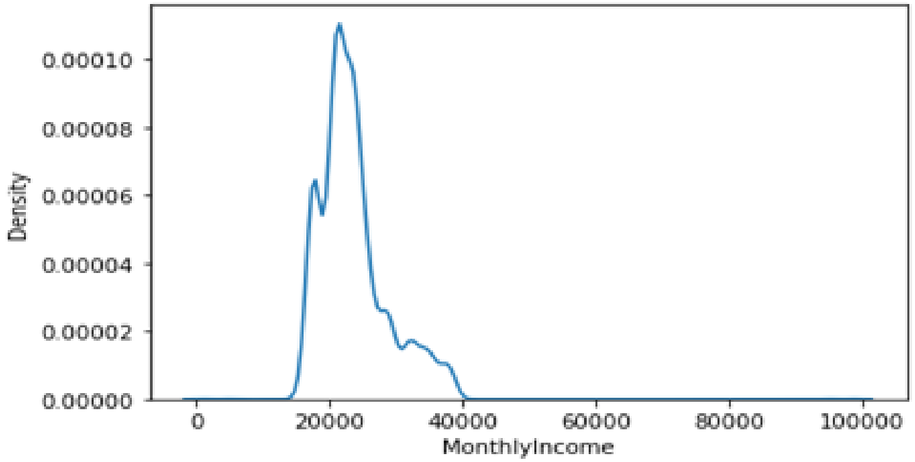


Occupation



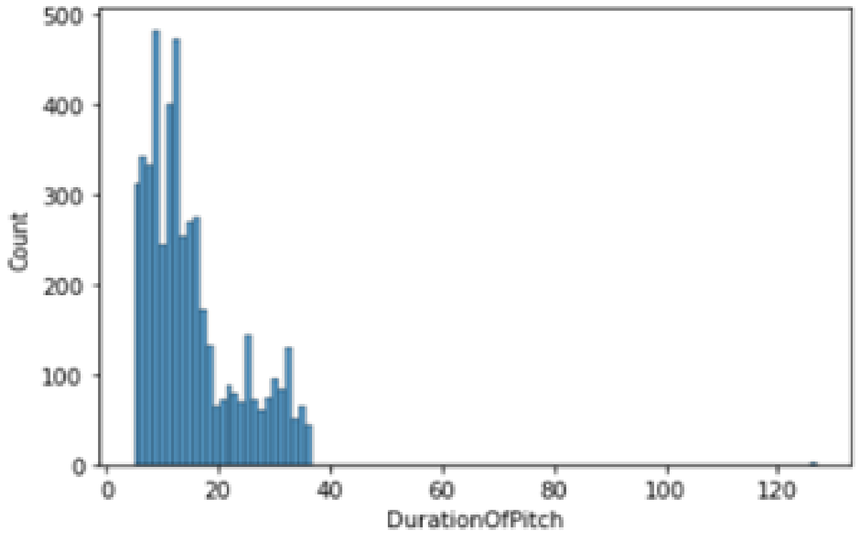


Rata-rata umur di 35 tahun

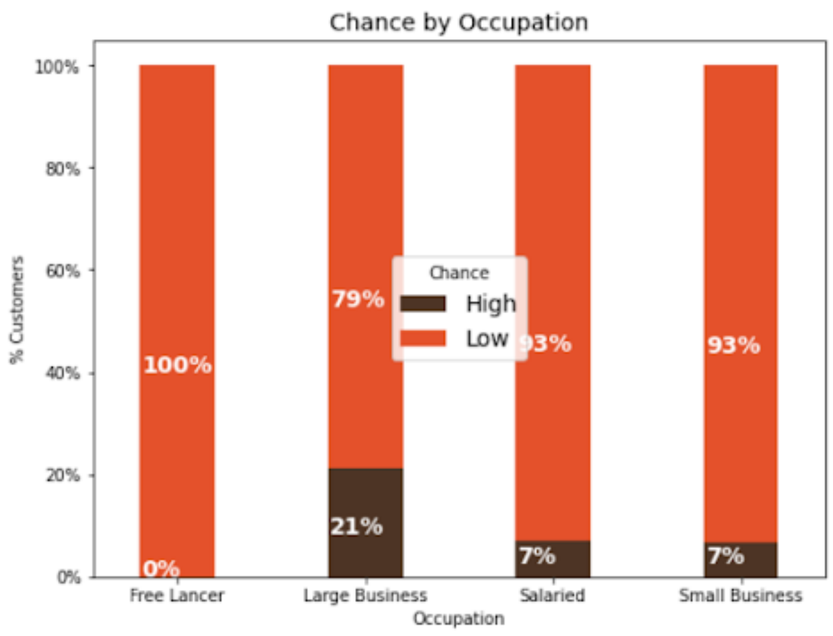


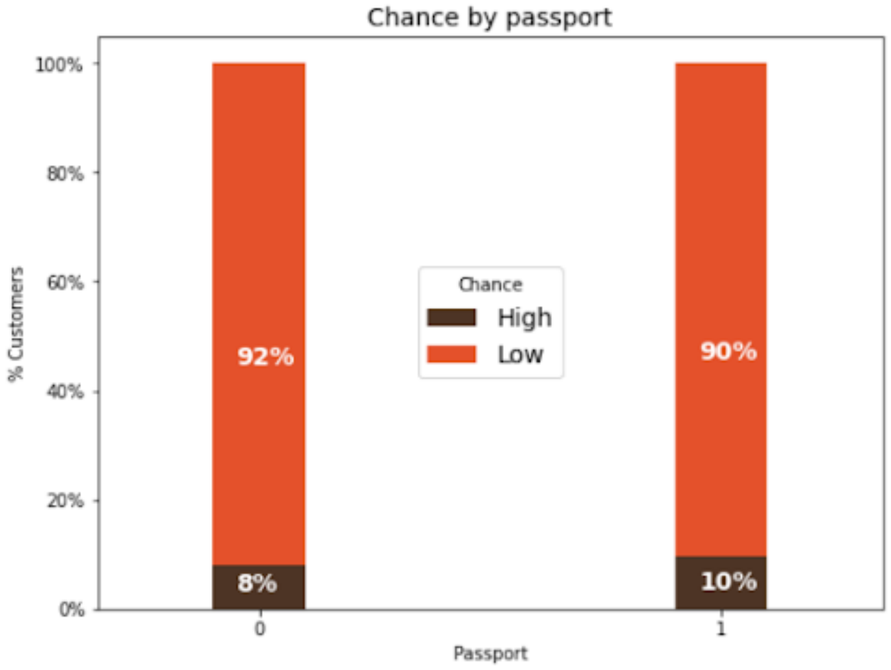
Rata-rata MonthlyIncome di $25.000

Exploratory Data Analysis

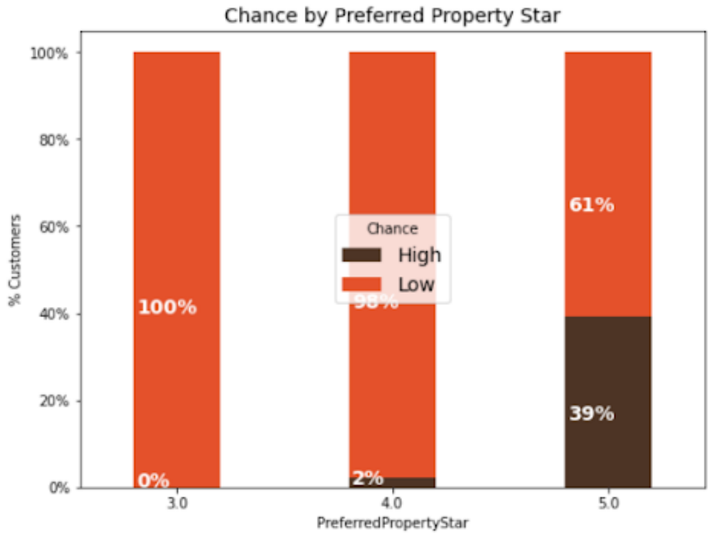


Rata-rata DurationOfPitch di 15 menit





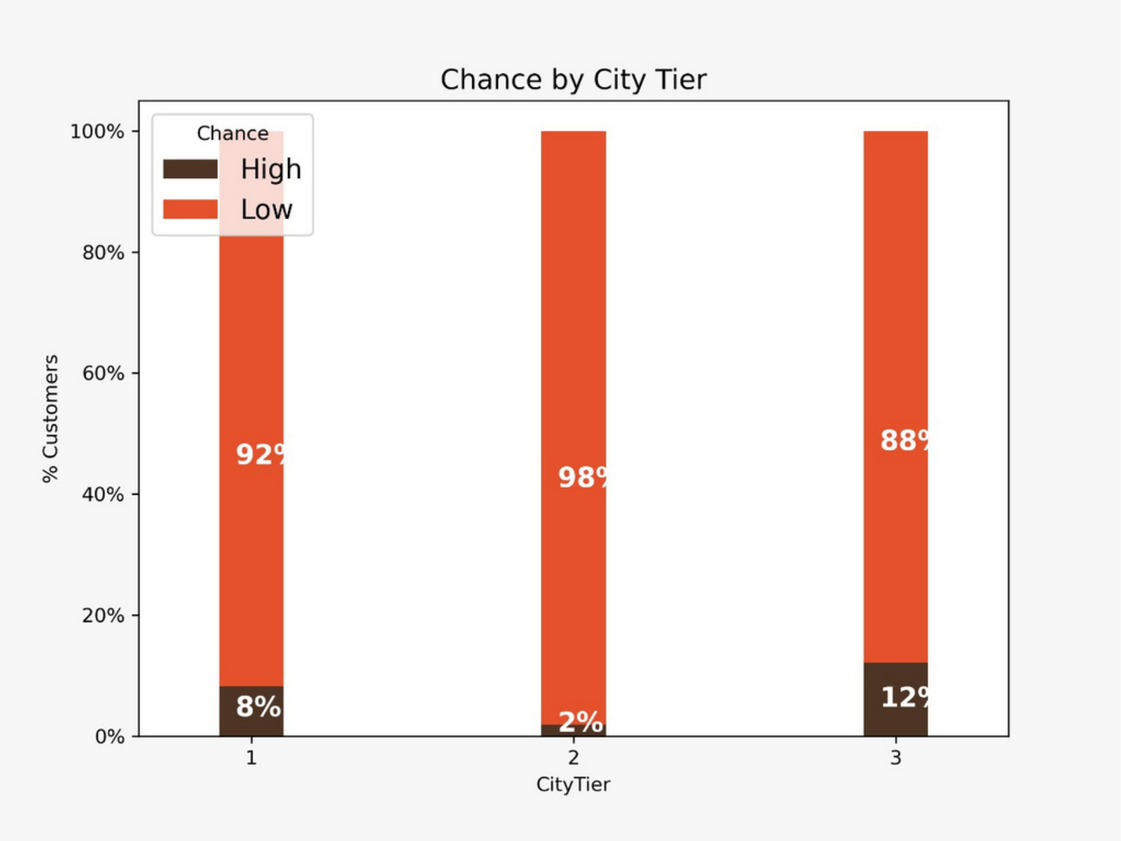
Pre-processing



De Jong and Hartog (2007) found that adequate allocation of time and financial resources as an indication of an innovation-supporting organization is very important to stimulate and also activate entrepreneurial behavior among employees.

