

# About Us

Dompét Kilat is a peer-to-peer lending company under PT Indo Fin Tek. We are registered under Otoritas Jasa Keuangan P2P regulation. At Dompét Kilat, we helped the underserved to get funding access. We create an innovative, agile, and convenient digital lending platform and take part in stimulating financial inclusion for a better Indonesia. Come work with energetic and dynamic individuals in Dompét Kilat.

## Candidate Test

Please do not try to write your answer on this document by requesting edit access. You can make a copy and write your answer on the copy. Submit your answer along with required files. Good luck!

## Technical Test

### Dataset

In this dataset, each entry represents a person who takes a credit by a bank. Each person is classified as good or bad credit risks according to the set of attributes. The attributes are:

- Age (numeric)
- Sex (text: male, female)
- Job (numeric: 0 - unskilled and non-resident, 1 - unskilled and resident, 2 - skilled, 3 - highly skilled)
- Housing (text: own, rent, or free)
- Saving accounts (text - little, moderate, quite rich, rich)
- Checking account (numeric, in DM - Deutsch Mark)
- Credit amount (numeric, in DM)
- Duration (numeric, in month)
- Purpose (text: car, furniture/equipment, radio/TV, domestic appliances, repairs, education, business, vacation/others)
- Risk (Value target - Good or Bad Risk)

## Questions A

Based on that dataset in this folder, please create:

1. EDA to understand distribution and pattern
2. Customer segmentation
3. Credit risk prediction model

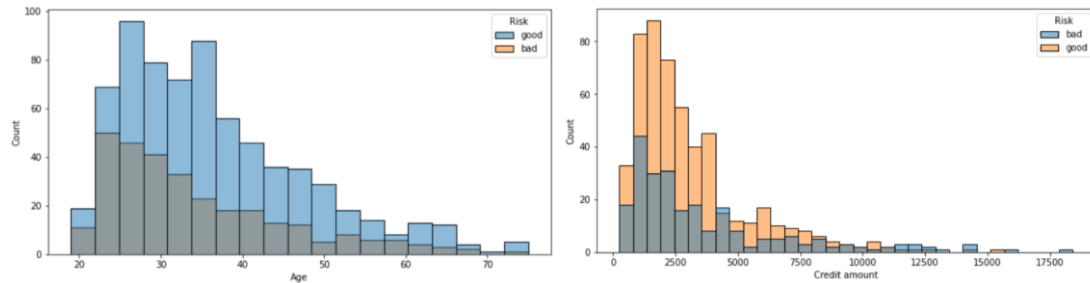
Please also provide:

1. Presentation of your findings and insights (the audience will be non-technical person)
2. Jupyter notebook explaining your approach to solve the problem for the technical audience

## ANSWER

### EKSPLORATORY DATA ANALYSIS

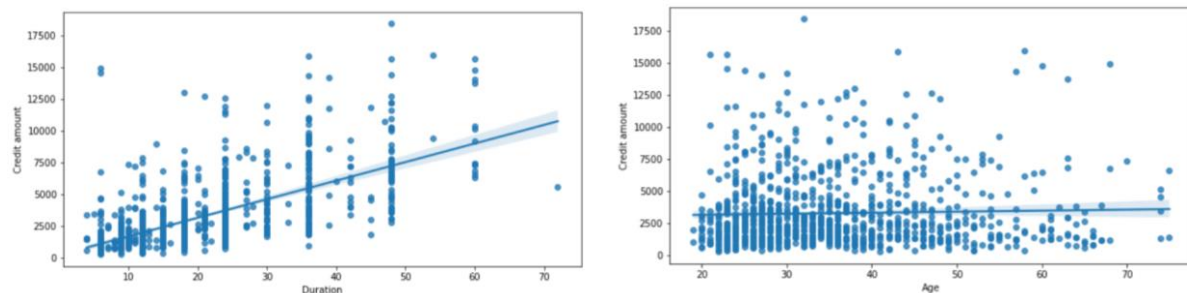
#### WHO ARE THE AVERAGE USER



Gambar 1. Plot Distribution About Age & Credit Amount

Berdasarkan Gambar 1, rata-rata user yang mengajukan kredit dari bank berusia 27 tahun sampai 42 tahun. Rata-rata user yang memiliki rentan usia tersebut mengajukan kredit sebesar Rp 2.333.000.000 dengan durasi pengembalian rata-rata 18 bulan. 69% dari user yang berada pada range usia tersebut diterima pengajuan kreditnya oleh bank.

