TESIS

ANALISIS REKOMENDASI PEMBUATAN PRODUK MENGGUNAKAN R STUDIO DAN SOSIAL MEDIA TWITTER

(Studi Kasus : PT GIT Solution)

Disusun oleh:

Nama	**: Flyas Mahananing Puri**
NIM	**: 18.52.1160**
Konsentrasi	**: Informatics Technopreneurship**

PROGRAM STUDI S2 TEKNIK INFORMATIKA

PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2022

TESIS

ANALISIS REKOMENDASI PEMBUATAN PRODUK MENGGUNAKAN R STUDIO DAN SOSIAL MEDIA TWITTER

(Studi Kasus : PT GIT Solution)

ANALYSIS OF PRODUCT RECOMMENDATION USING R STUDIO AND SOCIAL MEDIA TWITTER

(Case Study: PT GIT Solution)

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister
HALAMAN JUDUL

Disusun oleh:

Nama	**: Fiyas Mahananing Puri**
NIM	**: 18.52.1160**
Konsentrasi	**: Informatics Technopreneurship**

PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022

HALAMAN PENGESAHAN

ANALISIS REKOMENDASI PEMBUATAN PRODUK MENGGUNAKAN R STUDIO DAN SOSIAL MEDIA TWITTER
(Studi Kasus : PT GIT Solution)

ANALYSIS OF PRODUCT RECOMMENDATION USING R STUDIO AND SOCIAL MEDIA TWITTER
(Case Study: PT GIT Solution)

Dipersiapkan dan Disusun oleh

Fiyas Mahananing Puri

NIM 18.52.1160

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Selasa, 17 Mei 2022

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 17 Mei 2022
Rektor

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

ANALISIS REKOMENDASI PEMBUATAN PRODUK MENGGUNAKAN R STUDIO DAN SOSIAL MEDIA TWITTER
(Studi Kasus : PT GIT Solution)

ANALYSIS OF PRODUCT RECOMMENDATION USING R STUDIO AND SOCIAL MEDIA TWITTER
(Case Study: PT GIT Solution)

Dipersiapkan dan Disusun oleh

Fiyas Mahananing Puri

NIM 18.52.1160

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Selasa, 17 Mei 2022

Pembimbing Utama	Anggota Tim Penguji
Prof.Dr.Kusrini,M.Kom	Prof.Dr.Kusrini,M.Kom
NIK. 190302106	NIK. 190302106
	Dr. Arief Setyanto,S.Si.,M.T
Pembimbing Pendamping	NIK. 190302036
Emha Taufiq Luthfi,S.T,M.Kom	Dhani Ariatmanto,M.Kom,Ph.D
NIK. 190302215	NIK. 190302197

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 17 Mei 2022
Direktur Program Pascasarjana

Prof.Dr. Kusrini, M.Kom
NIK. 19030210

HALAMAN PERNYTAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fiyas Mahananing Puri
NIM : 18.52.1160
Konsentrasi : Informatics Technopreneurship

Menyatakan bahwa Tesis dengan judul berikut:
ANALISIS REKOMENDASI PEMBUATAN PRODUK MENGGUNAKAN R STUDIO DAN SOSIAL MEDIA TWITTER (Studi Kasus : PT GIT Solution)

Dosen Pembimbing Utama : Prof.Dr. Kusrini, M.Kom
Dosen Pembimbing Pendamping : Emha Taufiq Luthfi, S.T, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 17 Mei 2022
Yang Menyatakan,

METERAI TEMPEL

Fiyas Mahananing Puri

HALAMAN PERSEMBAHAN

Penulis mempersembahkan penelitian tesisi ini dan tidak lupa mengucapkan banyak terima kasih kepada seluruh pihak yang telah memberikan bantuan, bimbingan, masukan, dan arahan sehingga penulis dapat menyelesaikan penyusunan tesis ini :

1. Prof. Dr. M. Suyanto, M.M. selaku Ketua Universitas Amikom Yogyakarta;
2. Prof.Dr. Kusrini, M.Kom, selaku Direktur Program Pascasarjana Universitas Amikom Yogyakarta;
3. Prof. Dr. Kusrini, M.Kom., sebagai dosen pembimbing 1 yang telah meluangkan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penulisan tesis ini;
4. Emha Taufiq Luthfi, S.T., M.Kom, sebagai dosen pembimbing 2 yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing saya dalam penulisan tesis ini;
5. Suami dan Orang tua sebagai kepanjangan tangan dari Allah 'Azza Wajalla yang telah memberikan semuanya;
6. Seluruh rekan – rekan di Magister Teknik Informatika Universitas Amikom Yogyakarta;
7. Seluruh staff karyawan PT GIT Solution yang telah bersedia kerja sama dan memberikan bantuan dalam meyelesaikan tesis ini;

KATA PENGANTAR

Segala puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan tesis ini dengan sebaik – baiknya. Penulisan tesis ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Magister Teknik Informatika pada Universitas Amikom Yogyakarta.

Penulis tidak lupa mengucapkan banyak terima kasih kepada seluruh pihak yang telah memberikan bantuan, bimbingan, masukan, dan arahan sehingga penulis dapat menyelesaikan penyusunan tesis ini :

2. Prof. Dr. M. Suyanto, M.M. selaku Ketua Universitas Amikom Yogyakarta;
3. Dr. Kusrini, M.Kom, selaku Direktur Program Pascasarjana Universitas Amikom Yogyakarta;
5. Prof. Dr. Kusrini, M.Kom., sebagai dosen pembimbing 1 yang telah meluangkan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penulisan tesis ini;
6. Emha Taufik Luthfi, S.T., M.Kom, sebagai dosen pembimbing 2 yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing saya dalam penulisan tesis ini;
6. Suami dan Orang tua sebagai kepanjangan tangan dari Allah 'Azza Wajalla yang telah memberikan semuanya;
7. Seluruh rekan – rekan di Magister Teknik Informatika Universitas Amikom Yogyakarta;
8. Seluruh staff karyawan PT GIT Solution yang telah bersedia kerja sama dan memberikan bantuan dalam meyelesaikan tesis ini;

Yogyakarta, 17 Mei 2022
Penulis

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PENGESAHAN ii
HALAMAN PERSETUJUAN iii
HALAMAN PERNYATAAN KEASLIAN TESIS iv
HALAMAN PERSEMBAHAN v
KATA PENGANTAR vi
DAFTAR ISI vii
DAFTAR TABEL ix
DAFTAR GAMBAR xi
INTISARI xii
ABSTRACT xiii
BAB I PENDAHULUAN 1
1.1. Latar Belakang Masalah 1
1.2. Rumusan Masalah 4
1.3. Batasan Masalah 4
1.4. Tujuan Penelitian 5
1.5. Manfaat Penelitian 5
BAB II TINJAUAN PUSTAKA 6
2.1. Tinjauan Pustaka 6
2.2. Keaslian Penelitian 7
2.3. Landasan Teori 10
BAB III METODE PENELITIAN 26

3.1. Jenis, Sifat, dan Pendekatan Penelitian 26
3.2. Metode Pengumpulan Data 26
3.3. Metode Analisis Data 27
3.4. Alur Penelitian 28
BAB IV HASIL PENELITIAN DAN PEMBAHASAN 29
4.1. Tahap Pengumpulan Data 29
4.1.1 Crawling Data 29
4.1.2 Preprocessing Data 33
4.2.1 Pengelompokan Data 40
4.2.2 Pengujian Data Metode *Davies-Bouldin Index* (DBI) 52
4.2.3 Perhitungan Metode *Simple Additive Weighting* (SAW) 55
BAB V PENUTUP 64
5.1. Kesimpulan 64
5.2. Saran 64
DAFTAR PUSTAKA 65

DAFTAR TABEL

Tabel 2.1 Matriks Literatur Review dan Posisi Penelitian 7
Tabel 4. 1 Daftar Produk PT GIT Solution 30
Tabel 4.2 Hasil Crawling Data 40
Tabel 4. 3 Perhitungan jarak dengan 3 centroid acak 42
Tabel 4. 4 Tabel hasil perhitungan a(i) 42
Tabel 4. 5 Tabel hasil perhitungan b(i) 43
Tabel 4. 6 Perhitungan nilai indeks S(i) 43
Tabel 4.7 Pusat Awal Clustering 44
Tabel 4.8 Perhitungan Jarak Terpendek Iterasi 1 44
Tabel 4.9 Hasil Pengelompokan Iterasi 1 45
Tabel 4.10 Perhitungan Cluster Baru Iterasi 2 46
Tabel 4.11 Hasil Cluster Baru Iterasi 2 46
Tabel 4.12 Perhitungan Jarak Terpendek 46
Tabel 4.13 Hasil Pengelompokan Iterasi 2 47
Tabel 4.14 Perhitungan Cluster Baru Iterasi 2 47
Tabel 4.15 Hasil Cluster Baru Iterasi 3 48
Tabel 4.16 Hasil Pengelompokan Iterasi 3 49
Tabel 4. 17 Perhitungan Cluster Baru Iterasi 4 49
Tabel 4. 18 Hasil Cluster Baru Iterasi 4 50
Tabel 4. 19 Melakukan perhitungan Jarak Pusat Cluster Baru Itersi 3 50
Tabel 4.20 Pengelompokan Item dalam Cluster 51
Tabel 4. 21 Hasil Centroid 52

