

PERANCANGAN *BUSINESS INTELLIGENCE* DAN SEGMENTASI PELANGGAN MENGGUNAKAN *K MEANS CLUSTERING* BERDASARKAN *RFM MODEL*

Ahmad Karami*, Yusuf Widharto

Departemen Teknik Industri, Fakultas Teknik, Universitas Diponegoro,
Jl. Prof. Soedarto, SH, Kampus Undip Tembalang, Semarang, Indonesia 50275

Abstrak

PT XYZ adalah perusahaan penyedia solusi otomasi, informasi, dan pengukuran untuk industri terkemuka yang menggabungkan teknologi unggul dengan layanan teknik, manajemen proyek, konsultasi, dan pemeliharaan di seluruh wilayah Indonesia. Dengan cakupan dan jumlah pelanggan yang cukup banyak, PT XYZ berusaha untuk membuat sebuah sistem informasi yang dapat memudahkan dalam memantau bagaimana kontribusi dari pelanggan terhadap bisnis perusahaan. Namun terdapat berbagai tantangan untuk pengimplentasiannya, yang salah satunya adalah banyaknya jumlah jenis produk yang disediakan oleh perusahaan sehingga menambah kompleksitas dalam pengembangannya. Oleh karena itu diperlukan sebuah kerangka acuan yang tepat dalam pengembangannya, seperti Business Intelligence. Dengan mengaplikasikan cakupan dari Business Intelligence yaitu Data Mining yang dapat mengubah data menjadi informasi yang berharga bagi perusahaan, melalui salah satu teknik Data Mining yaitu K Means Clustering perusahaan dapat mengelompokkan pelanggan berdasarkan suatu kriteria tertentu. Kriteria yang dipilih adalah Model RFM yang terkenal mampu mendeskripsikan nilai dari suatu pelanggan. Dengan menggunakan excel sebagai platformnya dan python sebagai bahasa pemrogramannya, rancangan sistem informasi business intelligence ini didapatkan efektif membantu perusahaan dalam memahami kontribusi dari pelanggan.

Kata kunci: *business intelligence, data mining, k means clustering, model rfm, sistem informasi*

Abstract

PT XYZ is a leading industrial automation, information, and measurement solution provider that combines superior technology with engineering, project management, consulting, and maintenance services throughout Indonesia. With very high coverage and the number of customers, PT XYZ strives to create an information system that can make it easier to monitor how customers contribute to the company's business. However, there are various challenges for its implementation, one of which is a large number of types of products provided by the company, thus adding to the complexity of its development. Therefore we need a correct framework for the development, such as Business Intelligence. By applying the scope of Business Intelligence namely Data Mining, which can turn data into valuable information for the company, and through one of the Data Mining techniques namely K Means Clustering, companies can group customers based on certain criteria. The criteria chosen is the famous RFM Model which is able to describe the value of a customer. By using excel as the platform and python as the programming language, the design of this business intelligence information system was found to be effective in helping companies understand contributions from customers.

Keywords: *business intelligence, data mining, information system, k means clustering, rfm model*

1. Pendahuluan

PT XYZ adalah perusahaan penyedia solusi otomasi, informasi, dan pengukuran untuk industri terkemuka yang menggabungkan teknologi unggul dengan layanan teknik, manajemen proyek, konsultasi, dan pemeliharaan di seluruh wilayah

Indonesia. Dengan cakupan dan jumlah pelanggan yang banyak, PT XYZ berusaha untuk membuat sebuah sistem informasi yang dapat memudahkan perusahaan dalam memantau bagaimana kontribusi dari pelanggan terhadap bisnis perusahaan. Namun terdapat berbagai tantangan untuk pengimplentasiannya, yang salah satunya adalah banyaknya jumlah jenis produk yang disediakan oleh perusahaan sehingga menambah kompleksitas dalam pengembangannya. Oleh karena itu diperlukan sebuah

*Ahmad Karami
E-mail: ahmadkarami@students.undip.ac.id

kerangka acuan yang tepat dalam pengembangannya, seperti *Business Intelligence*.

Business intelligence (BI) adalah alat analisis yang digunakan untuk mengkonsolidasi data, menganalisis, menyimpan dan mengakses banyak data untuk membantu dalam pembuatan keputusan, seperti perangkat lunak untuk *query* database dan pelaporan alat untuk analisis data multidimensi, dan data mining. (Laudon, 2017). Manfaat dari *Business Intelligence* diantaranya meningkatkan nilai data dan informasi organisasi, memudahkan pengukuran kinerja organisasi, meningkatkan nilai investasi teknologi informasi yang sudah ada, dan meningkatkan efisiensi biaya (Turban. E et al., 2011).

Salah satu metrik yang dapat digunakan dalam mengukur karakteristik pelanggan Model RFM. Model RFM terdiri dari tiga kriteria utama yaitu *Recency*, *Frequency*, dan *Monetary*. *Recency* menggambarkan rentang pembelian terakhir pelanggan hingga saat ini, *Frequency* menggambarkan banyaknya frekuensi pembelian untuk rentang periode tertentu, dan *Monetary* menggambarkan banyaknya total nilai pembelian dari pelanggan untuk rentang periode tertentu (Tsiptsis & Chorionopoulos, 2010).

Selain menggunakan kriteria Model RFM, juga diperlukan metode yang dapat mengolah dan mengelompokkan pelanggan ke dalam grup-grup yang memiliki karakteristik yang sama. Untuk hal ini digunakan metode *clustering*. K-Means merupakan salah satu algoritma clustering dengan metode partisi (partitioning method) yang berbasis titik pusat (centroid) (Jain, 2010). Oleh karena itu, dengan menggunakan *Clustering* sebagai teknik pengolahan data dan Model RFM sebagai dasar kriterianya, maka informasi mengenai karakteristik dari pelanggan akan didapatkan.