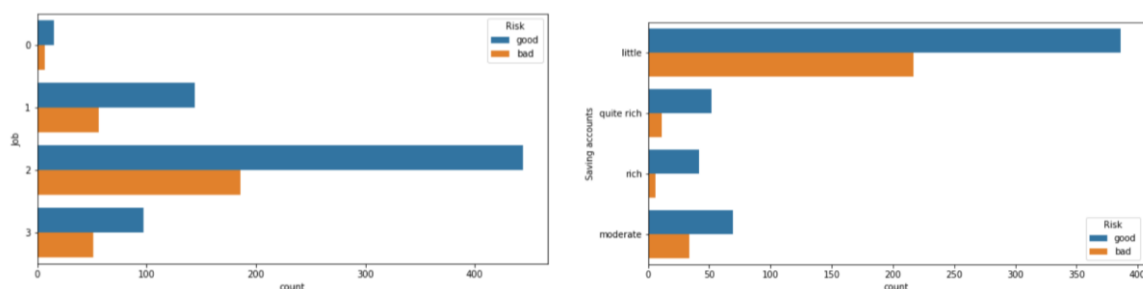
#### CORRELATION ABOUT CREDIT AMOUNT

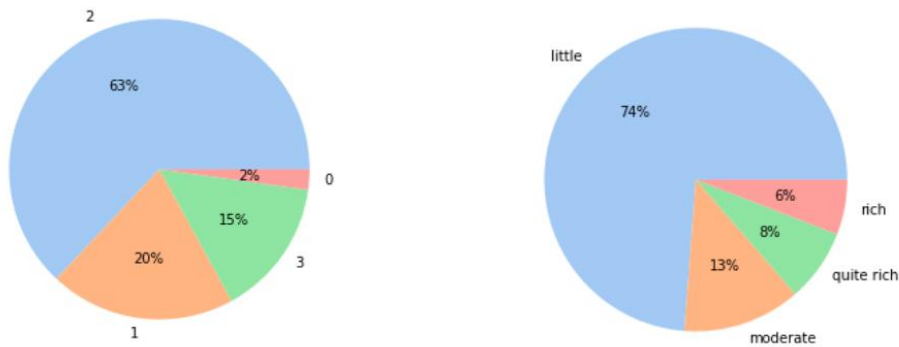


Gambar 2. Regression Plot About Credit Amount

Berdasarkan Gambar 2, terdapat hubungan antara total credit yang diajukan dengan durasi pengembalian, tetapi tidak ada hubungan antara total credit dengan umur user. Ini menunjukkan jika semakin banyak credit yang diajukan maka semakin banyak juga durasi pengembalian yang dibutuhkan.

**MY ASSUMPTION: GOOD USER FOR TAKE CREDIT HAVE JOB BUT NOT HAVE A LOT OF SAVING ACCOUNT. SO, LETS CHECK THE DATA.**





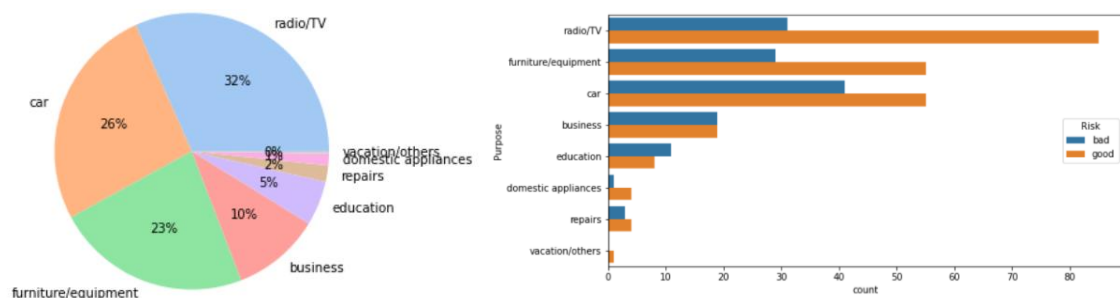
Gambar 3. Bar Chart About Job & Saving Account

Keterangan:

- 0 : unskilled and non-resident,
- 1 : unskilled and resident,
- 2 : skilled,
- 3 : highly skilled

Berdasarkan Gambar 3, user yang mengajukan kredit dari bank kebanyakan memiliki pengalaman dalam pekerjaannya tetapi memiliki tabungan yang sedikit.