Tabel 4. 22 Data Berdasarkan Cluster 52
Tabel 4. 23 Rincian Hasil SSW 53
Tabel 4. 24 Hasil Matrik SSB 54
Tabel 4. 25 Hasil Perhitungan Rasio 54
Tabel 4.26 Data Cluster II 56
Tabel 4.27 Daftar Bobot Atribut 56
Tabel 4.28 Daftar Kriteria Tweet 56
Tabel 4. 29 Daftar Kriteria User 57
Tabel 4.30 Konversi Data 57
Tabel 4.31 Hasil Normalisasi Terbobot 58
Tabel 4.32 Hasil Perangkingan dengan SAW 58
Tabel 4. 33 Skala Prioritas Pengembangan Produk 62
Tabel 4. 34 Data Pakar Pengembangan Produk 62

DAFTAR GAMBAR

Gambar 2.1 Bagian-bagian R Studio 19

Gambar 2.2 Perbedaan data terstruktur dan tidak terstruktur 20

Gambar 2.5 *Tahapan Text Mining dan Data Mining* 21

Gambar 2.6 *Pre Processing data* 22

Gambar 3.1Diagram Alur Penelititan 28

Gambar 4.1 API dan Token Twitter 31

Gambar 4.2 Visualisasi Hasil Crawling Data 32

Gambar 4.3 Fungsi Variabel Mingingtweets_text 33

Gambar 4.4 Source Stopwords 35

Gambar 4.5 Cleaning Kata 36

Gambar 4.6 Contoh Data Sebelum Dibersihkan Format .rds 37

Gambar 4.7 Contoh Rincian Data Sebelum Dibersihkan Format .rds 38

Gambar 4. 8 Data Setelah Dibersihkan 38

Gambar 4.9 Cleaning Data 39

Gambar 4.10 Contoh Hasil Data R Studio 39

Gambar 4. 11 Flowchart K-Means Clustering 41

Gambar 4.12 Flowchart *Simple Additive Weighting* 56

Gambar 4. 13 Bagan Penggunaan 2 Metode 62

Gambar 4. 14 Grafik Hasil Perangkingan Metode SAW & Data Pakar 63

INTISARI

Dalam arus persaingan bisnis saat ini internet menjadi andalan dalam meningkatkan penjualan dan pelayanan terhadap *customer*, termasuk juga system rekomendasi yang telah banyak dimanfaatkan salah satunya pada *e-commerce*. PT GIT Solution sebagai salah satu perusahaan pengembang perangkat lunak yang berskala nasional, telah mempunyai banyak *customer* dari berbagai bidang, baik pemerintah maupun swasta. Tujuan dari sistem rekomendasi yaitu untuk memberikan usulan produk yang terkostumisasi sesuai dengan keinginan dan ketertarikan setiap *customer*. Salah satu fakta terpenting adalah bahwa siklus pengembangan perangkat lunak sudah memasuki tahap tahunan bahkan mingguan. Kepuasan pelanggan/*customer* menjadi kewajiban utama oleh perusahaan. Twitter sebagai salah satu platform micro-blogging yang paling umum saat ini dengan lebih dari 200 juta pengguna aktif menjadi sumber big data yang dapat dimanfaatkan. Untuk mendapatkan hasil rekomendasi produk penelitian menggunakan algortima k-means clustering guna mengelompokkan data bersadarkan kedekatan kriteria masing-masing data kemudian melakukan pengujian dari tahap clustering tersebut dengan menggunakan Davies-Bouldin Index (DBI) dan hasil evaluasinya sebesar 0.70% mendekati 0 yang berarti sudah cukup baik. Perangkingan juga dilakukan untuk mendapatkan hasil data produk tertinggi dengan menggunakan algortima *simple additive weighting* (saw) dan didapatkan 6 peringkat produk yang dapat dijadikan rekomendasi.

Kata kunci: K-Means, Simple Additive Weighting, Twitter, RStudio, Clustering

ABSTRACT

In the current business competition, the internet has become a mainstay in increasing sales and service to customers, including a recommendation system that has been widely used, one of which is e-commerce. PT GIT Solution as one of the software development companies on a national scale, has had many customers from various fields, both government and private. The purpose of the recommendation system is to provide customized product proposals according to the wishes and interests of each customer. One of the most important facts is that the software development cycle has entered the yearly or even weekly stage. Customer/customer satisfaction is the main obligation of the company. Twitter as one of the most common micro-blogging platforms today with more than 200 million active users is a great source of big data that can be utilized. To get the results of research product recommendations using the k-means clustering algorithm to group data based on the proximity of the criteria for each data then carry out testing from the clustering stage using the Davies-Bouldin Index (DBI) and the evaluation results are 0.70% close to 0 which means it is enough good. Ranking was also carried out to obtain the highest product data results using the simple additive weighting (SAW) algorithm and obtained 6 product ratings that could be used as recommendations.

Keyword: K-Means, Simple Additive Weighting, Twitter, RStudio, Clustering

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Trend perkembangan teknologi yang cukup pesat pada saat ini dimana sebagian besar sistem di berbagai instansi/lembaga sudah berganti dari manual menjadi terkomputerisasi, menimbulkan peluang bisnis dibidang teknologi informasi itu sendiri. Pembuatan perangkat lunak yang *up to date* dan cepat menjadi salah satu kunci sukses pada bisnis ini. Salah satu fakta terpenting adalah bahwa siklus pengembangan perangkat lunak sudah memasuki tahap tahunan bahkan mingguan. Kepuasan pelanggan/*customer* menjadi kewajiban utama oleh perusahaan.

PT GIT Solution sebagai salah satu perusahaan pengembang perangkat lunak yang berskala nasional, telah mempunyai banyak *customer* dari berbagai bidang, baik pemerintah maupun swasta. Dalam arus persaingan bisnis saat ini internet menjadi andalan dalam meningkatkan penjualan dan pelayanan terhadap *customer*, termasuk juga system rekomendasi yang telah banyak dimanfaatkan salah satunya pada *e-commerce*. Divisi business consultant perusahaan yang bertanggung jawab dalam pemasaran produk PT GIT Solution, dalam melakukan riset target pasar masih belum memanfaatkan penggunaan big data, riset dilakukan dengan menggunakan historical data client dan dengan cara mem-follow up atau melakukan visit. Tujuan dari penelitian ini ialah memberikan rekomendasi dari hasil analisis data twitter menggunakan algoritma pengelompokan K-Means Clustering dan metode perangkingan Simple Additive Weighting untuk

memberikan usulan produk yang terkostumisasi sesuai dengan keinginan dan ketertarikan setiap *customer*. Disamping itu ketersediaan informasi di dunia maya yang semakin banyak dan mudah untuk didapat, tentunya bisa dimanfaatkan guna mendukung pembuatan produk perusahaan agar lebih efisien dan efektif.

Pemanfaatan big data pada saat ini (industri 4.0) sangat dibutuhkan termasuk dalam persaingan di banyak sektor salah satunya bisnis *software development*. Twitter sebagai salah satu platform micro-blogging yang paling umum saat ini dengan lebih dari 200 juta pengguna aktif menjadi sumber big data yang dapat dimanfaatkan. Twitter dipilih karena volume data yang sangat besar, 400 juta tweet yang dikirimkan ke jaringan Twitter per harinya. Dengan memanfaatkan fitur *Hashtag* yang sangat populer sering menjadi tren di Twitter. *Hashtag* atau tagar adalah kata kunci yang diawali dengan # dan dapat ditempatkan dimana saja di isi tweet untuk mengkategorikan atau menandai kata/frasa sebagai kata kunci yang terkait dengan tweet. Dengan mengklik tagar di tweet, pengguna dapat melihat semua tweet yang mengandung tanda pagar.

Dalam penelitiannya Sreekanth Madisetty (2019), yang berjudul "*Event Recommendation using Social Media*" disimpulkan bahwa mereka menggambarkan kontribusinya yaitu memprediksi popularitas masa depan suatu acara dan merekomendasikan acara tersebut kepada pengguna dengan menggunakan popularitas yang diprediksi dan mengusulkan metode untuk mengidentifikasi tagar yang relevan untuk acara yang direncanakan menggunakan algoritma pembelajaran untuk menentukan peringkat sebagai bagian dari langkah identifikasi konten[1].

Amrita Shelar dan Ching-yu Huang (2015) dalam penelitiannya berjudul "*Sentiment Analysis of Twitter Data*" menyajikan analisis eksplorasi data dari twitter. Mereka menerapkan teknik untuk analisis sentimen dan menemukan sentimen orang dalam bentuk polaritas. Mereka bermaksud untuk menemukan lebih banyak tentang pengguna dan bisnis untuk mempelajari tentang calon investor untuk organisasi nirlaba sebagai *roadmap* masa depan[2].

Xiaohui Hu, Zichao Mai, Haolan Zhang, Yun Xue, Weixin Zhou dan Xin Chen (2016) dalam penelitiannya berjudul "A Hybrid Recommendation Model Based on Weighted Bipartite Graph and Collaborative Filtering". Menghasilkan kesimpulan Model rekomendasi *hybrid* berdasarkan *item based CF* menunjukkan nilai sedikit lebih tinggi (86,10%) dibandingkan dengan rekomendasi *item based CF* tradisional (85,55%).t [3]

Gaojun Liu dan Xingyu Wu (2019) pada penelitian yang berjudul "Using Collaborative Filtering Algorithms Combined with Doc2Vec For Movie Recommendation". Dari hasil percobaan disimpulkan, berdasarkan MovieLens dataset memverifikasi bahwa algoritma *collaborative filtering* dengan model Doc2Vec memecahkan kekurangan dari model algoritma rekomendasi tradisional sampai pada batas tertentu dan meningkatkan efek rekomendasi. [4]

Mihuandayani, Herda Dicky Rahamdita, dan Ikhwan B. Sumafta (2019) dalam penelitiannya yang berjudul "Food Trend Based on Social Media Big Data Analysis Using K-Mean Clustering Algorithm (A Case Study on Yogyakarta Culinary Industry) menunjukkan hasil kesimpulan, penggalian tren makanan dari media

social Twitter menggunakan K-Means Clustering menghasilkan akurasi data 70%-80% dibandingkan dengan penjualan pada beberapa restoran di Yogyakarta.[5]

Dalam Penelitian yang berjudul A Scable Graph Analytics Framework for Programming With Big Data in R (pbdR) oleh S.M.Shamimul Hasan, Drew Schmidt, Ramakrishnan Kannan,dan Neena Imam (2019), Hasil eksperimen menunjukkan bahwa *Framework* yang diusulkan mampu melakukan pengembangan grafik paralel skala besar melalui Bahasa R yang sudah digunakan.[6]

1.2. Rumusan Masalah

Bagian Berdasarkan uraian latar belakang diatas, maka didapatkan fokus rumusan masalah dalam penelitian sebagai berikut :

a. Bagaimana hasil pengujian algoritma *K-Means Clustering* dalam memberikan rekomendasi produk IT untuk PT GIT Solution?
b. Bagaimana hasil algoritma *Simple Additive Weighting* (SAW) dalam melakukan perangkingan produk IT menggunakan data Twitter?