International Journal Academic Research in Business Social Sciences.

Pre-processing



OECD Labour Force Statistics working age 15 - 64. Survey Penduduk Antar Sensus 20 - 40 tahun umur produktif.

Details of Tier-wise Classification of Centers Based on Population

(i)	**Classification of centers (tier-wise)**	**Population (as per 2001 census)**
	Tier 1	1,00,000 and above
	Tier 2	50,000 to 99,999
	Tier 3	20,000 to 49,999
	Tier 4	10,000 to 19,999
	Tier 5	5,000 to 9,999
	Tier 6	Less than 5000
(ii)	**Population-group wise classification of centers**	
	Rural center	Population upto 9,999
	Semi-urban Center	From 10,000 to 99,999
	Urban Center	From 1,00,000 to 9,99,999
	Metropolitan Center	10,00,000 and above

OECD SUPAS2015

Pre-processing

Target Generator	• High :
	  ○ Executive, Have Passport, Large Business
	  ○ Age 15-30, Monthly Income 15000- 25000
	• Low

Missing Value	• Filling missing value with mean or median in numerical data • Filling missing value with mode in categorical data
Outlier	Handling outlier menggunakan IQR
One-Hot Encoding	Designation, Marital Status, Occupation, Product Pitched
Class Balancing	From 90:10 to 67:33 using SMOTE
Split Training & Testing	Proportion (80,20) Random State = 42

