



CERTIFIED
JUNIOR DATA SCIENTIST

CAPSTONE
PROJECT

TUJUAN

Memberi kesempatan pada para peserta untuk menerapkan semua materi yang telah dipelajari selama ini mulai dari exploratory data analysis (EDA), preprocessing, feature engineering, modelling, model evaluation hingga presentasi dan storytelling dari insight yang didapatkan.

SUBMISSION FORMAT

File yang diharapkan dari peserta adalah:

1. Jupyter Notebook

Diberikan penjelasan step-by-step dan alasannya.

2. PPT

Point: problem, tujuan, flow/schema olah data, EDA terkait keputusan modeling, feature engineering (feature selection & penambahan variabel), modeling, analisa akhir dari data baru, kesimpulan & saran.

Note: Tidak ada kodingan dalam PPT

DEADLINE

Pengumpulan akan dilakukan terakhir pada hari **Selasa tanggal 1 Februari 2021, 18:00 WIB via GRMDS**. Untuk pengumpulan setelah deadline akan dikenakan **sanksi pengurangan sebesar 10 poin**.

KEY PENILAIAN CAPSTONE PROJECT

Analisis Regresi

1. EDA	15%
2. Pre-processing	25%
3. Feature engineering	15%
4. Modelling & evaluation	15%
5. Presentation & storytelling	30%

Analisis Clustering

1. EDA	30%
2. Pre-processing	15%
3. Feature engineering	10%
4. Modelling & evaluation	15%
5. Presentation & storytelling	30%

Analisis Classification

1. EDA	15%
2. Pre-processing	10%
3. Feature engineering	20%
4. Modelling & evaluation	25%
5. Presentation & storytelling	30%



CASE STUDY



#1

REGRESSION



REGRESSION

Perhitungan gaji karyawan merupakan agenda rutin yang dilakukan Human Resources Department (HRD). Tentunya diharapkan upah yang diterima karyawan dapat disusun dengan adil dan layak. Banyak faktor yang mempengaruhi besaran gaji karyawan, seperti lama masa kerja, departemen, pendidikan, kompetensi, dan lain-lain. HRD ingin membuat sistem penilaian gaji karyawan yang adil dan layak dengan machine learning.

Bantu HRD untuk menganalisa dan **memprediksi gaji karyawan** dengan machine learning.

REGRESSION | METADATA

Data: **estimated monthly income.csv**

Variable	Description
Age	Umur karyawan.
BusinessTravel	Perjalanan dinas.
Department	Departemen.
DistanceFromHome	Jarak ke kantor.
Education	Pendidikan.
EducationField	Bidang pendidikan.
EnvironmentSatisfaction	Kepuasan terhadap lingkungan kerja.
Gender	Jenis kelamin.