1.3. Batasan Masalah

Pada penelitian ini, dibatasi oleh parameter-parameter sebagai berikut :

a. Data yang digunakan adalah data *social network* dari Twitter API (node *hashtag*).
b. Kata kunci yang digunakan ialah 12 kata yang merupakan produk IT PT GIT Solution.
c. Rekomendasi yang dihasilkan berupa hasil analisis dengan algoritma K-*Means Clustering* dan *Simple Additive Weighting* (SAW).

d. Rekomendasi yang dihasilkan akan digunakanan oleh PT GIT Solution dalam pembuatan produk *software & training IT.*

1.4. Tujuan Penelitian

Tujuan dari penelitian ini antara lain :

a. Menganalisis data social media Twitter untuk membuat rekomendasi.
b. Memberikan rekomendasi pembuatan produk kepada perusahaan lebih efektif dan efisien.

1.5. Manfaat Penelitian

Manfaat akhir dari penelitian ini nantinya yang diharapkan oleh penulis adalah sebagai berikut :

a. Secara teoritis, memperkaya penelitian di bidang data mining terutama yang berkaitan dengan Bahasa R atau penggunaan R Studio, dan penerapan algoritma *K-Means* dan *Simple Additive Weighting* (*SAW*).
b. Secara kelembagaan, membantu memberikan rekomendasi produk yang akan dibuat oleh perusahaan

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Perkembangan teknologi database dalam sistem informasi yang cepat membuat jumlah data yang disimpan maupun dikelola semakin membengkak. teknologi database yang diperlukan untuk penyimpanan dan pengolahan data pun dituntut untuk memenuhi tantangan tersebut.

Dalam penelitiannya Lutfi Ali Muharom, Alfian Futuhul Hadi, Dian Anggraeni (2017), yang berjudul "Rancang Bangun Data *Warehouse* dan R Studio Serta Pemanfaatanya dalam Peramalan Pola Konsumsi Masyarakat di Kabupaten Jember" disimpulkan bahwa Perancangan dan bangun data *warehouse* menggunakan database engine MySQL sudah memadai dalam melakukan aktifitas query data. Penggunaan R Studio sangat tepat apabila digunakan sebagai alat praktikum.

Penelitian oleh Gaojun Liu dan Xingyu Wu (2019) pada penelitian yang berjudul "Using Collaborative Filtering Algorithms Combined with Doc2vec For Movie Recommendation" menghasilkan kesimpulan bahwa, berdasarkan MovieLens dataset memverifikasi bahwa algoritma collaborative filtering dengan model Doc2Vec memecahkan kekurangan dari model algoritma rekomendasi tradisional sampai pada batas tertentu dan meningkatkan efek rekomendasi.

2.2. Keaslian Penelitian

Tabel 2.2 Matriks Literatur Review dan Posisi Penelitian

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Rancang Bangun Data *Warehouse* dan R Studio Serta Pemanfaatanya dalam Peramalan Pola Konsumsi Masyarakat di Kabupaten Jember	Lutfi Ali Muharom, Alfian Futuhul Hadi, Dian Anggraeni "Jurnal Sistem dan Teknologi Informasi Indonesia 2016"	Melakukan analisis data statistic yang digunakan dalam pengambilan keputusan dengan membangun suatu data *warehouse* sendiri dan mengolahnya menggunakan RStudio	Perancangan dan bangun data *warehouse* menggunakan database engine MySQL sudah memadai dalam melakukan aktifitas query data. Penggunaan RStudio sangat tepat apabila digunakan sebagai alat praktikum.	Session pengguna RStudio hanya satu kali, jika pengguna mencoba login dari perangkat lain, maka koneksi dari perangkat lain terputus	Persamaan dengan penelitian ini adalah pada penggunaan Bahasa R dengan RStudio sebagai tools dalam mengolah data untuk menghasilkan rekomendasi guna mengambil keputusan bagi stakeholder. Perberdaan ialah pada sumber data yang digunakan.
2	A Scable Graph Analytics Framework for Programming With Big Data in R (pbdR)	S.M.Shamimul Hasan, Drew Schmidt, Ramakrishnan Kannan, Neena Imam "International Conference on Big Data (Big Data) IEEE 2019"	Membuat *Framework* yang memanfaatkan ekosistem pemrograman dengan Big Data dalam Bahasa R (pbdR)	Hasil eksperimen menunjukkan bahwa *Framework* yang diusulkan mampu melakukan pengembangan grafik paralel skala besar melalui Bahasa R yang sudah digunakan.	Penulis Mengusulkan pengembangan kernel grafik lainnya dengan backend ComBLAS dan memperluas API grafik tingkat tinggi di pbdR ke backend selain ComBLAS.	Persamaan dengan penelitian ini ialah pada penggabungan penggunaan Big Data dan Bahasa R dalam menghasilkan sebuah usulan baru baik berupa *framework* maupun sebuah rekomendasi penunjang keputusan.
3	A Hashtag Recommendation System For Twitter Data Streams	Eriko Otsuka, Scott A. Wallace, & David Chiu. Springer.2016	Mengusulkan sistem rekomendasi tagar otomatis yang membantu pengguna	Menunjukkan bahwa HF-IHU dapat mencapai lebih dari 30% penarikan tagar ketika diminta	Penulis mengusulkan mengusulkan penggunaan struktur data indeks terbalik untuk	Penelitian ini berfokus pada *hastag* sebagai node data, jadi persamaan penelitian yang saya lakukan yaitu pada

Tabel 2.2 Matriks Literatur Review dan Posisi Penelitian (Lanjutan)

			menemukan tagar baru yang terkait dengan minat pengguna tersebut.	untuk mengidentifikasi 10 tagar yang relevan untuk sebuah tagar tertentu menciak. Selain itu, metode kami melakukan kNN, k-popularitas, dan Naïve Bayes oleh69, 54, dan 17%, masing-masing, mengingat 200 tagar teratas.	menyimpan dua peta frekuensi yang dibangun sebelum membentuk peringkat hashtag.	pemilihan salah node data berupa *hashtag* twitter sedangkan berbeda untuk tujuan penelitiannya.
4	User Profile Extraction From Twitter for personalized News Recommendation	Won-jo Lee, Kyo-Joong oh, Chae-Gyun Lim dan Ho-jin choi. International Conference on Web Intelligence Workshops IEEE/WIC/ACM. 2014	Menyelidiki metode berbeda untuk membangun profil pribadi menggunakan informasi yang diperoleh dari Twitter ke menyediakan layanan rekomendasi berita yang dipersonalisasi. Metode ini menggunakan tweets, re-tweets,dan tagar, tempat kata kunci penting diekstraksibangun profil pribadi.	Metode yang digunakan ini telah berhasil divalidasi oleh pengguna, studi bereksperimen atas rekomendasi berita prototipe pelayanan Kekuatan diskriminatif dari metode tersebut telah ditunjukkan dengan memeriksa perbedaan di antara profil pengguna,dan juga di antara daftar berita yang direkomendasikan.	Saran penelitian yaitu untuk pengembangkan data sebagai prediksi yang dapat diukur dalam hal rasio hit terhadap sekelompok kecil pengguna.	Persamaan penelitian yang saya buat dengan penelitian ini ialah pada tahap pre-processing data dari twitter untuk beberapa node yang digunakan.
5	A Hybrid Recommendation Model Based on Weighted Bipartite Graph and Collaborative Filtering	Xiaohui Hu, Zichao Mai, Haolan Zhang, Yun Xue, Weixin Zhou, Xin Chen,	Mengusulkan model rekomendasi *hybrid* yang	Model rekomendasi *hybrid* berdasarkan *item based CF* menunjukkan nilai sedikit lebih tinggi	Saran untuk mempertimbangkan faktor lingkungan seperti variable dimensi lain	Penelitian ini sejalan dengan penelitian yang saya buat yaitu penggunaan Collaborative filtering item-

Tabel 2.2 Matriks Literatur Review dan Posisi Penelitian (Lanjutan)

		"International Conference on Web Intelligence Workshops IEEE/WIC/ACM", 2016	menggabungkan *weighted bipartite network* berdasarkan *item based collaborative filtering* dan diimplementasikan pada Bookcrossing dataset.	(86,10%) dibandingkan dengan rekomendasi *item based CF* tradisional (85,55%).	yang dipertimbangkan secara komperhensif untuk membuat rekomendasi yang lebih akurat/tepat.	based, perbedaannya ialah, ditambahkan variable berupa node-node data pada twitter
6	Using Collaborative Filtering Algorithms Combined with Doc2 Vec For Movie Recommendation	Gaojun Liu, Xingyu Wu, "IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2019	Memberikan usulan model rekomendasi film berdasarkan fitur vector kata. Model yang diusulkan adalah Doc2Vec yang digunakan untuk mengekstrak semantic, tata bahasa dan urutan kata kalimat.	Dari hasil percobaan disimpulkan, berdasarkan MovieLens dataset memverifikasi bahwa algoritma *collaborative filtering* dengan model Doc2Vec memecahkan kekurangan dari model algoritma rekomendasi tradisional sampai pada batas tertentu dan meningkatkan efek rekomendasi	Kekurangannya yaitu masalah penyesuaian parameter dalam model data akan berdampak pada pengaruh model yang direkomendasikan dan *input* dari teks tunggal tidak efektif dalam meningkatkan akurasi rekomendasi. Sarannya untuk menambahkan poster film atau video terkait film untuk menganalisis dan memaksimalkan penggunaan model pembelajaran yang mendalam untuk lebih meningkatkan akurasi model rekomendasi.	Persamaan dengan penelitian ini yaitu mengenai algoritma rekomendasinya (collaborative filtering) dan pada penelitian yang saya lakukan mengikuti saran penulis yaitu untuk menyesuaikan parameter yang digunakan agar lebih akurat.