Model Training

Basic Model

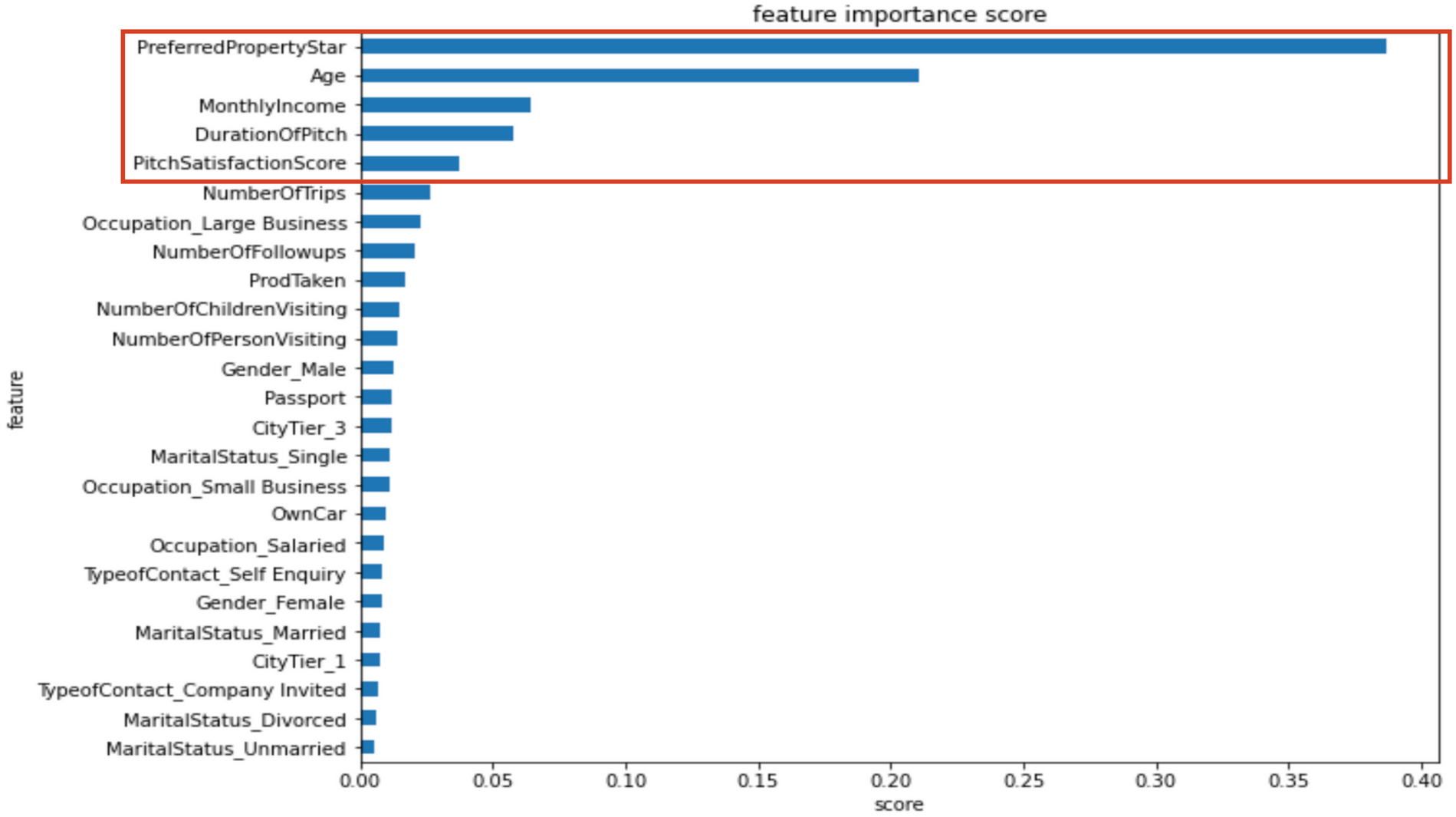
- K Nearest Neighbor
- Logistic Regression
- Decision Tree

Ensemble Model

- Random Forest
- AdaBoost
- XGBoost

	Accuracy	Precision	Recall	F-1 Score
Random Forest (5 top features)	0.99	0.93	0.96	0.95

Feature Importance



Simulasi dampak penggunaan Machine Learning

Studi kasus tanpa model

- Mengevaluasi Customer Excecutive di tahun 2020 sejumlah 1842 orang dan penjualan perusahaan hanya menjual 330 paket
- Biaya Marketing yang keluar menggunakan random marketing asumsi 10% dari harga produk
- Harga paket tour : $2000
- Random marketing : $200 x 1842 **$368.400**
- Paket terjual : $200 x 330 = $66.000 (marketing efektif hanya 17%)

Biaya marketing yang terbuang percuma

$302.400

Studi kasus dengan model

Menggunakan model kemungkinan prediksi marketing 96% terhadap pelanggan yang convert