REGRESSION | METADATA

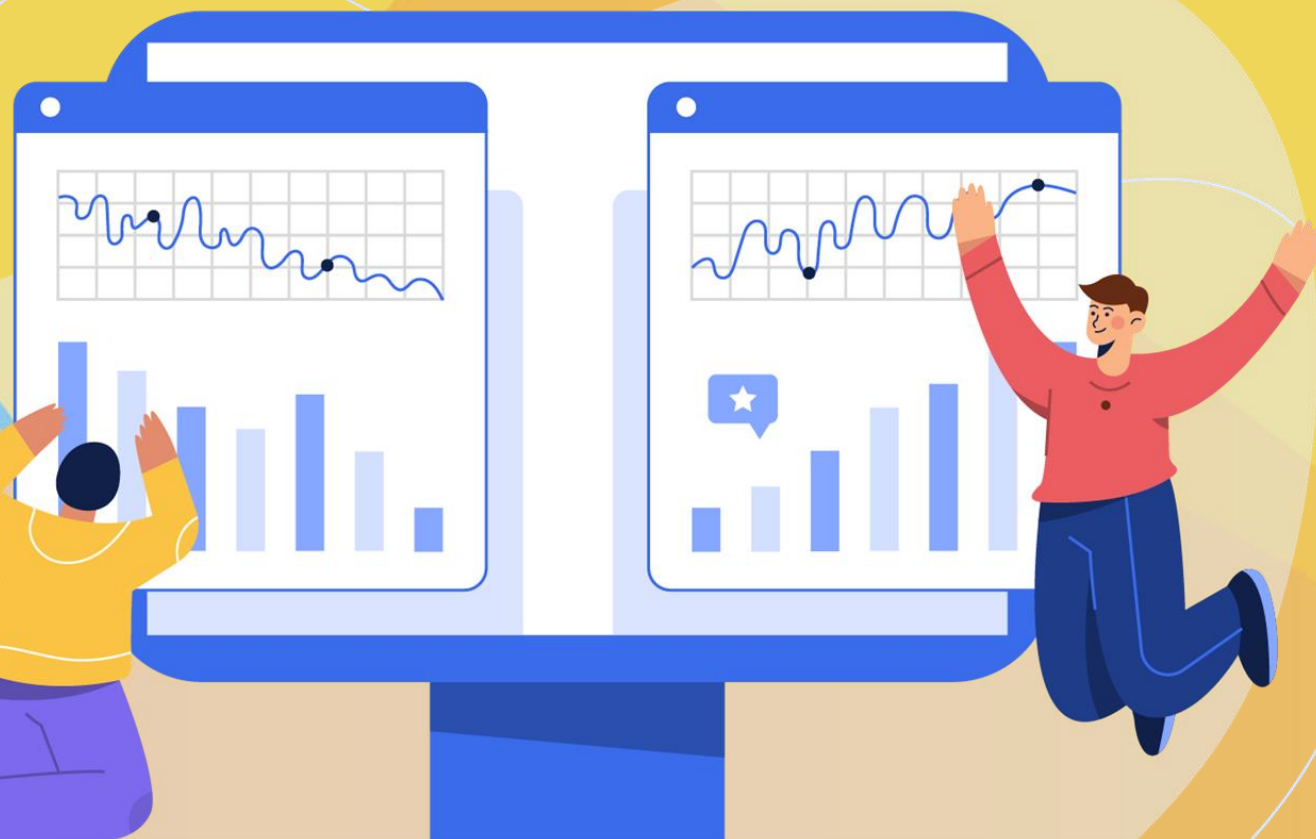
Variable	Description
JobLevel	Level jabatan.
JobRole	Posisi.
JobSatisfaction	Kepuasan terhadap pekerjaan.
MaritalStatus	Status pernikahan.
MonthlyIncome	Gaji bulanan.
NumCompaniesWorked	Jumlah perusahaan pernah bekerja.
OverTime	Lembur.
PercentSalaryHike	Kenaikan gaji.
PerformanceRating	Rating performa kerja.
RelationshipSatisfaction	Kepuasan terhadap hubungan rekan kerja.

REGRESSION | METADATA

Variable	Description
TotalWorkingYears	Lama pengalaman kerja.
TrainingTimesLastYear	Pelatihan yang diikuti tahun lalu.
WorkLifeBalance	Kualitas work life balance.
YearsAtCompany	Lama masa kerja.
YearsInCurrentRole	Masa kerja di jabatan saat ini.
YearsSinceLastPromotion	Masa kerja sejak promosi terakhir.
YearsWithCurrManager	Masa kerja dengan manajer saat ini.