2.3. Landasan Teori

a. *K-Means Clustering*

Merupakan salah satu algoritma *clustering* "*unsupervised machine learning*" yang melakukan pengelompokan data dengan sistem partisi berdasarkan karakteristik masing-masing objek. Pengelompokan mengacu pada observasi, data atau kasus berdasarkan kemiripan objek yang diteliti. Proses *clustering* bertujuan untuk meminimalkan terjadinya *object function* yang diset dalam proses *clustering* yang pada umumnya digunakan untuk memaksimalkan variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* taua dengan kata lain data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok lain.

Proses pengelompokkan dengan algoritma K-Means adalah sebagai berikut :

1. Menentukan banyaknya *cluster* untuk hasil yang diinginkan.
2. Mengalokasikan data-data sesuai dengan jumlah *cluster* yang sudah ditentukan.
3. Menghitung rata-rata/*centroid* dari data pada setiap *cluster*.
4. Menghitung jarak terdekat dengan menggunakan rumus Euclidean.
5. Melakukan perulangan tahap ke-3 sampai menemukan kesamaan iterasi dalam *clustering*.

Rumus Euclidean yang digunakan untuk menentukan jarak adalah sebagai berikut :

$$\text{dist} = \sqrt{\sum_{k=1}^{n=3} (P_k - q_k)^2}$$

Keterangan :

dist : Jarak obyek

n : Jumlah clsuter

p^K : Koordinat dari Objek p

q^k : Koordinat dari Objek q

k : Urutan dari koordinat

Kelebihan *K-Means Clustering* antara lain :

a. Mudah untuk diimplementasikan dan dijalankan

b. Waktu yang dbutuhkan untuk melakukan pembelajaran relatif cepat.

c. Sangat Flexible dan mudah diadaptasi.

d. Sangat umum digunakan

e. Menggunakan prinsip yang sederhana dan dapat dijelaskan dalam non-statistik

Sedangkan kelemahan dari algoritma K-Means ialah :

1. Penggunaan Cluster yang random menjadikan tidak ada jaminan untuk menemukan kumpulan cluster yang optimanl.
2. Jika terdapat beberapa titik sample data yang ada, maka hal yang mudah untuk melakukan perhitungan dan mencari jarak titik terdekat dengan titik yang telah dilakukan inisialisasi yang secara acak.
3. Apablia terjebak dalam kasus yang biasanya disebut dengan curse of dimensionality. Hal ini dapat terjadi jika salah satu data untuk melakukan pelatihan mempunyai dimensi yang sangat banyak.

b. Indeks Validitas Silhoutte

Adalah suatu ukuran statistik yang digunakan untuk menyelesaikan permasalahan penentuan jumlah cluster K optimal yang dapat memberikan presentasi grafis singkat seberapa baik setiap objek terletak dalam cluster tersebut. Penentuan jumlah cluster optimal dilakukan dengan melihat nilai rata-rata maksimum dari silhouette S(i). [6]

Indeks validitas silhouette dapat dijabarkan seperti persamaan berikut ini :

$$S(i) = \frac{b(i) - a(i)}{Max\{a(i), b(i)}$$

Dimana :

a(i) = Jarak rata-rata antara i dan semua objek lain dalam cluster yang sama

b(i) = Jarak rata-rata minimum antara I ke semua objek lain dalam cluster apapun

S(i) = Nilai silhoutte

Rata-rata nilai S(i) dari seluruh objek dalam suatu cluster menunjukkan tingkat kemiripan objek dalam suatu cluster yang juga menunjukkan seberapa tepat objek telah dikelompokan. Semakin dekat S(i) kepada 1, maka semakin baik pengelompokan objek. Sebaliknya, semakin dekat S(i) kepada -1, maka semakin buruk pengelompokan objek.

Tabel 2. 3 Tabel Niai Silhoutte Kaufman dan Rousseeuw

Nilai Silhoutte Coefficient	Struktur
0.7 < SC <= 1	Struktur Kuat
0.5 < SC <= 0.7	Struktur Sedang
0.25 < SC <=0.5	Struktur Lemah
SC <= 0.25	Tidak terstruktur

c. *Davies-Bouldin Index* (DBI)

David L. Davies dan Donald W. Bouldin memperkenalkan sebuah metode yang diberinama dengan nama mereka berdua, yaitu *Davies-Bouldin Index* (DBI) yang digunakanuntuk mengevaluasi *cluster* (Nawrin, et al. 2017). Evaluasi menggunakan *Davies- Bouldin Index* ini memiliki skema evaluasi internal *cluster*, dimana baik atau tidaknya hasil *cluster* dilihat dari kuantitas dan kedekatan antar data hasil *cluster* (Bates & Kalita 2016). *Davies-Bouldin Index* merupakan salah satu metode yang digunakan untuk mengukur validitas *cluster* pada suatu metode pengelompokan, kohesididefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari*cluster* yang diikuti. Sedangkan

separasi didasarkan pada jarak antar titik pusat *cluster*terhadap *cluster*nya. Pengukuran dengan *Davies-Bouldin Index* ini memaksimalkan jarak inter-*cluster* antara *cluster* C_i dan C_j dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah *cluster*. Jika jarak inter-*cluster* maksimal, berarti kesamaan karakteristik antar-masing-masing *cluster* sedikit sehingga perbedaan antar-*cluster* terlihat lebih jelas. Jika jarak intra-*cluster* minimal berarti masing-masing objek dalam *cluster* tersebut memiliki tingkat kesamaan karakteristik yang tinggi (Wani & Riyaz 2017). Tahapan dari perhitungan *Davies- Bouldin Index* adalah sebagai berikut:

1. *Sum of Square Within-cluster* (SSW)

   Untuk mengetahui kohesi dalam sebuah *cluster* ke-i adalah dengan menghitung nilai dari *Sum of Square Within-cluster* (SSW). Kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari sebuah *cluster* yang diikuti. Persamaan yang digunakan untuk memperoleh nilai *Sum of Square Withincluster* adalah sebagai berikut.

   $$SSW_i = \frac{1}{m_i}\sum_{j=i}^{m_i} d(x_j, c_i)$$

2. *Sum of Square Between-cluster* (SSB)

   Perhitungan *Sum of Square Between-cluster* (SSB) bertujuan untuk mengetahui separasi antar *cluster*. Persamaan yang digunakan untuk menghitung nilai *Sum of Square Between cluster* adalah sebagai berikut :

   $$SSB_{i,j} = d(c_i, c_j)$$

3. *Ratio* (Rasio)

   Bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j. Untuk menghitung nilai rasio yang dimiliki oleh masing-masing *cluster*, digunakanpersamaan berikut (Nawrin, et al. 2017).

4. Davies Bouldin Index

Nilai rasio yang diperoleh dari persamaan (2.6) digunakan untuk mencari nilai. Semakin kecil nilai *Davies-Bouldin Index* (DBI) yang diperoleh, maka semakin baik *cluster* yang diperoleh dari pengelompokkan menggunakan algortima *clustering* (Bates & Kalita 2016).

d. Simple Additive Weighting (SAW)

Merupakan metode mencari penjumlahan terbobot dari ratingkinerja pada setiap alternatif dari semua atribut. Metode SAW membutuhkan proses normalisasi matriks keputusan (X) ke suatu skala yang dapat diperbandingkan dengan semua rating alternatif yang ada (Kusumadewi, 2010).

Diberikan persamaan sebagai berikut :

Dimana

$$Vi = \sum_{j=1}^{n} w_j r_{ij}$$

Vi = nilai preferensi

Wj = bobot rangking

Fij = rating kinerja ternormalisasi

Nilai Vi yang lebih besar mengindikasikan bahwa alternatif Ai lebih terpilih. (Kusumadewi, 2010). Langkah-langkah dari metode SAW adalah :

1. Menentukan kriteria-kriteria yang akan dijadikan acuan dalam pengambilan keputusan, yaitu C.
2. Menentukan rating kecocokan setiap alternatif pada setiap kriteria.
3. Membuat matriks keputusan berdasarkan kriteria (C), kemudian melakukan normalisasi matriks berdasarkan persamaan yang disesuaikan dengan jenis atribut (atribut keuntungan ataupun atribut biaya) sehingga diperoleh matriks ternormalisasi R.
4. Hasil akhir diperoleh dari proses perangkingan yaitu penjumlahan dari perkalian matriks ternormalisasi R dengan vektor bobot sehingga diperoleh nilai terbesar yang dipilih sebagai alternatif terbaik (A) sebagai solusi (Kusumadewi, 2006).