2. Metodologi Penelitian

Berikut merupakan metode pengumpulan dan pengolahan data di PT XYZ:

1. Pengumpulan Data

Pada tahap ini dilakukan proses pengumpulan data yang berkaitan dengan topik penelitian yang dibahas. Data yang akan dikumpulkan berupa data *sales* dari tahun 2018 sampai 2020.

2. Pengolahan Data

Pengolahan data dilakukan ke dalam tiga tahapan utama yaitu:

a. *Preprocessing Data*

Pada tahap ini sebelum data akan diproses, data tersebut harus dilakukan beberapa perlakuan terlebih dahulu seperti pengintegrasian dan kombinasi data, transformasi data, reduksi data, dan pembersihan data agar data menjadi siap untuk diolah.

b. *Processing Data*

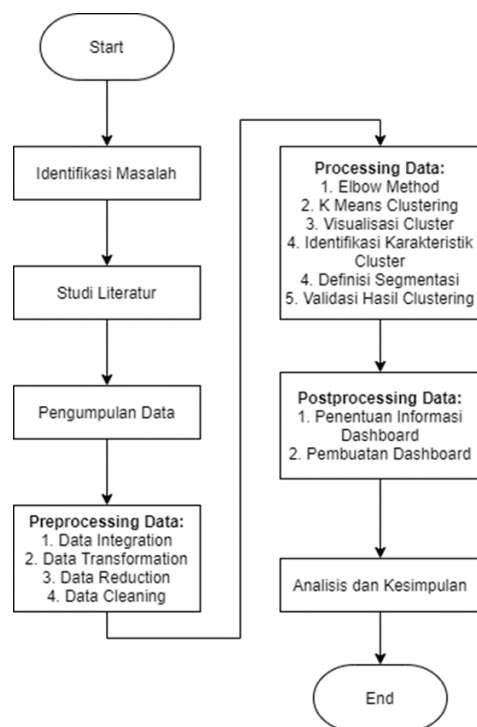
Pada tahap ini data dilakukan *data mining* atau penggalian informasi sehingga mendapatkan sebuah pola yang dapat menjadi pengetahuan bagi perusahaan. Pola yang akan dicari adalah

segmentasi pelanggan. Teknik *data mining* yang digunakan adalah K Means Clustering.

c. *Post Processing Data*

Pada tahap ini data yang sudah diproses diintegrasikan ke dalam sebuah *dashboard* sebagai visualisasi hasil proses. Untuk itu pertama-tama perlu dilakukan penentuan informasi apa saja yang ingin ditampilkan dalam *dashboard*. Setelah itu dilakukan perancangan tampilan dan pembuatan *dashboard*.

Berikut merupakan keseluruhan metodologi penelitian pada Gambar 1.



Gambar 1. Metodologi penelitian

2.1. Analisis Pemilihan Metode dan Tools

Berikut merupakan analisis pemilihan metode beserta *tools* yang akan digunakan dalam pembuatan sistem informasi *business intelligence* ini.

1. Platform rancangan sistem informasi keseluruhan

Platform yang akan dipilih adalah excel, hal ini dikarenakan data mentah berasal dari hasil export dari sistem ERP SAP ke excel. Alasan utama lain dalam pemilihan platform ini dibanding platform lain adalah excel pada umumnya sudah otomatis terinstall pada beberapa PC pengguna sehingga tidak perlu investasi tambahan dalam penggunaannya, excel juga saat ini digunakan oleh sebagian besar tim PT XYZ.

2. Metrik penilaian segmentasi pelanggan

Metrik yang digunakan untuk penilaian segmentasi pelanggan adalah RFM Model. RFM model dipilih dalam penelitian ini

dikarenakan beberapa kelebihan yang dimiliki oleh RFM model adalah sangat efektif dalam menentukan pelanggan yang berharga dan juga sangat memerlukan sedikit variabel (Kahan, 1998; Kaymak, 2001; Miglautsch, 2000) dan memudahkan pengambilan keputusan dengan memprediksi respons dari pelanggan (McCarty & Hastak, 2007).

3. Metode dan *tools* dalam *preprocessing data*

Dalam tahapan *Data Processing*, data input ditransformasikan kedalam format yang sesuai untuk dianalisa. Tahap ini dilakukan proses penggabungan data dari berbagai sumber, pembersihan data agar untuk menghilangkan noise data dan data ganda, dan memilih atribut data untuk proses data mining (Larose, 2005).

Pada tahapan *preprocessing data*, *tools* yang akan digunakan adalah *excel power query*, hal ini dikarenakan fitur-fitur dalam *excel power query* sangat cocok dengan tahapan ini yaitu terdiri dari empat fitur utama yaitu *connect*, *transform*, *combine*, dan *share*.

4. Metode dan *tools* dalam *processing data*

Untuk *processing data* digunakan teknik *data mining*. *Data Mining* adalah penambangan informasi dengan mencari pola atau aturan dari data yang sangat besar (Beynon & Davies, 2004).

Pada *processing data*, teknik *data mining* yang digunakan adalah *clustering*. *Clustering* digunakan untuk mengidentifikasi objek data yang mirip satu sama lain. Kemampuan tersebut sangat cocok dalam tujuan utama di penelitian ini yaitu untuk menentukan segmentasi pelanggan berdasarkan karakteristik model RFM.