**So, what do they need with the credit?**



Gambar 4. Presentase Tujuan Penggunaan Kredit

Berdasarkan Gambar 4, tipe user yang memiliki pengalaman dalam pekerjaan memiliki tabungan yang sedikit dengan tujuan pengajuan kredit yaitu untuk memenuhi kebutuhan radio/TV, mobil, dan furniture/equipment.

#### ACCEPTED PURPOSE & REJECTED PURPOSE

PURPOSE	RISK	PERCENTAGE
Business	Good	64%
	Bad	35%
Car	Good	68%
	Bad	31%
Domestic Appliances	Good	66%
	Bad	33%
Education	Good	61%
	Bad	38%
Furniture/Equipment	Good	67%
	Bad	32%

Radio/TV	Good	77%
	Bad	22%
Repairs	Good	63%
	Bad	36%
Vacation/others	Good	58%
	Bad	41%

Tabel 1. Percentage About Purpose by Risk

Berdasarkan Tabel 1, tujuan pengajuan kredit yang paling banyak diterima oleh bank adalah Radio/TV sedangkan tujuan pengambilan kredit yang paling banyak ditolak oleh bank adalah Vacation/others.

#### How many credits they take? Why they be accepted or rejected?

Category	Age	Job	Credit Amount	Duration
Average Good risk User by purpose Radio/TV	33	Skilled	1.589.000.000	15
Average Bad risk User by purpose Vacation/others	39	Highly skilled	11.938.000.000	24

Tabel 2. Compare between good risk user by Radio/TV and bad risk user by Vacation/others

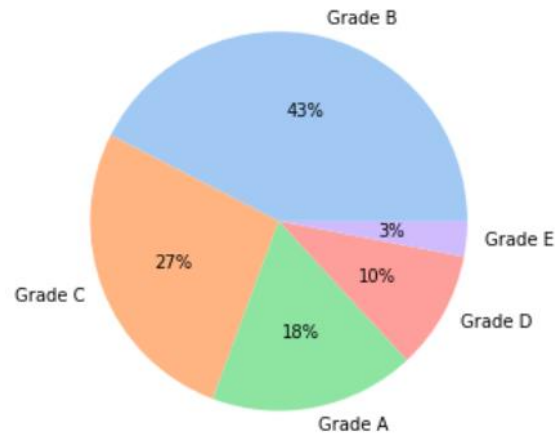
Berdasarkan Tabel 2, user yang pinjaman kreditnya ditolak oleh bank dikarenakan jumlah kredit yang diajukan cukup besar dengan tujuan pinjaman yaitu liburan/lainnya. User yang pinjaman kreditnya diterima oleh bank dikarenakan jumlah kredit yang diajukan sedikit dengan tujuan radio/TV.

## CONCLUTION

User potensial dari pengajuan kredit ini berusia 27 tahun sampai 42 tahun yang memiliki pengalaman dalam pekerjaannya dengan jumlah kredit yang diajukan sebesar Rp 2.333.000.000 dengan durasi pengembalian rata-rata 18 bulan. User yang mengajukan kredit merupakan orang yang ingin memenuhi kebutuhannya dalam hal radio/TV, mobil, dan furniture/equipment tetapi memiliki tabungan yang sedikit dalam akun banknya. Semakin besar pinjaman akan membuat durasi pengembalian semakin lama serta tujuan kredit yang tidak jelas membuat pihak bank menolak pengajuan kredit.

# USER SEGMENTATION

## USER SEGMENTATION BY PROBABILITY OF ACCEPTED CREDIT



Pengelompokkan user menggunakan metode regresi logistik dengan melihat peluang user diterima oleh bank, dengan terbagi menjadi 5 kelompok yaitu:

### Grade A (Very Good)

Name	Average
Age	47 tahun
Housing	Milik Sendiri
Saving Account	Cukup
Checking Account	Sedikit
Credit Amount	1.591.500.000
Duration	12 bulan
Purpose	Kebutuhan furniture

Tabel 2.1. Kelompok User dengan Grade A

Berdasarkan Tabel 2.1, dapat disimpulkan jika user dengan Grade A berusia 36 tahun sampai 56 tahun yang memiliki rumah pribadi dengan tabungan yang cukup serta pengajuan kredit sebesar Rp 1.591.500.000 dalam waktu pengembalian 12 bulan.