Kelebihan dari model *Simple Additive Weighting* (SAW) dibandingkan dengan model pengambilan keputusan yang lain terletak pada kemampuannya untuk melakukan penilaian secara lebih tepat karena didasarkan pada nilai kriteria dan bobot preferensi yang sudah ditentukan, selain itu SAW juga dapat menyeleksi alternatif terbaik dari sejumlah alternatif yang ada karena adanya proses perangkingan setelah menentukan nilai bobot untuk setiap atribut.

e. *Tweepy*

Tweepy merupakan library yang digunakan untuk berinteraksi dengan akun twitter dengan menggunakan twitter API (*application program interface*) dalam bahasa pemrograman phyton. Sistem interaksi tersebut menggunakan metode RESTful (*REpresentational State Transfer*) API

yang merupakan standar arsitektur komunikasi berbasis web yang diterapkan dalam pengembangan layanan berbasis web.

Untuk melakukan crawling data bahasa yang paling sering digunakan adalah python. Python memiliki banyak library yang masing-masing punya fungsi atau kegunaan yang berbeda-beda. Salah satu fungsinya untuk melakukan crawling data twitter yaitu menggunakan tweepy. Tweepy merupakan salah satu library python yang populer dan mudah digunakan untuk mengakses API dari twitter. Dengan Tweepy memudahkan kita untuk mendapatkan data dari twitter berdasarkan keyword yang digunakan. Kamu bisa mendapatkan data sekunder berupa kumpulan komentar atau text di twitter yang bisa kamu gunakan untuk bahan penelitian, misalnya kamu ingin mengetahui sentimen dan opini orang-orang terhadap suatu produk kosmetik tertentu dengan cara melakukan crawling data tweet atau komentar yang menyebutkan nama produk atau akun twitter produk tersebut.

f. R dan Data mining

R adalah sebuah bahasa pemrograman yang dikhususkan untuk mengolah data, statistik dan grafik. Data mining dapat diartikan sebagai suatu proses pencarian pola-pola yang tersembunyi (*hidden pattern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam database, data *warehouse*, atau media penyimpanan informasi yang lain.

R Merupakan bahasa yang digunakan dalam komputasi statistik yang pertama kali dikembangkan oleh Ross Ihaka dan Robert Gentlement di University of Auckland New Zealand yang merupakan akronim dari nama depan kedua pembuatnya. Sebelum R dikenal ada S yang dikembangkan oleh John Chambers dan rekan-rekan dari Bell Laboratories yang memiliki fungsi yang sama untuk komputasi statistik. Hal yang membedakan antara keduanya adalah R merupakan sistem komputasi yang bersifat gratis

R pada dasarnya dibangun oleh banyak bahasa programming mulai dari C++, C, fortran, python, dan sebagainya. Semua program itu dikemas dalam bentuk paket (package), yang kemudian bisa disertakan dalam setiap analisis data yang kita lakukan. Package pada dasarnya berisi fungsi-fungsi. misalkan untuk menganalisis ratusan bahkan ribuan table dalam format ascii, csv atau lain-lain kita bisa menggunakan package dplyr, data.table, dan lain-lain, untuk analisis data spasial seperti shapefile, raster, tiff, dan lain-lain, kita bisa menggunakan package sp, rgdal, raster, raster_view, dll. ketika menginstall R-studio pertamakali biasanya ada package-package yang sudah terinstall secara automatis, tapi banyak juga yang belum. Semua paket disimpan dalam library.

g. RStudio

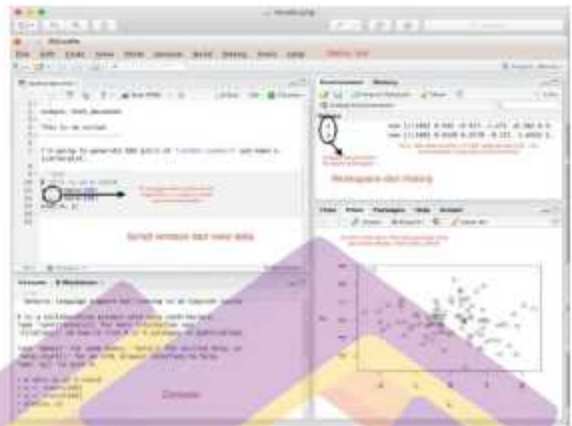
Aplikasi R pada dasarnya berbasis teks atau command line sehingga pengguna harus mengetikkan perintah-perintah tertentu dan harus hapal perintah-perintahnya. Setidaknya jika kita ingin melakukan kegiatan analisa

data menggunakan R kita harus selalu siap dengan perintah-perintah yang hendak digunakan sehingga buku manual menjadi sesuatu yang wajib adasaat berkeja dengan R.

Kondisi ini sering kali membingunkan bagi pengguna pemula maupun pengguna mahir yang sudah terbiasa dengan aplikasi statistik lain seperti SAS, SPSS, Minitab, dll. Alasan itulah yang menyebabkan pengembang R membuat berbagai frontend untuk R yang berguna untuk memudahkan dalam pengoperasian R.

RStudio merupakan salah satu bentuk frontend R yang cukup populer dan nyaman digunakan. Selain nyaman digunakan, RStudio memungkinkan kita melakukan penulisan laporan menggunakan Rmarkdown atau RNotebook serta membuat berbagai bentuk project seperti shyni, dll. Pada R studio juga memungkinkan kita mengatur working directory tanpa perlu mengetikkan sintaks pada Commander, yang diperlukan hanya memilihnya di menu RStudio. Selain itu, kita juga dapat meng-import file berisikan data tanpa perlu mengetikkan pada Commander dengan cara memilih pada menu Enviroument.

RStudio ialah salah satu *graphic user interface* untuk menjalankan pemrograman Bahasa R dengan salah satu keunggulan pada RStudio yaitu dapat dijalankan pada browser, sehingga *user* tidak perlu melakukan penginstalan R kecuali R *package* pemrograman yang sesuai dengan kebutuhan pengguna. Gambar 2.1 adalah bagian-bagian dari RStudio.



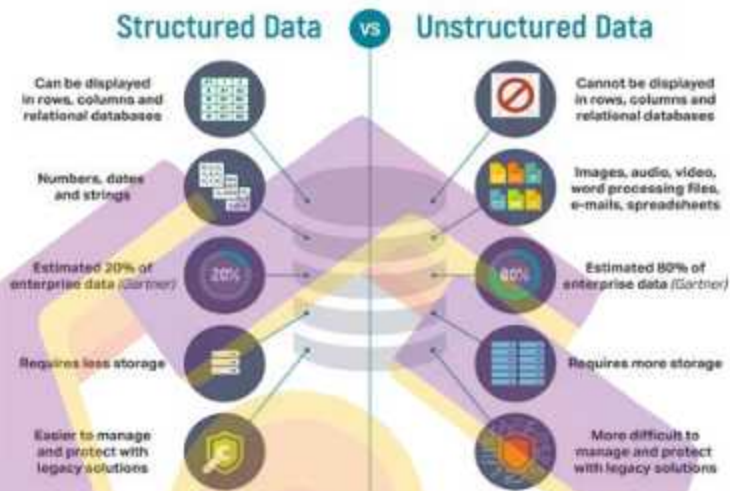
Gambar 2.1 Bagian-bagian R Studio

h. Text Mining

Text mining merupakan salah satu subjek data mining, karena teks juga adalah data. Namun karena cakupannya sangat luas sehingga penggunaan istilah Text Mining dan Data Mining dibedakan. Data mining menambang pengetahuan dari kumpulan data yang banyak dan biasanya terstruktur. Perbedaannya terdapat dari tugas data miningnya, seperti klasifikasi, klasterisasi, asosiasi, estimasi atau prediksi, bukan pada karakteristik data yang diolahnya.

Text mining merupakan penambangan pengetahuan dari data yang berupa teks dan data jenis ini sifatnya tidak terstruktur. Banyak data teks yang bisa kita temui dalam kehidupan sehari-hari, dan data tersebut bisa kita olah sesuai dengan tujuan penelitian kita. Data teks itu seperti artikel di media online, chat grup whatsapp, status atau tweet di media sosial dan lain sebagainya. Menurut penelitian, terdapat lebih dari 80% data yang ada di

internet bersifat tidak terstruktur, seperti data teks, video, audio, image dan lainnya.



Gambar 2.2 Perbedaan data terstruktur dan tidak terstruktur

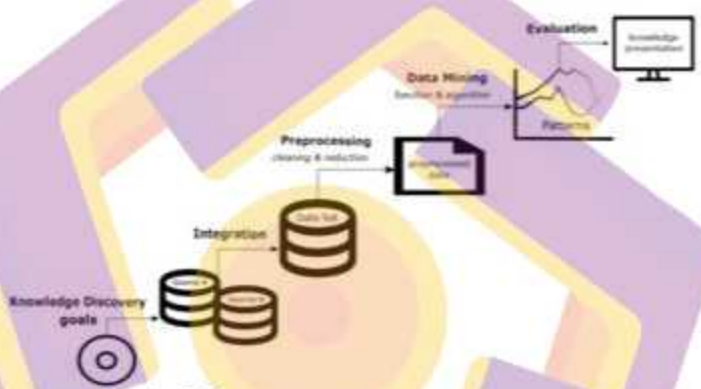
Fungsi dari text mining sangat banyak, karena cakupannya adalah teks, maka apapun yang kita ingin gali dari teks tersebut bisa kita lakukan dengan teknik Text Mining. Dari yang paling mudah adalah menganalisis sentimen berdasarkan chat / status / tweet dari banyak orang terhadap suatu kasus tertentu. Misalnya menganalisis sentimen para pengguna aplikasi game online Mobile Legend, lebih banyak mana orang yang suka atau orang yang tidak suka.