Metode yang digunakan dalam menentukan jumlah cluster yang optimal adalah *Elbow Method*. *Elbow Method* menurut (Kodinariya & Makwana, 2013) adalah metode visual yang digunakan untuk menentukan jumlah cluster yang tepat pada sebuah dataset

dengan cara kerja tahap awal memulai $k=2$ dan akan terus bertambah setiap step selagi menghitung nilai variabel. Grafik yang dihasilkan akan turun secara drastis sebelum akhirnya menjadi stabil dan disitulah nilai k yang optimum.

Metode untuk memvalidasi yang akan digunakan adalah Davies Bouldin Index. Indeks validitas Davies-Bouldin (DB) menghitung rata-rata nilai setiap titik pada himpunan data. Perhitungan nilai setiap titik adalah jumlah nilai *compactness* yang dibagi dengan jarak antara kedua titik pusat kluster sebagai *separation*. Jumlah kluster terbaik ditunjukkan dengan nilai DB yang semakin kecil (Davies & Bouldin, 1979).

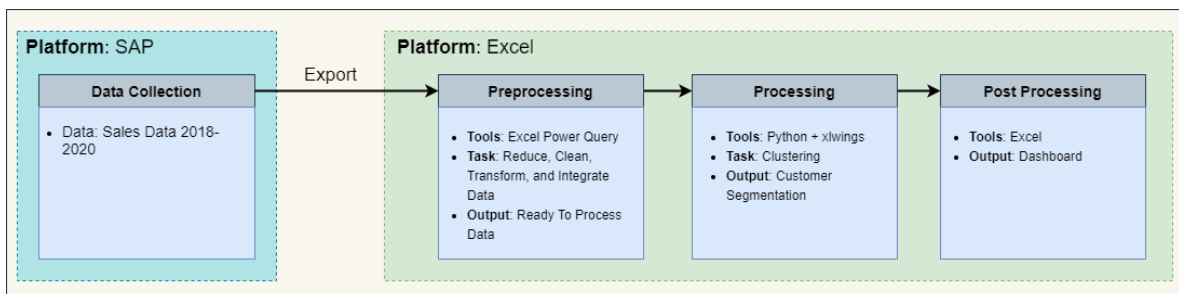
Tools yang akan digunakan untuk melakukan *clustering* adalah bahasa pemrograman Python. Python adalah bahasa pemrograman berorientasi objek yang diinterpretasikan yang mendapatkan popularitas di bidang ilmu data dan analitik. Python memiliki *library* standar yang sangat besar dan kuat yang dapat digunakan untuk menganalisis dan memvisualisasikan data (Butwall et al., 2019).

5. Metode dan *tools* dalam *postprocessing data*

Pada tahap *postprocessing data*, *tools* yang digunakan adalah *excel* berupa sebuah *dashboard*. *Dashboard* merupakan alat untuk menampilkan informasi dalam bentuk tabel, laporan serta indikator visual dan informasi tersebut akan ditampilkan secara visual di layar dengan cara yang memudahkan pengguna dalam mengawasi dan mengelola inisiatif bisnisnya (Wajong, 2015).

2.2. Rancangan Sistem Informasi *Business Intelligence*

Berikut ini merupakan skema dari rancangan sistem informasi *business intelligence* yang akan dibuat bisa dilihat pada Gambar 2.



Gambar 2. Skema sistem informasi *business intelligence*

3. Hasil dan Pembahasan

Pada bab ini menjelaskan bagaimana pengumpulan data dan pengolahan data dilakukan.

3.1. Raw data

Raw data dalam penelitian ini adalah data sales untuk tahun 2018 sampai tahun 2020. Data diexport dari database ERP SAP ke dalam bentuk excel. File terbagi menjadi tiga file yaitu untuk tahun 2018, 2019, dan 2020. Berikut merupakan atribut-atribut yang ada dalam raw data pada Gambar 3.



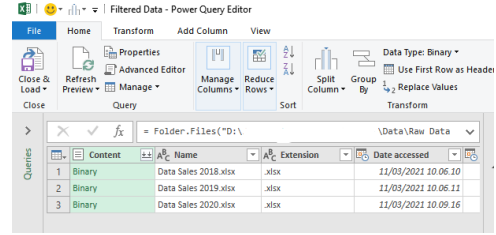
Gambar 3. Daftar atribut raw data

3.2. Preprocessing data

Pada tahapan ini dilakukan empat langkah utama dalam preprocessing data yaitu data integration, data transformation, data reduction, dan data cleaning.

1. Data integration

Tahap pertama dalam data integration adalah mengintegrasikan file raw data ke dalam satu file excel untuk dikombinasikan. Hal ini dapat dilakukan dengan menggunakan fitur dari excel power query. Power query memiliki fitur yang bisa menghubungkan file-file yang ingin digabungkan dengan cara menentukan folder yang menjadi sumbernya. Berikut merupakan hasil referensi dari file yang akan diolah pada Gambar 4.



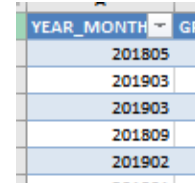
Gambar 4. Daftar atribut raw data

2. Data transformation

Pada tahapan ini data ditransformasi agar sesuai dengan metode pengolahan yang ingin dilakukan. Untuk itu maka diperlukan untuk menentukan data-data apa saja yang diperlukan. Hal tersebut dapat dilakukan dengan melihat metrik dari model RFM di awal yaitu recency, frequency, dan monetary. Berikut merupakan penjabaran dari atribut-atribut yang dibutuhkan.