REGRESSION | DATASET

Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EnvironmentSatisfaction	Gender	HourlyRate	JobLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	NumCompaniesWorked	OverTime	PercentSalaryHike	PerformanceRating	RelationshipSatisfaction	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearstinceLastPromotion	YearsWithCurrManager
18	Travel_Rarely	230	Research & Development	3	3	Life Sciences	3	Male	54	1	Laboratory Technician	3	Single	1420	1	No	13	3	3	0	2	3	0	0	0	0
18	Travel_Rarely	812	Sales	10	3	Medical	4	Female	69	1	Sales Representative	3	Single	1200	1	No	12	3	1	0	2	3	0	0	0	0
18y	Travel_Frequently	1306	Sales	5	3	Marketing	2	Male	69	1	Sales Representative	2	Single	1878	1	Yes	14	3	4	0	3	3	0	0	0	0
18	Non-Travel	287	Research & Development	5	2	Life Sciences	2	Male	73	1	Research Scientist	4	Single	1051	1	No	15	3	4	0	2	3	0	0	0	0
18	Non-Travel	247	Research & Development	8	1	Medical	3	Male	80	1	Laboratory Technician	3	Single	1904	1	No	12	3	4	0	0	3	0	0	0	0
18	Non-Travel	1124	Research & Development	1	3	Life Sciences	4	Female	97	1	Laboratory Technician	4	Single	1611	1	No	15	3	3	0	5	4	0	0	0	0
18y	Travel_Frequently	544	Sales	3	2	Medical	2	Female	70	1	Sales Representative	4	Single	1569	1	Yes	12	3	3	0	2	4	0	0	0	0
18	Non-Travel	1431	Research & Development	14	3	Medical	2	Female	33	1	Research Scientist	3	Single	2389.601929	1	No	16	3	3	0	4	1	0	0	0	0
19	Travel_Rarely	578	Sales	22	1	Marketing	4	Male	50	1	Sales Representative	3	Single	1675	1	Yes	19	3	4	0	2	2	0	0	0	0
19	Travel_Rarely	1181	Research & Development	3	1	Medical	2	Female	79	1	Laboratory Technician	2	Single	1483	1	No	14	3	4	1	3	3	1	0	0	0
19y	Travel_Frequently	602	Sales	1	1	Technical Degree	3	Female	100	1	Sales Representative	1	Single	2325	0	No	21	4	1	1	5	4	0	0	0	0
19	Travel_Rarely	303	Research & Development	2	3	Life Sciences	2	Male	47	1	Laboratory Technician	4	Single	1102	1	No	22	4	3	1	3	2	1	0	1	0
19	Travel_Rarely	489	Human Resources	2	2	Technical Degree	1	Male	52	1	Human Resources	4	Single	2564	1	No	12	3	3	1	3	4	1	0	0	0
19	Travel_Rarely	419	Sales	21	3	Other	4	Male	37	1	Sales Representative	2	Single	2121	1	Yes	13	3	2	1	3	4	1	0	0	0
19	Travel_Rarely	645	Research & Development	9	2	Life Sciences	3	Male	54	1	Research Scientist	1	Single	2552	1	No	25	4	3	1	4	3	1	1	0	0
19	Non-Travel	504	Research & Development	10	3	Medical	1	Female	96	1	Research Scientist	2	Single	1859	1	Yes	25	4	2	1	2	4	1	1	0	0
19	Travel_Rarely	265	Research & Development	25	3	Life Sciences	2	Female	57	1	Research Scientist	4	Single	2994	1	Yes	12	3	4	1	2	3	1	0	0	1
20y	Travel_Frequently	871	Research & Development	6	3	Life Sciences	4	Female	66	1	Laboratory Technician	4	Single	2926	1	Yes	18	3	2	1	5	3	1	0	1	0
20	Travel_Rarely	959	Research & Development	1	3	Life Sciences	4	Female	83	1	Research Scientist	2	Single	2836	1	No	13	3	4	1	0	4	1	0	0	0
20	Travel_Rarely	1362	Research & Development	10	1	Medical	4	Male	32	1	Research Scientist	3	Single	2484.257211	1	Yes	11	3	4	1	5	3	1	0	1	1
20	Travel_Rarely	500	Sales	2	3	Medical	3	Female	49	1	Sales Representative	3	Single	2044	1	No	13	3	4	2	3	2	2	2	0	2
20	Travel_Rarely	129	Research & Development	4	3	Technical Degree	1	Male	84	1	Laboratory Technician	1	Single	2973	1	No	19	3	2	1	2	3	1	0	0	0
20	Travel_Rarely	1097	Research & Development	11	3	Medical	4	Female	98	1	Research Scientist	1	Single	2600	1	Yes	15	3	1	1	2	3	1	0	0	0
20y	Travel_Frequently	769	Sales	9	3	Marketing	4	Female	54	1	Sales Representative	4	Single	2312.640088	1	Yes	14	3	2	2	3	3	2	2	0	0



#2

REGRESSION



REGRESSION

Apartemen/Rumah adalah salah satu bentuk properti yang beberapa tahun belakangan ini cukup digemari oleh masyarakat khususnya masyarakat perkotaan, kaum urban dan millennials. Tak jarang pula begitu banyak apartemen yang disewakan kepada customer yang ingin mencari rumah sewa.

Sebagai seorang data scientist **analisa** dan **prediksikan harga sewa rumah** pada data yang diberikan. Agar pemilik apartemen bisa mudah menyewakan apartemennya dan begitu pula customer bisa lebih mudah dalam mencari rumah/apartemen sewa.

REGRESSION | METADATA

Data: **Housing Rental Prices.csv**

Variable	Description
rent (Target Variable)	sewa bulanan dalam IDR
Number of bedrooms	jumlah kamar tidur
Number of bathrooms	jumlah kamar mandi
Number of Balconies	jumlah balkon
Brokerage amount	biaya real-estate dalam IDR (sekali dalam durasi perjanjian)
Deposit Amount	Jumlah jaminan pemilik dalam IDR. Ini akan dikembalikan sesuai kesepakatan (sekali dalam durasi kesepakatan)
Maintenance amount	Biaya perawatan gedung (bulanan)

REGRESSION | METADATA

Variable	Description
Built-Up Area	luas area dalam sq ft
Super Built-Up Area	area dalam sq.ft
Type of Furnishing	Menunjukkan jenis perabotan
Availability for	rumah/apartemen tersedia untuk keluarga, pria , wanita, sarjana
Address	tujuan lokasi
Floor Number	Nomor lantai apartemen
Home Facing	Arah pintu
Floor-type	Jenis material lantai yang digunakan ex.marmer dll
Gate Community	Keamanan gerbang tersedia atau tidak
Corner Property	Kolom Corner_pro menunjukkan properti milik lokasi sudut jalan