I. Tahapan Text Mining

Tahapan dalam text mining dapat dilakukan dengan cara sebagai berikut :

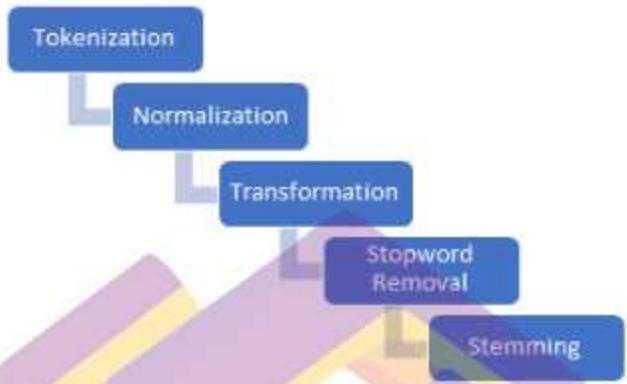
1. Knowledge Discovery Goal

2. Data Preparation
3. Data Pre processing
4. Data Modelling
5. Evaluation
6. Knowledge and Result



Gambar 2.3 *Tahapan Text Mining dan Data Mining*

Sebenarnya tidak banyak perbedaan antara tahapan dalam text mining dan data mining. Namun fokus text mining lebih banyak pada tahap kedua yaitu data pre processing.



Gambar 2.4 *Pre Processing data*

Pada tahapan data pre processing, data yang sudah disiapkan harus benar-benar bersih dari noise agar mendapatkan akurasi yang bagus saat dimodelkan. Tahapan pre processing pada text mining seperti Case Folding, yaitu membuat huruf menjadi kecil semua, tidak ada lagi huruf kapital. Selanjutnya ada Tokenization. Proses tokenisasi berfungsi untuk mengurutkan kata dalam kalimat.

Kemudian ada Stopword yang menghilangkan kata yang dianggap tidak bermakna, seperti yang, di, ke, saya, kamu, dan lain sebagainya.

Stemming untuk mengembalikan kata pada kata dasarnya, seperti "saw" menjadi "see", atau dalam bahasa Indonesia stemming bisa disebut dengan menghilangkan imbuhan seperti "mengerjakan" menjadi "kerja".

Ada juga pembobotan pada teks. Biasanya pembobotan ini digunakan untuk mencari kata kunci dari suatu dokumen. Misalnya, kita ingin mengkategorisasikan artikel berdasarkan kata kuncinya seperti Olah raga, agama, politik dan lain sebagainya. Cara kerjanya adalah dengan

melihat seberapa banyak istilah olah raga muncul dari suatu dokumen. Jika banyak, maka dianggap artikel olah raga begitu pula seterusnya. Pembobotan ini bisa menggunakan TF IDF. Namun biasanya TF IDF akan menghasilkan hasil yang sama antara satu dokumen dengan dokumen lainnya sehingga sulit untuk dikategorisasikan. Sehingga bisa dilanjutkan dengan algoritma Vector Space Model agar lebih optimal.

J. Twitter

Twitter merupakan media sosial yang banyak mendapatkan perhatian masyarakat Indonesia. Secara global, berdasarkan data pada Desember 2014, terdapat 284 juta pengguna aktif layanan tersebut. Jumlah pengguna Twitter di Indonesia sudah mencapai angka 50 juta. Jumlah ini diprediksi makin terus bertambah tiap tahun. Hal ini menyebabkan Indonesia merupakan pasar yang paling menguntungkan. Data yang dilansir Statistika berdasarkan hasil penelitian PeerReach menunjukan bahwa Indonesia tercatat sebagai pengguna Twitter terbanyak ketiga di dunia, dengan jumlah 6,5 persen setelah Amerika Serikat (24,3 persen), dan Jepang (9,3 persen). Sementara itu data dari perusahaan public relations, Webershandwick, menunjukan bahwa pengguna Twitter, berdasarkan data PT. Bakrie Telecom, twitter memiliki 19,5 juta pengguna di Indonesia dari total 500 juta pengguna global. Twitter menjadi salah satu jejaring sosial paling besar di dunia sehingga mampu meraup keuntungan mencapai USD 145 juta. Kebanyakan pengguna Twitter di Indonesia adalah konsumen,

yaitu yang tidak memiliki blog atau tidak pernah meng-upload video di Youtube namun sering update status di Twitter dan Facebook.

Chief Executive Officer Twitter, Dick Costolo, mengatakan bahwa Indonesia menjadi salah satu negara dengan pengguna Twitter terbanyak. Karena itulah Twitter pun akhirnya mendirikan kantor di Jakarta. Berdasarkan laporan Twitter di kuartal IV 2014, total pengguna aktifnya mencapai 288 juta per bulan. Menurut Costolo pengguna media sosial di Indonesia memiliki pengetahuan yang baik dengan dunia digital. Disamping itu pengguna Twitter di Indonesia dinilai sangat atraktif dan bersemangat dan dianggap sangat aktif menuliskan cuitan. Tidak jarang, hasil obrolan di lini masa menjadi Trending topic atau topik yang paling banyak dibicarakan di seluruh dunia.

Di Twitter kita bisa menuliskan sebuah status atau pesan yang nantinya akan dibaca oleh pengguna lainnya, fitur ini disebut tweet atau kicauan. Keunikan dari Twitter adalah membuat tweet atau postingan yang ada di Twitter dengan ukuran maksimum 140 karakter. Kelebihan pada media sosial ini salah satunya Twitter menyediakan API (Application Programming Interface) yang sangat baik, sehingga memudahkan setiap orang untuk mengambil data dari Twitter.

Pengumpulan data dari Twitter dapat digunakan untuk berbagai kebutuhan seperti, mengetahui popularitas kandidat pilkada atau pemilu, mendapat informasi mengenai popularitas suatu produk atau untuk yang

sederhana dapat digunakan untuk melihat semua mention, retweet atas suatu akun Twitter tertentu. Oleh karena itu, perusahaan semakin tertarik untuk mendapatkan informasi tentang apa yang dipikirkan dan dirasakan masyarakat tentang produk dan layanan mereka melalui media sosial salah satunya Twitter. Pada penelitian ini data yang diambil dari Twitter adalah data pendapat pelanggan tentang trasnportasi online Gojek dan Grab.

k. Crawling Data

Crawling adalah teknik pengumpulan data yang digunakan untuk mengindeks informasi pada halaman menggunakan URL (Uniform Resource Locator) dengan menyertakan API (Application Programming Interface) untuk melakukan penambangan dataset yang lebih besar. Data yang dapat dikumpulkan dapat berupa text, audio, video, dan gambar. peneliti dapat memulai dengan melakukan penambangan data pada API yang bersifat open source seperti yang disediakan oleh Twitter. Untuk melakukan crawling data di Twitter kamu dapat menggunakan library scrapy ataupun tweepy pada python.

Crawling data di Twitter adalah suatu proses untuk mengambil atau mengunduh data dari server Twitter dengan bantuan Application Programming Interface (API) Twitter baik berupa data user maupun data tweet. Crawling data ini dilakukan untuk mengambil data dari Twitter dimana data tersebut dibutuhkan untuk tugas akhir ini. Cara melakukan crawling data ialah dengan membuat program dengan memasukkan kata kunci untuk mencari tweet sesuai yang kita inginkan.

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Adapun jenis, sifat dan pendekatan penelitian yang akan dilakukan pada penelitian ini sebagai berikut :

a. Jenis Penelitian Eksperimen

   Penelitian ini merupakan penelitian eksperimen yaitu meliputi penerapan *R Studio, K-Means & SAW* dalam menghasilkan rekomendasi pembuatan produk *software* menggunakan data dari social media twitter.

b. Sifat Penelitian Deskriptif

c. Tujuan penelitian ini untuk mengetahui hasil rekomendasi dari penerapan *R Studio, K-Means & SAW* yang menggunakan twitter sebagai sumber data.

d. Pendekatan penelitian Kuantitatif

   Pada penelitian ini menggunakan pendekatan kuantitatif yang nantinya hasil dari penelitian ini merupakan informasi hasil rekomendasi pembuatan produk dalam meningkatkan pendapatan bagi perusahaan.

3.2. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah :

a. Untuk memperoleh data yang digunakan dalam analisa rekomendasi pembuatan produk yaitu dengan mengumpulkan *crawling* data set melalui *RDataMining Twitter Account*. Teknik-teknik yang digunak

antara lain *text mining, topic modelling*, sedangkan *tools* yang digunakan yaitu twitter API dan R Packages.

b. Untuk menjalankan algoritma K-*Means Clustering* dan *Simple Additive Weighting* (SAW) dalam sistem rekomendasinya, node data yang digunakan yaitu node *tweet* dan *user*.