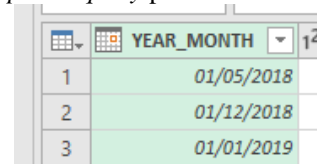
a) Recency

Recency adalah jarak waktu antara pembelian terakhir dari pelanggan dengan waktu saat ini. Atribut yang dapat menjelaskan metrik tersebut adalah tanggal dari pembelian, dalam kasus ini yaitu atribut tanggal pembelian dapat dilihat pada Gambar 5.



Gambar 5. Atribut tanggal pembelian

Data tersebut tetapi masih belum bisa dijadikan data untuk diproses dikarenakan format data tersebut masih berupa urutan angka, oleh karena itu perlu dilakukan transformasi data menjadi format tanggal. Berikut merupakan hasil dari perbaikan format ini menggunakan excel power query pada Gambar 6.



Gambar 6. Perbaikan format atribut tanggal pembelian

b) Frequency

Untuk menghitung jumlah frekuensi pembelian dari customer, diperlukan sebuah atribut yang dapat dijadikan sebagai identitas dari customer. Dalam kasus ini atribut yang cocok adalah nama perusahaan pembeli.

c) Monetary

Untuk menghitung jumlah nilai pembelian dari customer, diperlukan sebuah atribut yang merepresentasikan

nilai pembelian. Dalam kasus ini yaitu nilai mata uang dalam satuan dollar sehingga atribut yang sesuai adalah nilai pembelian.

d) Atribut tambahan

Dalam data, tersedia beberapa atribut yang bisa membantu dalam memetakan kontribusi dari pelanggan, atribut-atribut tersebut antara lain tipe proyek, *sales office*, segmentasi industri, jenis industri dan tahun fiskal.

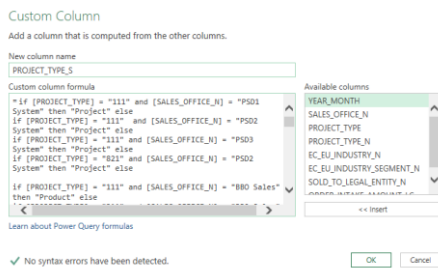
1) Tipe proyek

Dengan mengetahui tipe proyek pada pembelian, perusahaan dapat mengetahui seberapa besar kebutuhan pelanggan akan proyek tertentu beserta *value* dari tipe proyek tersebut. Dalam kasus ini yaitu kolom PROJECT_TYPE dan PROJECT_TYPE_N bisa dilihat pada Gambar 7.

PROJECT_TYPE	PROJECT_TYPE_N
999	Product M
999	Product M
999	Product M
111	Product I
999	Product M
999	Product M
999	Product M

Gambar 7. Data atribut tipe proyek

PT XYZ ingin membagi tipe proyek ke dalam bentuk yang lebih sederhana lagi yaitu menjadi empat tipe proyek diantaranya *product*, *project*, *service*, dan *solution*. Dalam *raw data* belum ada kolom tersebut maka diperlukan pembuatan kolom baru, pembuatan kolom baru ini bisa dilakukan dengan menggunakan *power query*. Berikut merupakan logika penentuan tipe proyek disederhanakan dengan *power query* pada Gambar 8.



Gambar 8. Logika tipe proyek disederhanakan dengan *power query*

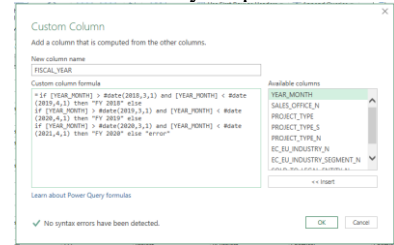
2) Segmentasi industri

Dengan mengetahui segmen industri pelanggan pada pembelian tertentu, perusahaan dapat menganalisis seberapa besar kontribusi untuk masing-masing segmen industri terhadap perusahaan.

3) Tahun Fiskal

Dapat dilihat bahwa di dalam *raw data* hanya ada atribut tanggal pembelian. Atribut tersebut belum bisa

menggambarkan apakah suatu *field* masuk ke dalam tahun fiskal ke berapa. Oleh karena itu diperlukan penambahan kolom baru sebagai atribut *fiscal year*. Berikut merupakan logika dari kolom untuk atribut *fiscal year* pada Gambar 9.



Gambar 9. Logika atribut *fiscal year*

3. Data reduction

Pada tahap ini atribut-atribut yang tidak dibutuhkan (selain yang disebutkan dalam rekapitulasi sebelumnya) maka akan dikeluarkan. Berikut merupakan hasil dari reduksi atribut data yang dapat dilihat dalam Gambar 10.



Gambar 10. Hasil reduksi atribut

4. Data cleaning

Pada tahap ini dilakukan perbaikan format dari data maupun perubahan-perubahan lain yang dapat memudahkan proses dan juga pembacaan dari data. Berikut merupakan langkah-langkah yang dilakukan pada tahap ini.

a) Pengeluaran data duplikat

Data duplikat merupakan data yang memiliki *value* yang sama di semua kolom. Hal bisa menyebabkan keanehan dalam pemrosesan, contohnya dalam kasus ini apabila terdapat duplikat dalam data tertentu maka proses akan menghitungnya menjadi dua kali frekuensi. Sehingga untuk menghilangkan hal tersebut data duplikat harus dikeluarkan.

b) Perbaikan format data

Di dalam *dataset* terdapat banyak format data yang belum sesuai atau belum siap untuk dilakukan pengolahan, berikut merupakan perbaikan-perbaikan yang dilakukan. Di dalam *excel power query* terdapat fitur untuk mendeteksi dan melakukan penyesuaian format data, setelah fitur tersebut dilakukan semua data menjadi sudah sesuai tetapi ada satu atribut yang masih belum sesuai yaitu

order intake, atribut tersebut masih berupa *numeric* oleh karena itu dilakukan perubahan menjadi *currency*.