REGRESSION | METADATA

Variable	Description
Parking Count	Berapa banyak kendaraan yang dapat diparkir
WheelChairFacility	ya/tidak
Pet-Friendly	ya/tidak
Agreement Duration	Dalam bulan
Electricity Bill	Siapa yang akan membayar
Power Backup Facility	Fasilitas listrik darurat
Property age in years	Usia Properti

REGRESSION | DATASET

bedroom	bathrooms	area	furnishing	available_for	address	floor_number	facing	floor_type	gate_community	corner_pro	parking	wheelchairadpti
2	2	1050	Unfurnished	All	Sadguru hights,	5	West	Marble	Yes	No	0	None
2	2	760	Unfurnished	All	Manav Silver Sp	5	East	Vitrified	Yes	Yes	2	None
3	3	0	Semifurnished	All	Saarrthi Souveni	1	South-West	Vitrified	Yes	No	1	None
1	1	628	Furnished	Family Only	Dhan Residency	3	East	Mosaic	Yes	No	1	None
2	2	668	Semifurnished	Family , Bachelo	Saptsiddhi Sava	6	South	Polished concre	Yes	Yes	1	None
2	2	950	Semifurnished	Family Only	Vimal Apartment	1	No Direction	Ceramic	No	No	1	None
3	3	1530	Semifurnished	Family Only	Atul Westernhills	2	East	Vitrified	Yes	No	2	Yes
2	2	900	Unfurnished	All	SRK Herambh, F	1	No Direction	Not provided	No	No	1	None
3	3	0	Semifurnished	Family Only	Tuscan Estate S	7	North	Not provided	Yes	No	1	None
3	3	1400	Unfurnished	All	Magarpatta City	2	North-West	Vitrified	Yes	No	1	None
2	2	0	Semifurnished	Family Only	On Request, Bib	3	South-East	Not provided	Yes	No	1	None
2	2	1000	Unfurnished	All	Supriya Gardens	2	North-East	Mosaic	Yes	Yes	1	None
3	3	1500	Semifurnished	All	Erande Arora Re	1	East	Not provided	Yes	No	1	None
2	2	0	Unfurnished	All	SBM West Wind	2	No Direction	Not provided	No	No	1	None
1	1	0	Unfurnished	Family Only	On Request, Bib	0	West	Not provided	No	No	1	None
1	1	0	Unfurnished	Family Only	Sai vilasita Apar	1	No Direction	Not provided	Yes	No	0	None
1	1	0	Unfurnished	All	maurya housing	1	No Direction	Not provided	Yes	No	0	None
1	1	445	Semifurnished	All	Puraniks Aldea,	1	East	Vitrified	Yes	No	1	None
2	2	836	Unfurnished	All	Anita Residency	8	North-East	Not provided	Yes	Yes	0	None
1	2	590	Semifurnished	Family Only	Geeta Golden P,	6	No Direction	Ceramic	Yes	No	1	None
3	3	1350	Semifurnished	Family Only	Paranjape Blue I	1	No Direction	Vitrified	Yes	No	1	None
1	1	350	Semifurnished	Family Only	Jal Vihar, Anand	2	East	Vitrified	Yes	No	1	None
2	2	870	Unfurnished	All	AG West One, V	7	West	Vitrified	Yes	No	1	None
2	2	53.44	Semifurnished	All	The Leaf, Yewak	5	West	Wood	Yes	Yes	1	Yes



#3

CLASSIFICATION



CLASSIFICATION

Analisa kredit adalah penilaian kelayakan terhadap pengaju kredit. Fungsi analisa kredit di antaranya sebagai dasar bagi bank dalam menentukan diterimanya pengajuan kredit, menentukan tingkat suku bunga kredit serta jaminan yang disyaratkan untuk dipenuhi nasabah, dan evaluasi. Banyak faktor yang dipertimbangkan analis kredit sebelum memberikan rekomendasi menerima atau menolak sebuah pengajuan kredit. **Bank A ingin membuat sistem penilaian kelayakan kredit secara otomatis** berdasarkan data yang mereka miliki. Lakukan analisis dan prediksikan **flag_kredit_macet** berdasarkan variabel yang ada.

CLASSIFICATION | METADATA

Data: **npl_credit_scoring.csv**

Variable	Description
jumlah_kartu	Jumlah kartu aktif yang dimiliki pelanggan.
outstanding	Total saldo pemakaian kartu kredit.
limit_kredit	Jumlah maksimum limit kredit yang dapat digunakan.
tagihan	Jumlah tagihan pada bulan terakhir.
total_pemakaian	Total pemakaian kartu kredit (tunai dan retail) pada bulan terakhir.
total_pemakaian_tunai	Total pemakaian transaksi tunai pada bulan terakhir (<i>cash advance</i>).
total_pemakaian_retail	Total pemakaian transaksi retail pada bulan terakhir.