Biaya Marketing yang keluar menggunakan prediksi model di marketing asumsi 10% dari harga produk

- Harga Paket tour : $2000
- Marketing by Model : $200 x 317 **$63.400**

(Marketing efektif 96%)

MARKETING COST PERUSAHAAN
$ 368.400 > $ 63.400

Marketing cost per customer $200

Without Model

With Model

$ 368.400

$ 63.400

Marketing cost =

1842x $200

317 x $200

Rekomendasi Bisnis & Service Improvement

RAKAMIN ACADEMY

Preferred property star ★★★★★

Hotel 4-5 star

Servis meliputi spa therapies dan aktivitas yang merujuk ke kesehatan, pampering dan rejuvinating

Age & Monthly Income

- Gen Z & Millenials (18 - 35) dengan pendapatan $20,000 perbulan
- Ketertarikan tinggi terhadap well balanced life, nature dan nurture lifestyle & Veganism melihat penelitian lifestyle trending (Creative Insights & Insider.co, 2021)

Rekomendasi *Service*

Training Pitching Method and Satisfaction Score (SOP)

- Pitching between 10-12 menit
- Senyum, salam, sapa
- Tanya apa yang ingin konsumen inginkan
- Tawarkan produk
- Kecuali atas request pelanggan, akomodasi yang ditawarkan hanya bintang 4 dan 5
- Menekankan keunggulan dari produk kita (personalized / bisa di adhoc)
- Aksesibilitas di dalam produk bisa dinikmati oleh berbagai usia
- Kemudahan waktu cancellation period 3-7 hari dengan minimum cancellation fee

Upgrade Feature on WEB

Servis Wellness bisa di tambahkan via website hanya dengan memasukan kode booking:

- Personal yoga
- Spa day package
- Boot Camp
- Variety Meals option
- Cruise

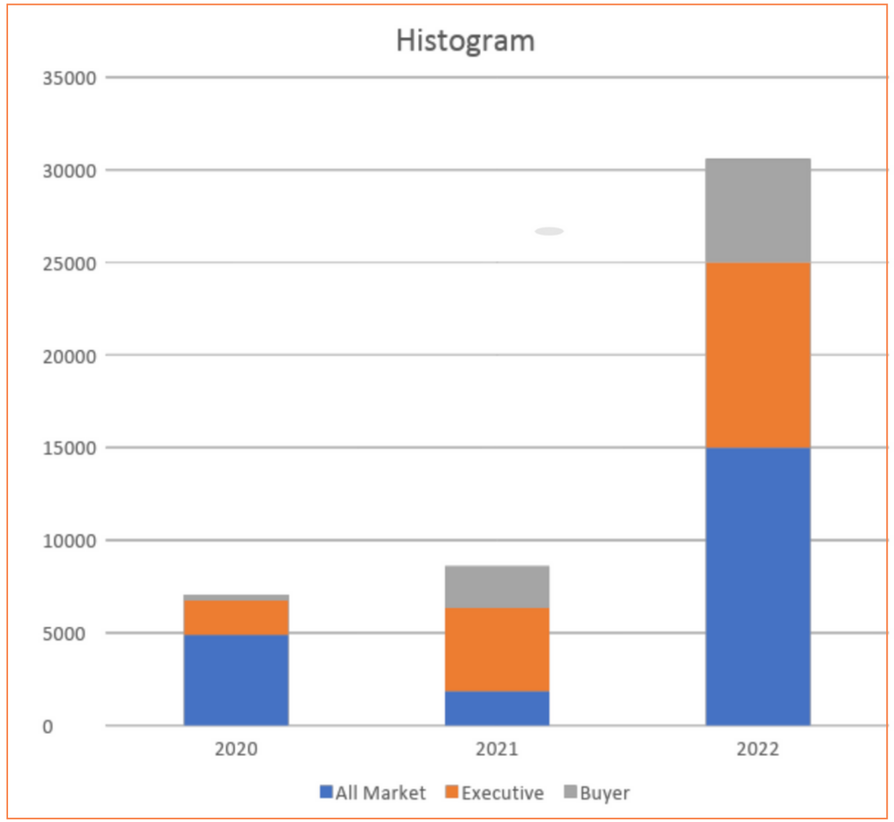
Revenue Simulation

Year	Data	Executive	Targeted buyer	Detected by model	Revenue by model
2020	4888	1842	330	317	$634.000
2021	10000	4500	3000	2250	$4.500.000
2022	15000	10000	7000	5600	$11.200.000

Berdasarkan data 2020 dimana 4888 pembeli dan Market executive sebesar 1842 (37 %) dengan true buyer 18%

Pada 2021 entri data baru sebesar 10.000 dengan designation esecutive sebanyak 4500 (45 %) dengan potensi buyer sebanyak 3000 (67%) paket wellness terjual (improvement of service) dengan confidence level 75 %.