3.3. Processing data

Pada tahap ini dilakukan pencarian pola segmentasi pelanggan dari dataset yang sudah dilakukan *preprocessing* dengan menggunakan teknik *clustering*. Berikut merupakan *flowchart* dari tahapan *processing data* pada Gambar 11.

Pengembangan *code* untuk tahapan *processing data* dilakukan terlebih dahulu dalam *software Pycharm*. Dalam *code*, pertama-tama dilakukan *import package* yang nantinya dibutuhkan. Berikut merupakan daftar *package* yang diimpor, dapat dilihat pada Tabel 1.

Tabel 1. Daftar *package* yang digunakan

<i>Package</i>	Keterangan kegunaan
pandas	Aggregasi data
math	Alat kalkulasi
datetime	Membantu menghitung metrik <i>recency</i>
numpy	Memberikan skor RFM
sklearn	- Algoritma <i>K Means Clustering</i> - Algoritma <i>Davies Bouldin Index</i>
seaborn	Visualisasi plot data
matplotlib	Plot data
xlwings	Integrasi <i>Excel</i> dengan <i>Python</i>

1. Referensi, Agregasi, dan Scoring Data

Pertama-tama dilakukan referensi terlebih dahulu terhadap *file excel* yang sudah dilakukan *preprocessing*. Referensi file dilakukan dua kali yaitu pertama untuk *package xlwings* dan yang kedua untuk *package pandas*.

Selanjutnya dilakukan agregasi data untuk metrik model RFM. Atribut yang digunakan adalah tanggal pembelian, nama perusahaan pembeli, dan nilai pembelian.

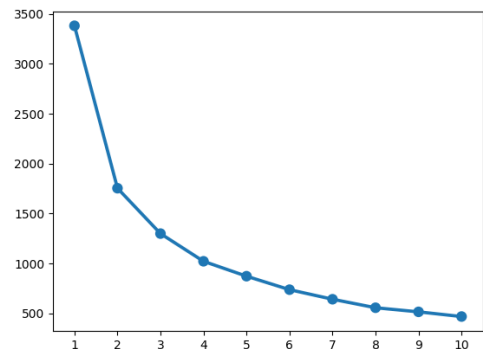
Tahapan selanjutnya adalah *scoring data*, *scoring* dilakukan dengan membagi masing-masing metrik tersebut menjadi 4 kuartil, maka skor terkecil yang bisa didapatkan oleh suatu metrik adalah 1 dan skor terbesar adalah 4. *Scoring* ini dilakukan dengan menggunakan *package* dari *numpy*.

Berikut merupakan output hasil dari proses-proses diatas sebelumnya hingga saat ini dapat dilihat dalam Gambar 12.

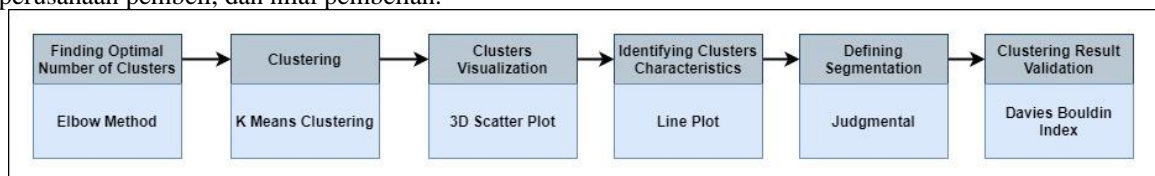
2. Menentukan jumlah *cluster* yang optimal

Setelah data dilakukan agregasi dan *scoring*, langkah selanjutnya adalah menentukan jumlah *cluster* yang optimal dengan menggunakan *elbow method* dengan menggunakan *package sklearn* dan digunakan *package seaborn* untuk memvisualisasikan hasil dari *elbow method* tersebut yang bisa dilihat pada Gambar 13.

Dapat dilihat bahwa $k=4$ merupakan jumlah *cluster* yang optimal hal ini dikarenakan setelah titik tersebut kemampuan data dalam mengelompokkan dan mendeskripsikan data tidak terjadi perubahan yang terlalu signifikan sehingga nilai k lebih dari 4 tidak lebih berharga karena kompleksitas juga meningkat.



Gambar 13. Visualisasi hasil *elbow method*



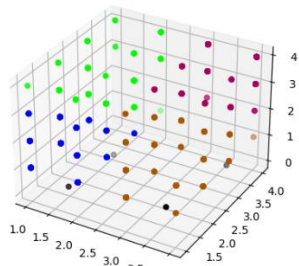
Gambar 11. *Data processing flowchart*

	Recency	Frequency	Monetary	r_score	f_score	m_score
SOLD_TO_LEGAL_ENTITY_N						
R664	11	2	-24520	2	2	1
G227	3	1	-1510	4	2	1
J349	11	1	-43	2	1	1
P594	1	4	-4	4	3	1
P558	10	6	-3	2	4	1

Gambar 12. Output agregasi dan scoring data

3. K means clustering

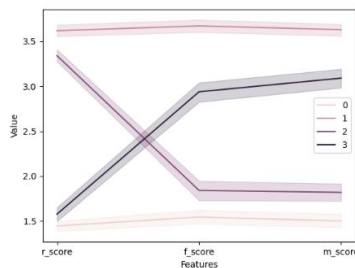
Setelah didapatkan jumlah cluster yang optimal maka langkah selanjutnya adalah melakukan *clustering*. K Means Clustering dapat dilakukan menggunakan *package* sklearn. Setelah dilakukan *clustering* terhadap data maka langkah selanjutnya adalah membuat visualisasi dari hasil cluster tersebut. Pertama-tama ditambahkan variabel *centers* sebagai pusat dari masing-masing *cluster*. Setelah itu ditambahkan variabel *ax* yaitu axis yang nantinya digunakan untuk *plotting*. Langkah terakhir adalah menyebarkan hasil *clustering* terhadap axis proyeksi 3D tersebut untuk masing-masing metrik. Setelah proses tersebut dilakukan maka selanjutnya adalah memvisualisasikannya dengan menggunakan *package* matplotlib yang dapat dilihat pada Gambar 14.