CLASSIFICATION | METADATA

Variable	Description
sisa_tagihan_tidak_terbayar	Jumlah tagihan tidak terbayar pada bulan terakhir.
rasio_pembayaran	Rasio perbandingan jumlah yang dibayar dengan tagihan pada bulan terakhir.
presentasi_overlimit	Presentasi overlimit.
nomor_pelanggan	ID pelanggan.
utilisasi_3bulan	Utilisasi kartu kredit selama 3 bulan terakhir.
pemakaian_3bln_per_limit	Rata-rata pemakaian selama 3 bulan terakhir dibagi total limit.
rasio_pembayaran_3bulan	Rata-rata rasio pembayaran 3 bulan terakhir.
utilisasi_6bulan	Utilisasi kartu kredit selama 3 bulan sebelum 3 bulan terakhir.

CLASSIFICATION | METADATA

Variable	Description
pemakaian_6bln_per_limit	Rata-rata pemakaian selama 3 bulan sebelum 3 bulan terakhir dibagi total limit.
rasio_pembayaran_6bulan	Rata-rata rasio pembayaran 3 bulan sebelum 3 bulan terakhir.
skor_pembayaran	Nilai yang dihitung berdasarkan jumlah minimum pembayaran yang dilakukan pelanggan dalam 6 bulan terakhir.
sisatagihan_per_jumlah_kartu	Jumlah tagihan yang tidak terbayar pada bulan terakhir dibagi jumlah kartu aktif.
sisatagihan_per_limit	Jumlah tagihan yang tidak terbayar pada bulan terakhir dibagi total limit.
total_pemakaian_per_limit	Total pemakaian kartu kredit (tunai dan retail) pada bulan terakhir dibagi total limit.
jumlah_tahun_sejak_pembuatan_kredit	Jumlah kredit sejak pembuatan kartu kredit pertama kali.
flag_kredit_macet	Flag yang menandakan apakah kartu kredit macet atau tidak.

