

# Big Data

# Final Assignment



# Final Assignment #1

## Recency, Frequency,

## Monetary (RFM) Analysis



# Business Understanding

---



## Ruang Lingkup Bisnis : Perbankan & Ritel

Salah satu alat pembayaran yang umum digunakan oleh masyarakat saat ini adalah menggunakan kartu debit/kredit. Kemudahan pembayaran ini merupakan salah satu produk layanan yang disediakan oleh perbankan. Kartu debit/kredit dapat digunakan untuk segala macam transaksi pelanggan, salah satunya transaksi pembelanjaan pada toko ritel. Dataset yang digunakan pada analisis ini merupakan kumpulan data transaksi penggunaan kartu kredit salah satu bank ternama di Indonesia, dimana tujuan dari analisis ini adalah untuk mengetahui segmentasi pelanggan menggunakan RFM analisis berdasarkan data transaksi yang ada.





# Data Understanding

| Field Name | Description               | Data Type |
|------------|---------------------------|-----------|
| CardID     | <i>Nomor kartu kredit</i> | int64     |
| TrxDate    | <i>Tanggal transaksi</i>  | object    |
| Amount     | <i>Total transaksi</i>    | int64     |



# Data Preparation

## ◆ Menyesuaikan tipe data setiap kolom

Melakukan pengecekan tipe data pada semua kolom, dalam hal ini kolom TrxDate memiliki tipe data object. Tipe data pada kolom ini kemudian di-convert menjadi tipe data datetime.

## ◆ Mengecek jumlah baris dan kolom

Melakukan pengecekan jumlah baris dan kolom yang ada pada dataset. Dari proses ini dapat diketahui bahwa terdapat 199 data dan 3 kolom.

## ◆ Mengecek data dengan nilai null

Melakukan pengecekan data pada keseluruhan dataset, apakah terdapat data yang kosong/null. Apabila terdapat nilai null atau data kosong maka dapat dilakukan berbagai cara seperti, diisi dengan nilai 0 atau nilai mean/median, atau baris dengan data kosong tersebut dihapus dari dataset.

# Modeling

---

## ✓ Melakukan Grouping Data

Grouping data dilakukan untuk menghitung jumlah hari pada TrxDate sebagai nilai R, menghitung jumlah transaksi sebagai nilai F, dan menghitung total transaksi sebagai nilai M. Perhitungan menggunakan CardID sebagai aggregate.

## ✓ Mengecek Nilai pada Setiap Quantile

Melakukan perhitungan nilai quantile pada setiap kolom. Hasil perhitungan ini bertujuan untuk menentukan kelompok data pada setiap quantile.

## ✓ Menentukan Kelompok RFM

Setelah dilakukan grouping data maka akan didapatkan nilai R, F, dan M untuk setiap data CardID. CardID yang terdapat pada transaksi adalah sebanyak 71 data kartu.





# Evaluation

---

- Melakukan Kombinasi Data R, F, dan M

Melakukan penggabungan nilai R, F, dan M agar menjadi sebuah kombinasi untuk menentukan kelompok pelanggan. Kombinasi yang ditemukan adalah sebanyak 28 kombinasi.

- Menghitung Nilai RFM

Setiap kombinasi yang ditemukan kemudian dihitung jumlah datanya. Hal ini untuk mengetahui jumlah pelanggan pada setiap kombinasi data.

# Result

---

| Group             | Recency (R)    | Frequency (F) | Monetary (M)                |
|-------------------|----------------|---------------|-----------------------------|
| 4 - Tinggi        | 1 - 94 hari    | > 7 kali      | Rp 1.732.768 - Rp 4.640.126 |
| 3 - Sedang        | 95 - 168 hari  | 4 - 6 kali    | Rp 1.342.489 - Rp 1.732.767 |
| 2 - Rendah        | 169 - 202 hari | 3 kali        | Rp 854.568 - Rp 1.342.488   |
| 1 - Sangat Rendah | 203 - 315 hari | 1 - 2 kali    | Rp 1 - Rp 854.567           |



# Recomendation Campaign

---

- 1 - Sangat Rendah

Segmen ini menunjukan bahwa pelanggan memiliki nilai RFM sangat rendah, dimana pelanggan tersebut hanya 1 kali melakukan transaksi di toko Giant. Untuk itu, perusahaan dapat memberikan promosi berupa memberikan cicilan 0% untuk transaksi apa saja dan dimana saja.

Dari data yang telah dianalisis, customer yang termasuk dalam segmen ini adalah sebanyak 35 orang.