Gambar 14. Visualisasi hasil *clustering*

4. Identifikasi karakteristik dari *cluster*

Tahapan selanjutnya adalah mengidentifikasi karakteristik untuk tiap metrik (*recency*, *frequency*, dan *monetary*) dari tiap *cluster*. Untuk itu digunakan line plot sebagai visualisasi. Hal tersebut dapat dilakukan dengan menggunakan *package* seaborn. Berikut merupakan *output* visualisasi untuk tahap ini yang dapat dilihat pada Gambar 15.



Gambar 15. Visualisasi karakteristik *cluster*

Tabel 2. Rekapitulasi pendefinisian segmentasi

Rank	Cluster	Segmen	Recency	Frequency	Monetary
1	1	High Value Customer	High	High	High
2	2	New Customer	High	Low	Low-High
3	3	Almost Lost Customer	Low	Low-High	Low-High
4	0	Low Value Customer	Low	Low	Low

5. Pendefinisian segmentasi

Setelah karakteristik *cluster* sudah didapatkan maka langkah selanjutnya adalah melakukan pendefinisian segmentasi yang sesuai dengan karakteristik tersebut. Berikut merupakan hasil segmentasi dengan cara *judgemental*.

a) Cluster 0

Cluster 0 memiliki karakteristik bahwa semua nilai metriknya adalah rendah, sehingga *cluster* ini disegmentasi sebagai *low value customer*.

b) Cluster 1

Cluster 1 memiliki karakteristik bahwa semua nilai metriknya adalah tinggi, sehingga *cluster* ini disegmentasi sebagai *high value customer*.

c) Cluster 2

Cluster 1 memiliki karakteristik bahwa nilai *recency* yang tinggi tetapi memiliki nilai *frequency* yang rendah. Oleh karena itu *cluster* ini didefinisikan sebagai pelanggan baru atau *new customer*.

d) Cluster 3

Cluster 3 memiliki karakteristik bahwa nilai *recency* yang rendah dan nilai *frequency* dan *monetary* bervariasi dari rendah hingga tinggi. Oleh karena itu *cluster* ini disegmentasi sebagai pelanggan yang hampir hilang dikarenakan sudah lama tidak melakukan pembelian.

Berikut merupakan rekap hasil pendefinisian segmentasi tersebut yang dapat dilihat pada Tabel 2.

6. Validasi hasil *clustering*: Davies Bouldin Index

Langkah terakhir adalah memvalidasi hasil *clustering*, yaitu mengetahui apakah hasil *cluster* yang didapatkan sudah valid dari sisi jumlah *cluster*. Dalam validasi ini digunakan 3 jumlah *cluster* yang berbeda dan berdekatan yaitu $k=3$, $k=4$, dan $k=5$. Hasil *Index Davies Bouldin* dapat dilihat pada Tabel 3.

Tabel 3. Rekapitulasi pendefinisian segmentasi

No	Jumlah Cluster	Nilai DBI
1	3	1,1451
2	4	0,9967
3	5	1,0058

Berdasarkan Tabel 3 terlihat bahwa nilai DBI *terkecil* ada pada jumlah *cluster* k=4, hal ini juga berkorelasi dengan hasil metode elbow yang dilakukan diawal.

7. Integrasi *Excel* dengan *Python*

Langkah terakhir adalah memindahkan hasil *clustering* tersebut ke dalam *excel*. Hal ini bisa dilakukan dengan menggunakan *package xlwings*. Hasil integrasi tersebut dapat dilihat pada Tabel 4.

3.4. *Post processing data*

Dalam tahap ini, data yang sudah dilakukan *processing* dikumpulkan ke dalam sebuah visualisasi yang mudah untuk dibaca oleh analis. Visualisasi ini dibuat dalam bentuk *dashboard* dalam *excel*. Untuk itu diperlukan rancangan terlebih dahulu tentang informasi-informasi apa saja yang ingin ditampilkan dalam *dashboard*. Berikut merupakan rancangan daftar informasi yang ingin ditampilkan dalam *dashboard*, dapat dilihat pada Tabel 5.

Setelah informasi-informasi tersebut dirancang maka langkah terakhir adalah mendesain *dashboard* tersebut. Berikut merupakan tampilan hasil akhir *dashboard* tersebut yang dapat dilihat dalam Gambar 16.