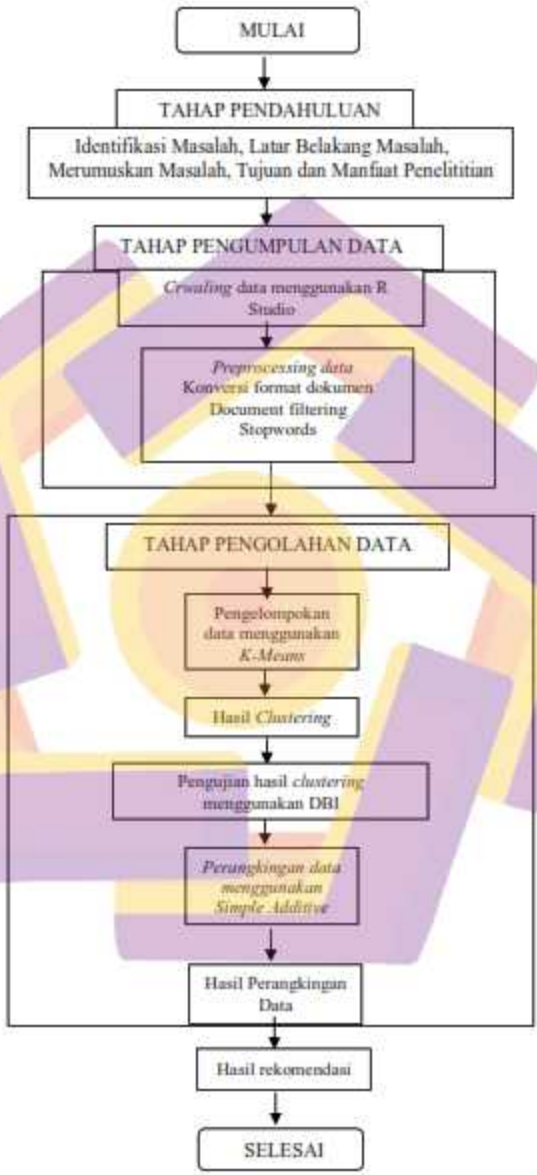
3.3. Metode Analisis Data

Metode analisis yang digunakan untuk menghasilkan rekomendasi dalam pada penelitian yaitu dengan tahapan sebagai berikut :

a. Pada tahap *preprocessing* data dilakukan dengan penyeleksian dokumen dari proses *crawling data* sebelumnya dan dilanjutkan dengan *tokenizing* untuk memisahkan suatu teks/kalimat menjadi kata tunggal dengan menghapus angka dan tanda baca atau sesuai dengan kamus yang telah ditentukan. *Stopwords* untuk proses menghilangkan kata tidak penting dalam suatu teks.

b. Pada pengolahan data dilakukan dengan dua tahap yaitu pengelompokan data berdasarkan kriteria objek menggunakan algoritma yaitu *K-Means Clustering* dan dilakukan perangkingan yang menghasilkan rekomendasi produk menggunakan metode *Simple Additive Weighting* (SAW). 5 kriteria data dari perangkingan yang digunakan yaitu nilai data dengan rate terendah, rendah, cukup, tinggi dan sangat tinggi dengan menghasilkan data dengan rangking tertinggi. Hasil data tersebut menjadi rekomendasi yang diberikan kepada perusahaan sebagai bahan pertimbangan maupun pendukung keputusan dalam penjualan produk.

3.4. Alur Penelitian



Gambar 3.1Diagram Alur Penelititan

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Tahap Pengumpulan Data

4.1.1 Crawling Data

Sebelum melakukan analisis terhadap data yang akan di olah, maka tahap yang paling pertama dilakukan adalah mengumpulkan data, dalam tahapan ini data diambil dari opini pengguna dari social media twitter dengan metode crawling data dengan rentang waktu satu bulang terakhir dengan jumlah pengambilan sample data sebanyak dua puluh ribu twit, oleh karena itu pengambilan data diaplikasikan pada keadaan yang sebenarnya agar pada tahap analisis benar-benar menghasilkan data yang akurat.

Atribut atau variabel yang digunakan dalam penitilian ini ialah node Hashtag didalam tweet dan User. Pertimbangan pemilihan hashtag sebagai atribut perhitungan yaitu dengan menambahakan tanda hashtag suatu thread/konten di twitter dapat lebih mudah diorganisir, yang dapat memudahkan pengguna twitter dalam mencari kata kunci, hal ini akan sangat baik untuk dapat meningkatkan *brand awareness* hingga *leads* dari suatu produk. Dengan kata lain hashtag akan membantu memudahkan target market perusahaan dalam mencari keyword berdasarkan kategori maupun konsep yang umum dan spesifik. Pemilihan atribut User twitter diperlukan tentu saja sebagai patokan pasti seberapa banyak pengguna yang membahas atau mendiskusikan suatu topik produk, dalam hal ini produk yang berhubungan dengan kemajuan teknologi informasi.

Data yang diambil menggunakan kata kunci yang merupakan nama atau layanan produk dari PT GIT Solution, sesuai dengan kebutuhan manajemen tim *business consultant* dan divisi teknis perusahaan untuk dapat membantu memberikan rekomendasi produk yang dijadikan prioritas dalam penjualan layanan. 12 kata kunci tersebut yaitu :

Tabel 4. 1 Daftar Produk PT GIT Solution

No	Produk
1	2d game
2	3d game
3	Augmented Reality
4	E-budgeting
5	E-planning
6	Website
7	Sistem Informasi Manajemen
8	Training IT
9	Photo VR
10	E-ticketing
11	Game simulator
12	Virtual Reality

Crawling adalah semacam pengambilan data dari media sosial kemudian dikumpulkan menjadi satu untuk dievaluasi dan di bentuk agar menjadi sebuah penelitian. Prosesnya cukup mudah dalam penelitian ini peneliti akan melakukan crawling data dari twitter dengan menggunakan *API* (*Application Programming Interface)*.

API Key Twitter adalah *Application Programming Interface*(API) dalam API ini suatu layanan berisi sekumpulan perintah, fungsi, komponen dan juga protokol yang disediakan untuk mempermudah programme pada saat membangun suatu sistem perangkat lunak. *API Key Twitter* itu sendiri memiliki suatu *consumer keys,*

consumer secret, access key, dan access secret. Consumer keys, access key, dan *access secret* tersebut digunakan untuk mengakses data Twitter yang dibutuhkan oleh programme pada Gambar 4.1 di bawah ini

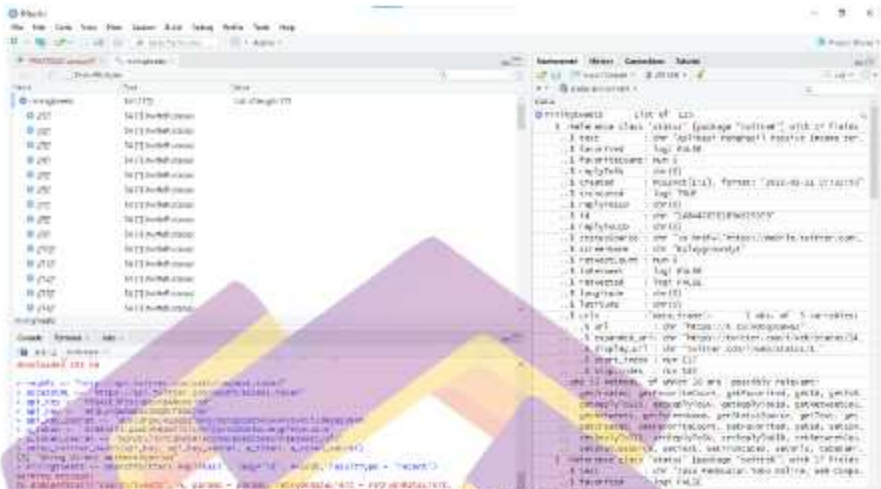
```
accessURL <- "https://api.twitter.com/oauth/access_token"
api_key <- "f8QAtE9P5sIvpUvYpA8uxxlsd"
api_key_secret <- "x3OwLXIOoxVTos1hbThibCrgPl4cZsghNaiEX8IvuyyPtVOaby"
a_token <- "1192644329140998145-Dypm394tg9iGuVb6POHkhPj1q87qP8"
a_token_secret <- "ek4MHmguoxG3Hf5MFrjIhQMM36eQsLStqeqjtyGxSZOJS"
```

Gambar 4.1 API dan Token Twitter

Bagian Hasil Penelitian dan Pembahasan merupakan bagian yang paling penting dari Tesis karena memuat semua temuan ilmiah yang diperoleh sebagai data hasil penelitian. Bagian ini diharapkan dapat memberi penjelasan ilmiah yang secara logis dapat menerangkan alasan diperolehnya hasil penelitian tersebut. Bagian ini juga harus menyediakan serangkaian alasan dalam menjawab rumusan masalah.

Pada bagian ini, peneliti menyusun secara sistematis disertai argumentasi yang rasional tentang informasi ilmiah yang diperoleh dalam penelitian, terutama informasi yang relevan dengan masalah penelitian. Pembahasan terhadap hasil penelitian yang diperoleh dapat disajikan dalam bentuk uraian teoritis, baik secara kualitatif maupun kuantitatif.

Dalam pelaksanaannya, bagian ini dapat digunakan untuk memperbandingkan hasil-hasil penelitian yang diperoleh dalam penelitian yang sedang dilakukan terhadap hasil-hasil penelitian yang dilaporkan oleh peneliti terdahulu yang diacu pada penelitian ini. Secara ilmiah, hasil penelitian yang diperoleh dalam penelitian dapat berupa temuan baru atau perbaikan, penegasan, atau penolakan interpretasi suatu fenomena ilmiah dari peneliti sebelumnya.



Gambar 4.2 Visualisasi Hasil Crawling Data

Pada gambar 4.2 menampilkan hasil crawling data menggunakan RStudio dengan 12 kata kunci yaitu #2dgame, #3dgame, #AugmentedReality, #TrainingIT, #SistemInformasiManajemen, #Website, #Virtual Reality, #E-planning, #E-Budgeting, #E-ticketing dan #gamesimulator. Gambar tersebut menunjukkan bahwa dari nilai 5000 tweet terbaru yang di ingin dikumpulkan, semuanya sukses di kumpulkan dan dapat diolah sebagai dataset, 5000 tweet ini adalah jumlah maksimal tweet yang dapat di ambil dengan menggunakan API twitter, hal ini merupakan pembatasan dari pihak twitter dengan menggunakan API gratis yang disediakan, dari twit mentah ini berisikan semua informasi twit dari masing-masing user yang membuat tweet di twitter, dengan kata lain bahwa twit tersebut masih mengandung informasi-informasi mentah dan memuat frasa yang mentah, oleh karena itu diperlukan pemisahan kata untuk membuang kata-kata yang tidak perlu ataupun kata-kata singkatan agar informasi yang didapatkan merupakan kata yang

berkaitan dengan tagar #aplikasi untuk mengetahui topik apa saja yang berkaitan dengan aplikasi.