- 2 - Rendah

Segmen ini menunjukan bahwa pelanggan memiliki nilai RFM rendah, pelanggan tersebut mungkin melakukan transaksi 1-2 kali di toko, namun total transaksi pelanggan tersebut rendah dan cukup lama waktu terakhir pelanggan tersebut berbelanja. Untuk itu, perusahaan dapat memberikan promosi berupa voucher belanja yang dapat digunakan pada merchant yang telah ditunjuk

Dari data yang telah dianalisis, customer yang termasuk dalam segmen ini adalah sebanyak 14 orang.



# Recomendation Campaign

---

- 3 - Sedang

Segmen ini menunjukan bahwa pelanggan memiliki nilai RFM lebih baik dibanding 2 segmen sebelumnya, dimana pelanggan tersebut hanya memiliki frekuensi belanja lebih sering namun tingkat resensi dan total transaksi masih rendah.

Untuk itu, perusahaan dapat memberikan promosi berupa hadiah langsung apabila transaksi mencapai batas total transaksi yang telah ditentukan.

Dari data yang telah dianalisis, customer yang termasuk dalam segmen ini adalah sebanyak 5 orang.

- 4 - Tinggi

Segmen ini menunjukan bahwa pelanggan memiliki nilai RFM paling tinggi, pelanggan tersebut untk melakukan transaksi lebih sering dengan total transaksi dan tingkat resensi yang cukup tinggi. Untuk kelompok pelanggan ini tidak perlu dilakukan adanya promosi, namun perusahaan dapat memberikan apresiasi ucapan terima kasih dengan memberikan kemudahan, misalnya kasir khusus pemegang kartu kredit ini yang minim antrian panjang.

Dari data yang telah dianalisis, customer yang termasuk dalam segmen ini adalah sebanyak 11 orang.



# Final Assignment #2

## Principal Component

## Analysis + Multi Linear

## Regression



# Business Understanding

Ruang Lingkup Bisnis : Perbankan

KPR merupakan salah satu produk pembiayaan atau pinjaman yang diberikan kepada pembeli rumah dengan skema pembiayaan sampai dengan persentase tertentu dari harga rumah atau properti. Dataset yang digunakan pada analisis ini merupakan kumpulan data pengajuan limit kredit KPR salah satu bank ternama di Indonesia, dimana tujuan dari analisis ini adalah untuk mengetahui segmentasi pelanggan menggunakan PCA dan prediksi jumlah limit kredit yang dapat diberikan menggunakan analisis multi linear regression.





# Data Understanding

| Field Name            | Description                           | Data Type |
|-----------------------|---------------------------------------|-----------|
| Customer ID           | <i>Nomor ID pelanggan</i>             | int64     |
| Umur                  | <i>Usia pelanggan</i>                 | int64     |
| Income                | <i>Pendapatan pelanggan</i>           | int64     |
| Product_holding       | <i>Produk perbankan yang dimiliki</i> | int64     |
| Saving                | <i>Jumlah tabungan</i>                | int64     |
| Deposit               | <i>Jumlah deposito</i>                | int64     |
| KK                    | <i>Tagihan kartu kredit</i>           | int64     |
| Tab Bisnis            | <i>Tabungan bisnis</i>                | int64     |
| Limit Kredit Mortgage | <i>Limit kredit yang diberikan</i>    | int64     |



# Data Preparation

## ◆ Menyesuaikan tipe data setiap kolom

Melakukan pengecekan tipe data pada semua kolom, dalam hal ini kolom TrxDate memiliki tipe data object. Tipe data pada kolom ini kemudian di-convert menjadi tipe data datetime.

## ◆ Mengecek jumlah baris dan kolom

Melakukan pengecekan jumlah baris dan kolom yang ada pada dataset. Dari proses ini dapat diketahui bahwa terdapat 199 data dan 3 kolom.

## ◆ Mengecek data dengan nilai null

Melakukan pengecekan data pada keseluruhan dataset, apakah terdapat data yang kosong/null. Apabila terdapat nilai null atau data kosong maka dapat dilakukan berbagai cara seperti, diisi dengan nilai 0 atau nilai mean/median, atau baris dengan data kosong tersebut dihapus dari dataset.

# Modeling

---

## ✓ Melakukan Split Data

Dataset dibagi menjadi 2 bagian, yaitu data training dan data testing dengan komposisi data training 70% dan data testing 30%.

## ✓ Melakukan Normalisasi Data

Normalisasi data dilakukan terhadap data training menggunakan fungsi StandardScaler()

## ✓ Pemodelan menggunakan PCA dan LinearRegression

Data kemudian dimodelkan menggunakan fungsi PCA() dimana hasil pemodelan PCA ditentukan variabel dependen dan independen. Kemudian variabel tersebut dimodelkan menggunakan fungsi LinearRegression()





# Evaluation (Testing and Accuracy)

---

## • Menghitung Prediksi

Dalam menentukan prediksi, data testing digunakan dalam pemodelan yang telah dibuat menggunakan fungsi predict().