Untuk 2022 perusahaan memiliki 15.000 data baru dengan 10.000 executive dengan 70% (7000) dari angka tersebut sebagai buyer dan dengan confidence level 80 % untuk mendeteksi $11.200.000 revenuenya



Project Evaluation

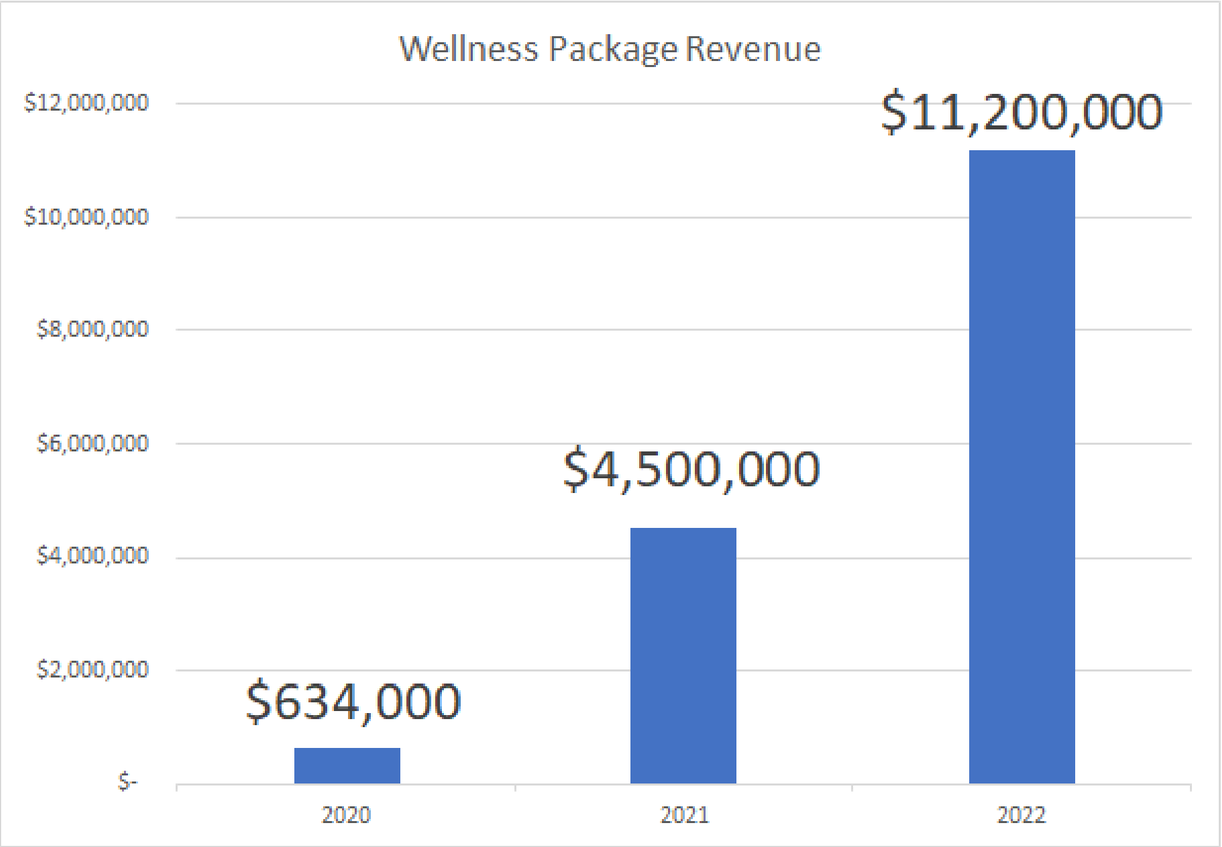
$ 8.000.000 that we lost in 2020

↓

Improvement of Service and Model Learning Service

↓

$ 11.200.000

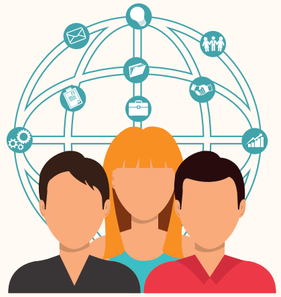


Recover the loss by Improving our Service + Model applied even get more revenue

Lost potential sales di 2020 tercover dengan revenue dengan tambahan $3.200.000

Business Feasibility

Kolaborasi dengan retreat perusahaan (incentive)



Implemetasi model ini pada produk lain yang ada dan produk baru yang akan ada di tahun mendatang.