### Grade B (Good)

Name	Average
Age	34 tahun
Housing	Milik Sendiri
Saving Account	Sedikit
Checking Account	Sedikit
Credit Amount	1.919.000.000
Duration	15 bulan
Purpose	Kebutuhan furniture

Tabel 2.2. Kelompok User dengan Grade B

Berdasarkan Tabel 2.2, dapat disimpulkan jika user dengan Grade B berusia 28 tahun sampai 40 tahun yang memiliki rumah pribadi dengan tabungan yang sedikit serta pengajuan kredit sebesar Rp 1.919.000.000 dalam waktu pengembalian 15 bulan.

### Grade C (Normal)

Name	Average
Age	29 tahun
Housing	Milik Sendiri
Saving Account	Sedikit
Checking Account	Sedikit
Credit Amount	3.136.000.000
Duration	24 bulan
Purpose	Kebutuhan furniture

Tabel 2.3. Kelompok User dengan Grade C

Berdasarkan Tabel 2.3, dapat disimpulkan jika user dengan Grade C berusia 25 tahun sampai 36 tahun yang memiliki rumah pribadi dengan tabungan yang sedikit serta pengajuan kredit sebesar Rp 3.136.000.000 dalam waktu pengembalian 24 bulan.

#### Grade D (Bad)

Name	Average
Age	26 tahun
Housing	Sewa
Saving Account	Sedikit
Checking Account	Sedikit
Credit Amount	4.280.000.000
Duration	33 bulan
Purpose	Mobil

Tabel 2.4. Kelompok User dengan Grade D

Berdasarkan Tabel 2.4, dapat disimpulkan jika user dengan Grade D berusia 23 tahun sampai 31 tahun yang memiliki rumah sewa dengan tabungan yang sedikit serta pengajuan kredit sebesar Rp 4.280.000.000 dalam waktu pengembalian 33 bulan.

#### Grade E (Very Bad)

Name	Average
Age	27 tahun
Housing	Menumpang
Saving Account	Sedikit
Checking Account	Cukup
Credit Amount	7.578.500.000
Duration	48 bulan
Purpose	Mobil

Tabel 2.5. Kelompok User dengan Grade E

Berdasarkan Tabel 2.5, dapat disimpulkan jika user dengan Grade E berusia 24 tahun sampai 32 tahun yang memiliki rumah menumpang dengan tabungan yang sedikit serta pengajuan kredit sebesar Rp 7.578.500.000 dalam waktu pengembalian 48 bulan.

## CONCLUSION

- Pengajuan kredit akan semakin baik jika kredit yang diajukan kecil, sebaliknya pengajuan kredit akan semakin buruk jika kredit yang diajukan besar.

- Pengajuan kredit akan semakin baik jika durasi pengembalian cepat, sebaliknya pengajuan kredit akan semakin buruk jika durasi pengembalian lama.
- Pengajuan kredit akan semakin baik jika jumlah tabungan banyak, sebaliknya pengajuan kredit akan semakin buruk jika jumlah tabungan sedikit.

## CREDIT RISK PREDICTION MODEL

### ALL METHOD

Semua metode klasifikasi tersebut menggunakan data test sebesar 33% dari total data.

#### Logistic Regression

	precision	recall	f1-score	support
0	0.65	0.15	0.24	117
1	0.67	0.96	0.79	213
accuracy			0.67	330
macro avg	0.66	0.55	0.51	330
weighted avg	0.66	0.67	0.59	330

Accurasi yang terbentuk pada metode regresi logistik sebesar 79%.

#### Decision Tree

	precision	recall	f1-score	support
0	0.42	0.38	0.40	117
1	0.68	0.71	0.69	213
accuracy			0.59	330
macro avg	0.55	0.54	0.55	330
weighted avg	0.58	0.59	0.59	330

Accurasi yang terbentuk pada metode decision tree sebesar 59%.