CLASSIFICATION | DATASET

X	jumlah_kartu	outstanding	limit_kredit	tagihan	total_pemakaian_t unai	total_pemakai an_retail	sisatagihan_t idak_terbayar	kode_cabang	rasio_pembay aran	persentase_ov erlimit	rasio_pembay aran_3bulan	rasio_pembay aran_6bulan	skor_delikuen si	flag_kredit_m acet	jumlah_tahun_s ejak_pembukaa n_kredit	total_pema kain	sisatagihan_ per_jumlah_k artu	sisatagihan_ per_limit	total_pemakai an_per_limit	pemakaian_3 bln_per_limit	pemakaian_6 bln_per_limit	utilisasi_3bula n	utilisasi_6bula n
1	2	36158	7.00E+06	23437	0	94	263231		102.19	0	74.78	100	0	0	15.41666667	94	131615	0.003760429	1.34E-05	0.01171875	0.017809524	0.013228036	0.021949333
2	2	268691	1.00E+07	254564	0	1012	0A		0	0	0	0	0	0	0.75	1012	0	0	0.0001012	0	0	0.004231625	3.00E-04
3	3	6769149	2.80E+07	4159779	0	0	0A		100	0	100	100.91	0	0	10.75	0	0	0	0.040517857	0.047703226	0.249389277	0.267852798	
4	4	3496732	2.10E+07	111231	0	2536660	581334G		100	0	25.01	22.64	0	0	19.75	2536660	145333.5	0.027682571	0.120793333	0.055970595	0.016850794	0.101912107	0.346634603
5	2	9402085	1.00E+07	6099283	0	2666558	5951865A		95.99	0	97.49	99.84	0	0	1.666666667	2666558	2975932.5	0.5951865	0.2666558	0.323027	0.131162467	0.707865025	0.336571367
6	2	6227439	8.00E+07	2081248	0	3690250	4613435A		0	0	28.46	42.03	0	0	10.25	3690250	2306717.5	0.057667938	0.046128125	0.017184375	0.117513096	0.038963606	0.181863421
7	2	3906290	4.00E+06	2043682	0	230400	3314046A		18.47	0	24.95	17.89	0	0	4.666666667	230400	1657023	0.8285115	0.0576	0.01875	0.166666668	1.014775813	0.977230833
8	4	9534837	2.00E+07	3692028	0	9327612	7881069A		58.55	0	105.24	166.67	0	0	0.75	9327612	1970267.25	0.39405345	0.4663806	0.2392893	0.241598694	0.261753338	0.22402625
9	2	4145065	5.00E+06	4021399	0	335680	4122425B		15.24	0	0	33.33	0	0	5.666666667	335680	2061212.5	0.824485	0.067136	0.19604205	0	0.47002435	0
10	4	1818606	7.00E+06	1765911	0	0	1627786A		100	0	58.4	62.68	0	0	8.5	0	406946.5	0.232540857	0	0.075	0.094800619	0.315145179	0.626447904
11	4	55520052	8.40E+07	12626615	0	31036235	0A		100	0	100	100	0	0	14.5	31036235	0	0	0.369478988	0.164046339	0.19530254	0.396655988	0.230336444
12	3	7004631	2.30E+07	610083	0	0	0A		100.13	0	74.99	18.52	0	0	10.66666667	0	0	0	0.057634076	0	0.342642826	0.497703464	
13	2	5664700	9.00E+06	2334602	0	2878725	4494421B		30.8	0	68.4	100	0	0	20.83333333	2878725	2247210.5	0.499380111	0.319858333	0.098427194	0.297481481	0.402411556	0.46713437
14	3	7691430	7.00E+06	6463709	0	0	7398073E		0	10.02	13.04	26.66	0	1	1	0	2466024.333	1.056867571	0	0.067480357	0.079714286	1.039140607	0.929415857
15	3	8640053	7.00E+06	8035169	0	326100	8269767A		17.53	23.26	29.47	24.24	0	0	6.666666667	326100	2756589	1.181395286	0.046585714	0.250685714	0.239319047	1.163175571	1.416155476
16	2	188958	3.00E+06	189816	0	722	0B		0	0	65.25	100	0	0	2	722	0	0	0.000240667	0.016566	0	0.02255475	0.016172
17	4	39700203	1.00E+08	5366000	0	3101354	0A		99.9	0	75.01	74.41	0	0	11.58333333	3101354	0	0	0.03101354	0	0.212714983	0.388039143	0.486339247
18	3	5959226	5.00E+06	5579255	0	380000	5902767G		20.97	19.54	18.29	13.26	0	0	2.916666667	380000	1967589	1.1805534	0.076	0.137606	0.112229534	1.0560083	0.8936248
19	2	5309091	5.00E+06	5309314	0	0	5310555A		9.17	6.21	7.43	14.26	0	1	2.833333333	0	2655277.5	1.062111	0	0.02369	0.2685276	0.98833345	0.9642602
20	2	701019	4.00E+06	306224	0	457725	0A		100	0	100	100	0	0	2.75	457725	0	0	0.11443125	0.140283438	0.0415	0.13060025	0.07637
21	3	13895389	1.30E+07	13552786	0	1583400	0A		164.09	0	98.12	100	0	0	12.41666667	1583400	0	0	0.1218	0.677190654	0.216974436	0.637737731	0.205
22	2	1178925	3.00E+06	879730	0	331790	0F		95.69	0	49.57	37.33	0	0	2	331790	0	0	0.110596667	0	0.292907223	0.545604083	



#4

CLASSIFICATION



CLASSIFICATION

Happy Customer Bank adalah bank swasta menengah yang menangani semua jenis produk perbankan, seperti tabungan, giro, produk investasi, serta produk kredit, di antara penawaran lainnya. Bank ini juga melakukan cross-selling produk kepada nasabah yang sudah ada dan untuk melakukannya mereka menggunakan berbagai jenis komunikasi seperti telecalling, e-mail, rekomendasi tentang net banking, mobile banking, dan lain-lain. Dalam hal ini, Happy Customer Bank ingin melakukan cross-selling kartu kreditnya kepada nasabah yang sudah ada. Bank telah mengidentifikasi satu set pelanggan yang memenuhi syarat untuk mengambil kartu kredit ini. Sekarang, bank meminta bantuan Anda dalam **mengidentifikasi pelanggan yang dapat menunjukkan ketertarikan yang lebih tinggi terhadap kartu kredit yang direkomendasikan.** Lakukan analisis dan prediksikan **Is_Lead** berdasarkan variabel yang ada.

CLASSIFICATION | METADATA

Data: **credit_card_lead_prediction.csv**

Variable	Description
ID	Pengenal unik untuk satu baris.
Gender	Jenis kelamin pelanggan.
Age	Usia pelanggan (dalam tahun).
Region_Code	Kode wilayah untuk pelanggan.
Occupation	Jenis pekerjaan untuk pelanggan.
Channel_Code	Kode channel akuisisi untuk pelanggan (encoded).
Vintage	Vintage untuk pelanggan (dalam bulan).