## • Menghitung Akurasi Pemodelan

Dari hasil prediksi menggunakan data testing 30%, maka menghasilkan nilai akurasi sebesar -4.05936.

# Final Assignment #3

## Support Vector Machine



# Business Understanding

---



## Ruang Lingkup Bisnis : Perbankan

KPR merupakan salah satu produk pembiayaan atau pinjaman yang diberikan kepada pembeli rumah dengan skema pembiayaan sampai dengan persentase tertentu dari harga rumah atau properti. Dataset yang digunakan pada analisis ini merupakan kumpulan data pengajuan percepatan pelunasan kredit pada salah satu bank ternama di Indonesia, dimana tujuan dari analisis ini adalah untuk mengetahui apakah nasabah dimaksud dapat disetujui atau tidak untuk melakukan percepatan pelunasan KPR dengan menggunakan analisis Support Vector Machine (SVM).





# Data Understanding

| Field Name  | Description                                 | Data Type |
|-------------|---|-----------|
| Customer ID | <i>Nomor ID nasabah</i>                     | int64     |
| Age         | <i>Usia nasabah</i>                         | int64     |
| Income      | <i>Kategori pendapatan nasabah</i>          | int64     |
| Education   | <i>Kategori pendidikan nasabah</i>          | int64     |
| Occupation  | <i>Kategori bidang pekerjaan nasabah</i>    | int64     |
| Income.1    | <i>Kategori pendapatan tambahan nasabah</i> | int64     |



# Data Understanding

| Field Name       | Description                       | Data Type |
|------------------|-----------------------------------|-----------|
| Average Balance  | <i>Sisa saldo kredit nasabah</i>  | int64     |
| Appraisal        |                                   | int64     |
| Industri         |                                   | int64     |
| Pelunasan Kredit | <i>Pengajuan pelunasan kredit</i> | int64     |



# Data Preparation

## ◆ Menyesuaikan tipe data setiap kolom

Melakukan pengecekan tipe data pada semua kolom, dalam hal ini kolom TrxDate memiliki tipe data object. Tipe data pada kolom ini kemudian di-convert menjadi tipe data datetime.

## ◆ Mengecek jumlah baris dan kolom

Melakukan pengecekan jumlah baris dan kolom yang ada pada dataset. Dari proses ini dapat diketahui bahwa terdapat 199 data dan 3 kolom.

## ◆ Mengecek data dengan nilai null

Melakukan pengecekan data pada keseluruhan dataset, apakah terdapat data yang kosong/null. Apabila terdapat nilai null atau data kosong maka dapat dilakukan berbagai cara seperti, diisi dengan nilai 0 atau nilai mean/median, atau baris dengan data kosong tersebut dihapus dari dataset.

# Modeling

---

## ✓ Melakukan Split Data

Dataset dibagi menjadi 2 bagian, yaitu data training dan data testing dengan komposisi data training 70% dan data testing 30%.

## ✓ Melakukan Normalisasi Data

Normalisasi data dilakukan terhadap variabel independen pada data training menggunakan fungsi StandardScaler()

## ✓ Pemodelan Menggunakan SVC

Data dimodelkan menggunakan fungsi SVC dari library SVM. Pemodelan dilakukan terhadap data training 70%.





# Evaluation (Testing and Accuracy)

---

## • Menghitung Prediksi

Dalam menentukan prediksi, data testing digunakan dalam pemodelan yang telah dibuat menggunakan fungsi predict().

## • Menghitung Akurasi Pemodelan

Dari hasil prediksi menggunakan data testing 30%, maka menghasilkan nilai akurasi sebesar 0.497 ~ 49.7%

```
3   require File.expand_path("../config/environment", __FILE__)
4   # Prevent database truncation if the schema has changed
5   abort("The Rails environment is running in production mode") if Rails.env.production?
6   require 'spec_helper'
7   require 'rspec/rails'
8
9   require 'capybara/rspec'
10  require 'capybara/rails'
11
12  Capybara.javascript_driver = :webkit
13  Category.delete_all; Category.create!(name: "Electronics")
14  Shoulda::Matchers.configure do |config|
15    config.integrate do |with|
16      with.test_framework :rspec
17      with.library :rails
18    end
19  end
20
21  # Add additional requires below this line
22
23  # Requires supporting files within the "spec/support/" directory
24  # in addition to the rails ones
25  # run as spec files by default
26  # in _spec.rb will both be
27  # run twice. It is recommended
28  # end with _spec.rb. You can
29  # opt out on the command line.
30  # No results found for 'mongoid'
```

# Deployment

---

Tahap ini dilakukan menggunakan bahasa pemrograman Python dengan IDE Kaggle.

Code secara keseluruhan terdapat pada link Github :

<https://github.com/dyanaagustina/Learn-BigData/tree/Big-Data/UAS>

# Thank You

---