#### Naïve Bayes

	precision	recall	f1-score	support
0	0.56	0.30	0.39	117
1	0.69	0.87	0.77	213
accuracy			0.67	330
macro avg	0.63	0.59	0.58	330
weighted avg	0.65	0.67	0.64	330

Accurasi yang terbentuk pada metode naïve bayes sebesar 67%

#### K-Nearest Neighbor

	precision	recall	f1-score	support
0	0.58	0.06	0.11	117
1	0.65	0.98	0.78	213
accuracy			0.65	330
macro avg	0.62	0.52	0.45	330
weighted avg	0.63	0.65	0.54	330

Accurasi yang terbentuk pada metode k-nearest neighbor sebesar 65%

**If we look at recall row 0, we know the best method for credit risk prediction model**

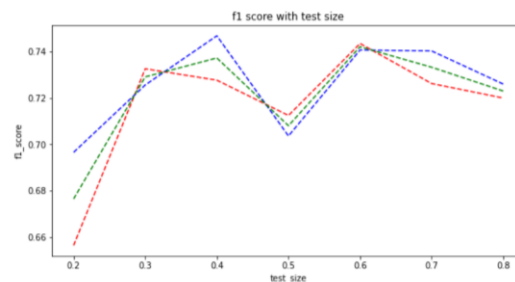


Recall pada baris 0 adalah perbandingan antara data prediksi yang ada di kelas “bad” dengan data yang seharusnya ada di kelas “bad”. Ini berarti metode logistic regression dan k-nearest neighbor tidak cukup baik dalam melakukan prediksi terhadap kelas “bad”. Maka model yang dipilih menggunakan metode decision tree dan naïve bayes.

## Decision Tree VS Naïve Bayes

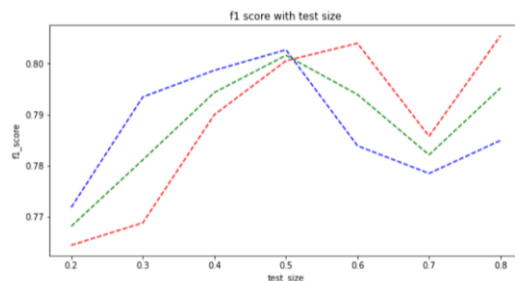
Model yang baik adalah model yang tidak overfit ataupun underfit, yang berarti model tersebut harus mampu beradaptasi dengan segala macam kondisi.

### Decision Tree



Minimum of Error	0.002797
Average of Error	0.013972
Maximum of Error	0.040056

### Naïve Bayes



Minimum of Error	0.002141
Average of Error	0.012935
Maximum of Error	0.024589

If we look at maximum of error, we know the best model is naïve bayes because error of naïve bayes is smaller than decision tree.

## CONCLUTION

Metode yang paling baik untuk data ini adalah Naïve Bayes untuk credit risk prediction model dengan tingkat akurasi 70%.

## Questions B

1. Can you tell us about your passion?
2. What is your goal in life?
3. If you join with Dompot Kilat as a Data Engineer, what is your expectation? Please be clear about this as we want to fully support your journey with us

## ANSWER

1. **I have passion about data and bussiness.** Passion data berawal ketika saya menjalani mata kuliah pemrograman seperti algoritma pemrograman, struktur data, perancangan analisis algoritma, dll. Saya sangat ambisius pada mata kuliah tersebut dan saya mencoba untuk mengajari rekan-rekan yang lain karena memang saya bisa. Dari mata kuliah tersebut saya mengenal data science, yang mana saya mencoba untuk belajar mandiri terkait machine learning, data visualisasi, data processing, dll. Passion bisnis berawal ketika saya menyadari bahwa data science bukan hanya tentang mengolah data tetapi tentang mengembangkan bisnis melalui data.
2. **I want to be a millioner with all my ability.** Passion yang saya miliki kemungkinan akan membawa saya menggapai posisi yang tinggi dalam suatu perusahaan/organisasi. Harapan yang saya impikan saat menggapai posisi tersebut adalah dapat meningkatkan kualitas hidup saya dan keluarga.
3. **I want to grow up.** Saat ini saya sedang mencari titik aman pertama dalam kehidupan setelah lulus kuliah, dimana saya pikir posisi data engineer di Dompot Kilat merupakan titik aman pertama yang bagus. Harapan saya jika diterima disini adalah saya harus belajar mengenai lebih banyak tools terkait data dan bisa meningkatkan skill saya dalam data analysis, data scientist, data engineer, dan bagaimana cara startup berkembang. Saya juga ingin menghilangkan kelemahan yang saya miliki seperti tidak lancar dalam bicara inggris dan ingin memperluas wawasan saya terkait bisnis serta menambah circle pertemanan.