CLASSIFICATION | METADATA

Variable	Description
Credit_Product	Apakah nasabah memiliki produk kredit aktif (KPR, personal loan, kartu kredit, dan lain-lain).
Avg_Account_Balance	Rata-rata saldo rekening nasabah dalam 12 bulan terakhir.
Is_Active	Apakah pelanggan aktif dalam 3 bulan terakhir.
Is_Lead	0 : Pelanggan tidak tertarik 1 : Pelanggan tertarik

CLASSIFICATION | DATASET

	ID	Gender	Age	Region_Code	Occupation	Channel_Code	Vintage	Credit_Product	Avg_Account_Balance	Is_Active	Is_Lead
0	KXFDHGFP	Female	77	RG283	Other	X3	31	No	903545	Yes	0
1	7ZCSVMZV	Male	44	RG254	Self_Employed	X2	21	Yes	975386	No	0
2	SNPMKRHF	Male	27	RG261	Salaried	X1	20	No	736875	No	0
3	PTYGLPWY	Male	54	RG268	Self_Employed	X2	98	No	886404	Yes	0
4	IVNJGNPN	Female	39	RG274	Salaried	X1	19	Yes	614740	No	0
...
9995	HUNORYGV	Female	52	RG284	Self_Employed	X2	92	No	703197	Yes	1
9996	CHM5WUGJ	Male	57	RG275	Self_Employed	X3	86	NaN	662827	No	1
9997	QTZYPHRJ	Male	41	RG274	Self_Employed	X3	15	NaN	1097236	No	1
9998	UA3VZFKY	Male	46	RG254	Other	X2	91	NaN	622973	No	1
9999	EPOBTHHQ	Male	40	RG279	Other	X2	33	Yes	879373	Yes	1

10000 rows × 11 columns



#5

CLUSTERING



CLUSTERING

Manajemen inventory adalah hal vital bagi supermarket, tujuannya untuk pengelolaan stok barang. Sederhananya, supermarket harus memastikan ketersediaan semua barang yang mereka jual. Kekosongan stok adalah kerugian karena supermarket kehilangan potensi penjualan. Namun, supermarket juga memiliki keterbatasan gudang untuk menyimpan barang-barang dagangannya, sehingga diperlukan stock control untuk mengelola inventory supermarket.

Bantulah pemilik supermarket untuk melakukan **stock control**. Lakukan analisis dan tentukan **manakah produk yang harus distok banyak dan sedikit**.

CLUSTERING | METADATA

Data: **stock keeping unit.xlsx**

Variable	Description
KODE	Kode produk.
NAMA	Nama produk.
QTY	Jumlah terjual.
HARGA	Harga jual.
DISC (%)	Discount persen.
DISC (Rp)	Discount harga.
SHELF LIFE	Masa penyimpanan dalam bulan.

CLUSTERING | DATASET

KODE	NAMA	QTY	HARGA	DISC (%)	DISC (Rp)	SHELF LIFE
017447	INDOMI AYAM SPC/CR	-5.000	32300.00	0.000	0	36
017447	INDOMI AYAM SPC/CR	-1.000	32300.00	0.000	0	28
024057	BAWANG PUTIH TIMB.	0.038	9100.00	0.000	0	10
024249	PERMEN PURE MRT	0.038	30000.00	0.000	0	6
024543	BAWANG BOMBAY TIMB.	0.056	5500.00	0.000	0	11
024241	PERMEN XINO XIXI/MM	0.068	31000.00	0.000	0	7
026070	S&W CHERY BUAH	0.071	48750.00	0.000	0	13
024238	PERMEN WHITE RABBIT	0.082	34000.00	0.000	0	12
024286	PERMEN TOFFE	0.084	33000.00	0.000	0	6
024057	BAWANG PUTIH TIMB.	0.088	9100.00	0.000	0	31
024543	BAWANG BOMBAY TIMB.	0.094	5500.00	0.000	0	35
024238	PERMEN WHITE RABBIT	0.098	34000.00	0.000	0	14
024250	PERMEN BIGHTO MRT	0.114	30000.00	0.000	0	14
024057	BAWANG PUTIH TIMB.	0.118	9100.00	0.000	0	29
024359	TELUR AYAM NEGRI RAS	0.122	8200.00	0.000	0	6
024543	BAWANG BOMBAY TIMB.	0.124	5500.00	0.000	0	7
024247	PERMEN ASEM JERUK	0.126	20000.00	0.000	0	27
024543	BAWANG BOMBAY TIMB.	0.128	5500.00	0.000	0	22
024057	BAWANG PUTIH TIMB.	0.132	9100.00	0.000	0	22
024246	PERMEN JELLY MALAYSA	0.132	21000.00	0.000	0	25
026021	SAWO	0.134	4700.00	0.000	0	7
026201	APEL BLUE CHELAND	0.142	16400.00	0.000	0	30
024057	BAWANG PUTIH TIMB.	0.146	9100.00	0.000	0	23
024057	BAWANG PUTIH TIMB.	0.148	9100.00	0.000	0	24
024533	TOMAT MRT	0.161	4700.00	0.000	0	34
024120	BAWANG MERAH TIMB.	0.162	14700.00	0.000	0	34