Tabel 4. Hasil integrasi *output clustering python* ke *excel*

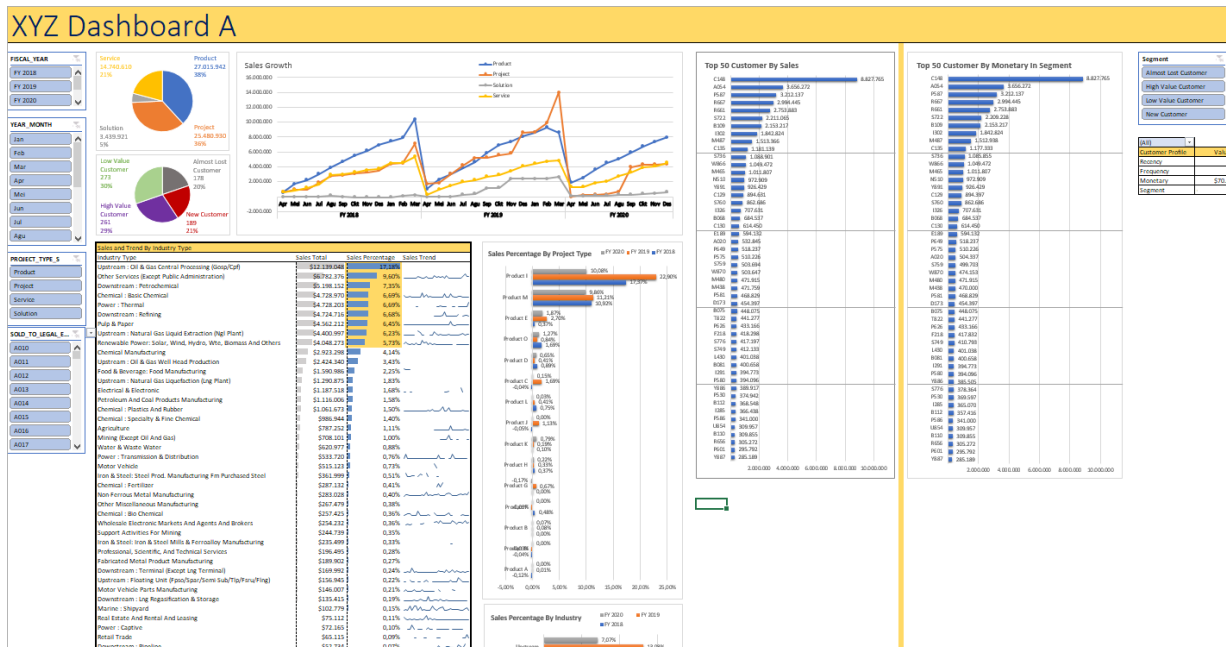
SOLD_TO_LEGAL_ENTITY_N	Recency	Frequency	Monetary	r_score	f_score	m_score	Cluster	Segment
A001	14	1	\$512	2	1	1	0	Low Value Customer
A002	5	1	\$6.539	3	1	3	2	New Customer
A003	6	1	\$286	3	1	1	2	New Customer
A004	25	1	\$130	1	2	1	0	Low Value Customer
A005	27	1	\$2.188	1	1	2	0	Low Value Customer
A006	2	3	\$5.463	4	3	3	1	High Value Customer
A007	14	6	\$138.387	2	4	4	3	Almost Lost Customer
A008	1	7	\$8.268	4	4	3	1	High Value Customer
A009	17	1	\$574	2	1	1	0	Low Value Customer
A010	13	1	\$6.387	2	1	3	0	Low Value Customer
A011	14	5	\$2.615	2	3	2	3	Almost Lost Customer
A012	8	2	\$1.101	3	2	2	2	New Customer
A013	9	2	\$2.190	3	2	2	2	New Customer
A014	1	3	\$2.951	4	3	2	2	New Customer
A015	9	2	\$1.594	3	2	2	2	New Customer
A016	6	1	\$303	3	1	1	2	New Customer
A017	14	2	\$11.678	2	3	3	3	Almost Lost Customer

Tabel 5. Rekapitulasi elemen *dashboard*

No	Nama	Atribut	Bentuk
1	<i>Sales Growth</i>	- <i>Legal Entity Name</i> - <i>Order IntakeAmount</i> - <i>Simplified Project Type</i> - <i>Fiscal Year</i>	<i>Cumulative Line Chart</i>
2	<i>Project Type Pie Chart</i>	- <i>Simplified Project Type</i> - <i>Order IntakeAmount</i>	<i>Pie Chart</i>
3	<i>Sales By Project Type</i>	- <i>Project Type</i> - <i>Order IntakeAmount</i> - <i>Fiscal Year</i>	<i>Vertical Bar Chart (percentage)</i>
4	<i>Sales By Industry Segment</i>	- <i>Industry Segment</i> - <i>Order IntakeAmount</i> - <i>Fiscal Year</i>	<i>Vertical Bar Chart (percentage)</i>
5	<i>Sales Total and Trend By Industry Type</i>	- <i>Industry Type</i> - <i>Order IntakeAmount</i> - <i>Simplified Project Type</i>	- <i>Data Bars (value and percentage)</i> - <i>Sparklines</i>
6	<i>Top 50 Customer By Sales</i>	- <i>Legal Entity Name</i> - <i>Order IntakeAmount</i>	<i>Bar Chart</i>

Tabel 5. Rekapitulasi elemen dashboard (lanjutan)

No	Nama	Atribut	Bentuk
7	Top Customer Share By Count Percentage	- Legal Entity Name - Simplified Project Type	Table
8	Top Customer Share By Sales Percentage	- Legal Entity Name - Order IntakeAmount	Table
9	Top Customer Share By Sales	- Simplified Project Type	Table
10	RFM Pie Chart	- Segmentasi Pelanggan - Cluster	Pie Chart
11	Top 50 Customer By Monetary In Segment	- Legal Entity Name - Monetary	Bar Chart
12	Customer Profile Segments	- Recency - Frequency	Table
13	Characteristics Summary	- Monetary - Segment	Table



Gambar 16. Hasil tampilan Dashboard A

4. Kesimpulan

Berikut merupakan kesimpulan dari penelitian yang telah dilakukan:

1. Untuk dapat memberikan sebuah pemetaan peran dan kontribusi dari pelanggan, maka dibutuhkan sebuah alat analisis yang dapat memetakan hal-hal tersebut. *Business intelligence* merupakan salah satu kerangka acuan yang dapat membantu mengolah data ke dalam sebuah bentuk yang mudah untuk dianalisis dalam membuat keputusan.
2. *Data mining*, salah satu cakupan dari *business intelligence* merupakan alat yang tepat untuk memudahkan proses peningkatan nilai data dan juga cocok dalam memantau kinerja perusahaan dalam kasus ini yaitu kontribusi pelanggan. *Data mining* terdiri dari tiga tahapan yaitu *preprocessing*, *processing*, dan *post processing*. Pada *preprocessing* terdapat empat langkah

utama yaitu *data integration*, *data transformation*, *data reduction*, dan *data cleaning*. Pada tahap *processing* dilakukan penentuan jumlah *cluster*, *clustering*, identifikasi karakteristik *cluster*, definisi segmentasi *cluster*, dan uji validasi hasil *cluster*. Pada tahap *post processing* dilakukan visualisasi hasil data ke dalam sebuah *dashboard*.

3. Metode RFM merupakan metode pembobotan pelanggan berdasarkan skor *recency*, *frequency*, dan *monetary*. Alasan pemilihan metode ini yaitu ketiga kriteria tersebut sangat cocok dalam mengidentifikasi kontribusi dari pelanggan dan hanya memerlukan sedikit variabel data. Teknik *clustering* yang digunakan adalah dengan *K Means Clustering*, metode ini cocok dikarenakan dapat mempartisi kelompok data berdasarkan skor kriteria dari RFM.

Diharapkan dari dashboard ini dapat bermanfaat bagi perusahaan sebagai berikut:

1. Memudahkan dalam membuat keputusan strategi *marketing* terhadap hasil segmentasi dari pelanggan.
2. Memudahkan dalam mengelola dan menjalin hubungan dengan pelanggan, terutama pelanggan yang bernilai tinggi.
3. Memudahkan untuk mengamati perkembangan sales dalam target pasar tertentu.
4. Dapat digunakan untuk mengidentifikasi pelanggan dari tipe industri dan tipe proyek mana saja yang menjadi prioritas.
5. Dapat digunakan untuk *profiling* pelanggan, perusahaan dapat melacak besar kontribusi pelanggan untuk tipe proyek tertentu.

Berikut merupakan pengembangan yang dapat dilakukan setelah penelitian yang telah dilakukan ini:

1. Dapat diketahui bahwa integrasi Excel dan Python bisa dilakukan lebih baik lagi yaitu bagaimana agar seorang analis mampu menentukan jumlah *cluster* langsung pada excel tidak perlu menggunakan IDE Python.
2. Di dalam *dataset* terdapat banyak nilai negatif, nilai negatif ini disebabkan karena koreksi pada *database*. Oleh karena itu sebaiknya dibuat sebuah sistem informasi dimana sistem dapat mendeteksi dan mengeluarkan data koreksi sebelum dilakukan *processing*.
3. Berdasarkan dataset awal yang didapatkan terdapat suatu atribut yaitu sales person. *Sales person* ini bisa dijadikan sebuah atribut untuk dapat menganalisis kinerja dari *sales person* tertentu.
4. Analisis yang sudah dilakukan pada tulisan ini bisa menjadi sebuah *guideline* untuk memudahkan perusahaan dalam mengimplementasikan sistem informasi *business intelligence* terkait kontribusi pelanggan jika PT XYZ ingin melakukan investasi ke sebuah *platform* baru yang lebih *powerful*.

5. Daftar Pustaka

- Beynon, P., & Davies. (2004). *Database Systems* (3rd ed.). PARGRAVE MACMILLAN. <https://link.springer.com/book/10.1007/978-1-349-13722-0>
- Butwall, M., Ranka, P., & Shah, S. (2019). Python in Field of Data Science: A Review. *International Journal of Computer Applications*, 178(49), 20–24. <https://doi.org/10.5120/ijca2019919404>
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2). <https://doi.org/10.1109/TPAMI.1979.4766909>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8).

<https://doi.org/10.1016/j.patrec.2009.09.011>

- Kahan, R. (1998). Using database marketing techniques to enhance your one-to-one marketing initiatives. *Journal of Consumer Marketing*, 15(5). <https://doi.org/10.1108/07363769810235965>
- Kaymak, U. (2001). Fuzzy target selection using RFM variables. *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, 2. <https://doi.org/10.1109/nafips.2001.944748>
- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6).
- Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. In *Discovering Knowledge in Data: An Introduction to Data Mining*. <https://doi.org/10.1002/0471687545>
- McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research*, 60(6). <https://doi.org/10.1016/j.jbusres.2006.06.015>
- Miglautsch, J. R. (2000). Thoughts on RFM scoring. *Journal of Database Marketing & Customer Strategy Management*, 8(1). <https://doi.org/10.1057/palgrave.jdm.3240019>
- Tsipsis, K., & Chorianopoulos, A. (2010). Data Mining Techniques in CRM: Inside Customer Segmentation. In *Data Mining Techniques in CRM: Inside Customer Segmentation*. <https://doi.org/10.1002/9780470685815>
- Turban, E, Rainer, R., & Potter, R. (2011). *Introduction to Information Technology*. John Wiley & Sons.
- Wajong, A. M. R. (2015). Applying performance dashboard in hospitals. *International Journal of Software Engineering and Its Applications*, 9(1). <https://doi.org/10.14257/ijseia.2015.9.1.19>