Travel & Trip

Your peace of mind is our business

Recharge the body, Nourish the soul

APPENDIX

What happens after?

- Pocket Travel App (SOUL)
- Melakukan training machine learning untuk improvisasi hasil

Mengapa hanya memilih Executive?

Melihat mayoritas dari proporsi Designation sebesar 38%. Daripada harus mengaplikasikan ke seluruh segmen, lebih baik hanya di fokuskan pada satu segmentasi yang di masa depan dapat juga diaplikasikan pada segmen / designation yang lainnya.

Duplicate value

Based on ID = none

based on every feature except ID = 395 out of 4888 (8% of total)

Model evaluations

Basic Model

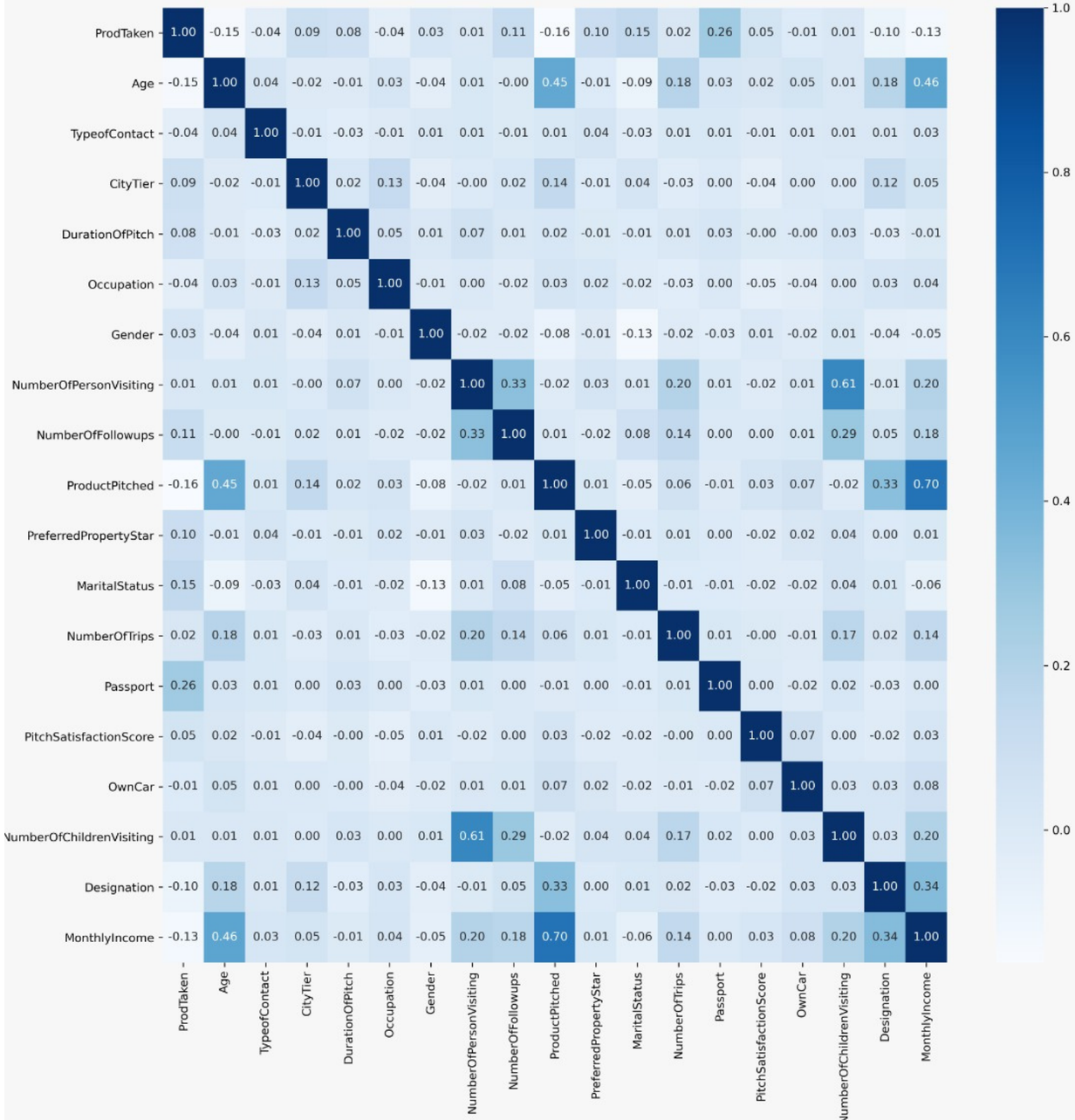
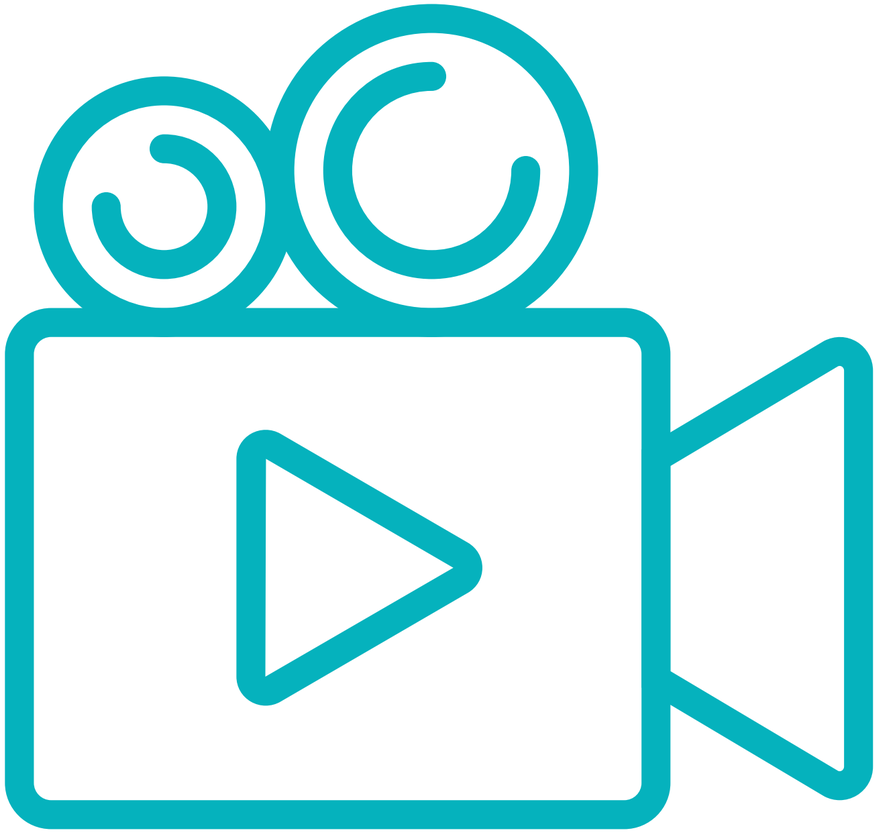
Model	Accuracy	Precision	Recall	F-1 Score
Logistic Regression	0.92	0.5	0.44	0.47
Logistic Regression with Randomized Search	0.96	0.7	0.85	0.77
KNN	0.92	0.5	0.44	0.47
Decision Tree	1	1	1	1
Decision Tree with Randomized Search	0.93	0.54	0.56	0.55
Decision Tree with top 6 feature importance	0.98	0.92	0.81	0.86

Ensembled Model

Model	Accuracy	Precision	Recall	F-1 Score
Bagging : Random Forest	1.0	1.0	1.0	1.0
Bagging : Random Forest (5 top features)	0.99	0.93	0.96	0.95
Boosting : AdaBoost	1.0	1.0	1.0	1.0
Boosting : AdaBoost (5 top features)	0.93	0.53	0.33	0.41
Boosting : XGBoost	1.0	1.0	1.0	1.0
Boosting : XGBoost (5 top features)	1.0	1.0	1.0	1.0

Why Random Forest?

Setelah melihat hasil training dari basic model terlihat bahwa ada overfitting pada model decision tree, menambahkan bias pada model menggunakan model random forest merupakan solusi yang tepat. Namun, setelah melakukan training dengan hyperparameter tuning performa dari model tersebut tetap overfit sehingga untuk mengatasi hal tersebut dengan melakukan training lagi menggunakan 5 top feature.

menggunakan gabungan dari subtree dari decision tree dapat menambahkan bias dengan randomness dari sub tree tersebut. Sehingga metode bagging ini dirasa paling cocok karena ditemukannya overfit pada training untuk model dasar yakni decision tree.

Correlation