#6

CLUSTERING



CLUSTERING

Data yang digunakan adalah perilaku pengguna kartu kredit. Data mencakup jumlah transaksi, nominal transaksi, hingga tenor berlangganan. Anda merupakan pelamar yang akan mengisi posisi sebagai data analyst. Kemudian, Anda ditugaskan untuk memberikan hal-hal menarik dari data tersebut. User memberikan tugas kepada Anda untuk melakukan **segmentasi pelanggan dengan data yang ada**, sehingga perusahaan dapat memberikan treatment kepada pelanggan sesuai dengan segmentasinya.

CLUSTERING | METADATA

Data: **credit_card_user_segmentation.csv**

Variable	Description
Cust_ID	Identitas pengguna.
Saldo	Jumlah saldo yang tersisa dalam akun mereka untuk pembelian (USD).
Frekuensi_Saldo	Seberapa sering saldo diperbarui.
Pembelian	Jumlah pembelian yang dilakukan dari akun (USD).
Pembelian_Sekali_Jalan	Jumlah pembelian maksimum dilakukan dalam sekali jalan (USD).
Pembelian_Angsuran	Jumlah pembelian dilakukan secara angsuran (USD).
Uang_Dibayar_Dimuka	Uang tunai di muka yang diberikan oleh pengguna (USD).

CLUSTERING | METADATA

Variable	Description
Frekuensi_Pembelian	Seberapa sering pembelian dilakukan.
Frekuensi_Pembelian_Satu_Kali	Seberapa sering pembelian dalam sekali jalan terjadi.
Frekuensi_Pembelian_Angsuran	Seberapa sering pembelian secara mencicil dilakukan.
Frekuensi_Uang_Dibayar_Dimuka	Seberapa sering uang tunai di muka dibayarkan.
Transaksi_Uang_Dibayar_Dimuka	Jumlah transaksi yang dilakukan dengan "uang tunai di muka".
Transaksi_Pembelian	Jumlah transaksi pembelian yang dilakukan.

CLUSTERING | METADATA

Variable	Description
Batas_Kredit	Batas kartu kredit untuk pengguna (USD).
Pembayaran	Jumlah pembayaran yang dilakukan oleh pengguna (USD).
Pembayaran_Minimal	Jumlah pembayaran minimum yang dilakukan oleh pengguna (USD).
Persen_Pembayaran_Penuh	Persentase pembayaran penuh yang dibayarkan oleh pengguna (persentase).
Tenor	Jangka waktu layanan kartu kredit untuk pengguna.

CLUSTERING | DATASET

	Cust_ID	Saldo	Frekuensi_Saldo	Pembelian	Pembelian_Satu_Kali	Pembelian_Angsuran	Uang_Dibayar_Dimuka	Frekuensi_Pembelian	Frekuensi_Pembelian_Satu_Kali
0	C10001	40.900749	0.818182	95.40	0.00	95.40	0.000000	0.166667	0.000000
1	C10002	3202.467416	0.909091	0.00	0.00	0.00	6442.945483	0.000000	0.000000
2	C10003	2495.148862	1.000000	773.17	773.17	0.00	0.000000	1.000000	1.000000
3	C10004	1666.670542	0.636364	1499.00	1499.00	0.00	205.788017	0.083333	0.083333
4	C10005	817.714335	1.000000	16.00	16.00	0.00	0.000000	0.083333	0.083333
...
8945	C19186	28.493517	1.000000	291.12	0.00	291.12	0.000000	1.000000	0.000000
8946	C19187	19.183215	1.000000	300.00	0.00	300.00	0.000000	1.000000	0.000000
8947	C19188	23.398673	0.833333	144.40	0.00	144.40	0.000000	0.833333	0.000000
8948	C19189	13.457564	0.833333	0.00	0.00	0.00	36.558778	0.000000	0.000000
8949	C19190	372.708075	0.666667	1093.25	1093.25	0.00	127.040008	0.666667	0.666667

8950 rows × 18 columns

KONSULTASI

1. Rabu, 26 Januari (19.00 - 21.00 WIB)
(Reggresion)
1. Jumat, 28 Januari 2021 (19.00 - 21.00 WIB)
(Classification dan Clustering)





LET'S HANDS-ON!