4.1.2 Preprocessing Data

Pada saat proses preprocessing ini terdiri dari beberapa tahapan untuk menjadikan kalimat pada tweet menjadi bahasa yang baku, karena tidak sepenuhnya tweet tersebut hasil crawling menggunakan kata baku. Selain itu proses ini berguna sebagai untuk menghilangkan beberapa bagian dari kalimat yang tidak berguna. Proses *preprocessing* ini dikerjakan menggunakan bantuan dari library pada Rstudio. Untuk mengerjakan proses *preprocessing* terdapat 4 tahapan proses untuk memperoleh hasil yang maksimal, sebagai berikut:

a) Konversi Format Dokumen

Data yang telah didapatkan terlebih dahulu harus dikonversi dalam bentuk fungsi agar RStudio dapat mengenali syntax yang akan digunakan dalam pemrosesan data sebelum dilakukan filtering data, fungsi yang ditentukan pertama adalah untuk mendeskripsikan dataset sebelumnya yang bernama "miningtweets" yang telah dikumpulkan menjadi nama "miningtweets_text".

```
29 miningtweets <- searchTwitter('#aplikasi', lang="id", n=1000, resultType = "recent")
30 miningtweets_text <- sapply(miningtweets,function(x) x$getText())
31 str(miningtweets_text)
32 BB_corpus <- Corpus(VectorSource(miningtweets_text))
33 inspect(BB_corpus)
```

Gambar 4.3 Fungsi Variabel Mingingtweets_text

Pada Gambar 4.3 baris pertama menggunakan keluarga fungsi yang sama (apply, lapply, dan sapply). Fungsi-fungsi ini, secara umum, menerapkan fungsi di berbagai data terstruktur. Sapply melintasi matriks, melakukan fungsi untuk mengambil teks

dari tweet, dan mengubah teks menjadi objek daftar, bigdata_list. Baris kode berikutnya menggunakan fungsi Corpus untuk mengubah daftar tweet menjadi corpus. Corpus merupakan abstraksi dalam R untuk mewakili kumpulan dokumen yang didefinisikan sebagai BB_corpus.

b) Document Filtering dan Stop Words

Pada bagian ini, dilakukan tahap filtering data yang bertujuan untuk mengoreksi ataupun menghapus data yang tidak perlu. Data yang dibersihkan tersebut adalah data yang salah, rusak, tidak akurat, tidak lengkap dan salah format. pada tahap filtering data ini, yang perlu dihapus diantaranya: link html, tanda baca, nomor, dan kelebihan ruang (spasi/tab,dll), mengubah semua kata menjadi huruf kecil, menghapus semua tanda baca, dan menghapus stopword. Stopwords adalah kata-kata yang sangat umum yang tidak

memiliki banyak arti, seperti kata-kata singkatan dan serapan. Daftar stopword yang lebih lengkap dapat ditemukan di http://www.ranks.nl/resources/stopwords.html.

Gambar 4.4 Source Stopwords

Pada *proses filtering* dan *stop words* ini berguna untuk mengurangi atau membersihkan data tweet dari kata atau kalimat yang tidak diperlukan seperti tanda baca, unicode, dan lain-lain. Proses ini terdapat 4 tahapan yang akan dilakukan oleh sistem untuk memperoleh hasil yang maksimal, seperti di bawah ini:

1. Membersihkan tanda baca
2. Membersihkan angka
3. Merubah huruf besar menjadi huruf kecil semua
4. Membersihkan kelebihan spasi

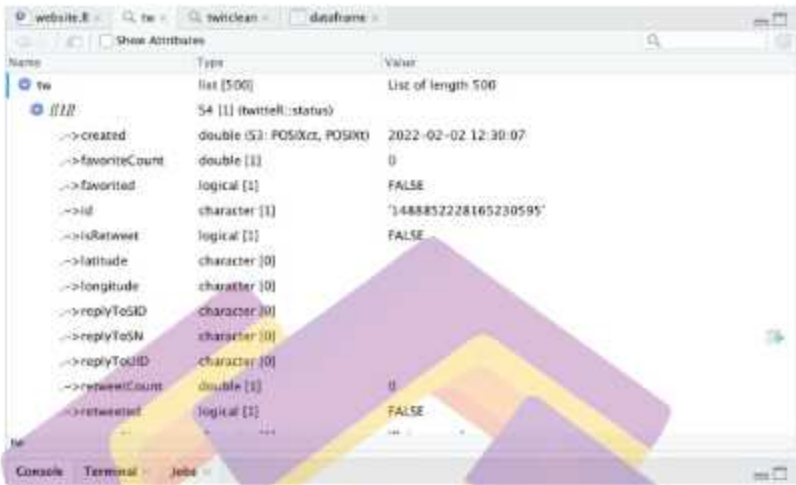
Beberapa kode program yang mengimplementasikan cleaning data dapat dilihat pada Gambar 4.5

Gambar 4.5 Cleaning Kata

Pada gambar 4.5 diperlihatkan bahwa hasil dari filtering dokumen berhasil tanpa ada error yang artinya kata-kata yang didefinisikan sebagai kata yang tidak diperlukan berhasil di pisahkan dari database, namun ada beberapa peringatan setiap syntax twitclean dijalankan, namun tidak berkaitan dengan filtering data yang dilakukan.

Tahap awal untuk mengolah data yang dilakukan adalah melihat melihat hasil *crawling* data twitter yang dilakukan di R Studio. Hasil tersebut dapat dilihat kedalam 2 bentuk yaitu dengan format .rds, seperti terlampir pada gambar 4.6 dan data berikut :

Gambar 4.6 Contoh Data Sebelum Dibersihkan Format .rds



Gambar 4.7 Contoh Rincian Data Sebelum Dibersihkan Format .rds

Pada gambar 4.6 dan 4.7 memperlihatkan contoh hasil crawling data dengan kata kunci #website. Data tersebut diatas merupakan data mentah sebelum dilakukan *cleaning csv* di RStudio. Setelah dilakukan *cleaning* data dapat terlihat seperti pada gambar 4.8

Gambar 4. 8 Data Setelah Dibersihkan

Proses ekstraksi dilakukan untuk mendapatkan representasi angka seperti pada gambar 4.9

Gambar 4.9 Cleaning Data

Gambar 4.10 Contoh Hasil Data R Studio

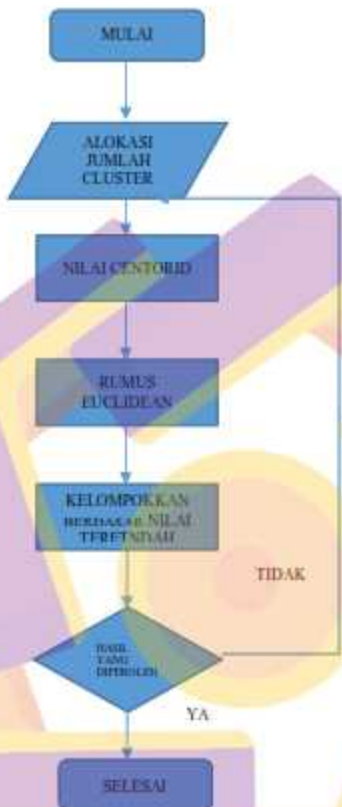
Pada proses ekstraksi fitur tersebut, dilakukan secara otomatis dan berulang terhadap masing-masing kata kunci yang telah ditetapkan sehingga mendapatkan data dengan representasi angka. Dari keseluruhan data yang diperoleh selanjutnya akan dilakukan perhitungan yaitu dengan 2 tahap, pengelompokkan menggunakan *K-Means Clustering* dan perangkingan data menggunakan *Simple Additive Weighting* (SAW). Data keseluruhan yang diperoleh seperti pada tabel 4.2 berikut :

Tabel 4.2 Hasil Crawling Data

NO	Produk	User	Tweet
1	2d game	254	381
2	3d game	122	183
3	Augmented Reality	315	473
4	E-budgeting	48	72
5	E-planning	36	54
6	Website	433	650
7	Sistem Informasi Manajemen	153	230
8	Training IT	252	378
9	Photo VR	112	168
10	E-ticketing	145	218
11	Game simulator	156	234
12	Virtual Reality	298	447

4.2.1 Pengelompokan Data

K-Means Clustering adalah salah satu teknik untuk menentukan cluster optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal Euclidean untuk jarak antar vector. Proses *clustering* bertujuan untuk meminimalkan terjadinya object function yang diset dalam proses *clustering* yang pada umumnya digunakan untuk meminimalisasikan variasi antar *cluster* dan memaksimalkan variasi antar *cluster* atau dengan kata lain data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok lain.



Gambar 4. 11 Flowchart K-Means Clustering

Dalam pengelompokan data digunakan algoritma *K-Means Clustering* dengan Langkah sebagai berikut :

1. Menghitung jumlah K (cluster) optimal yang digunakan dengan menggunakan indeks validitas silhoutte.
   a. Menghitung Jarak Pusat Cluster menggunakan rumus Euclidean dengan hasil (3 centroid acak)

Tabel 4. 3 Perhitungan jarak dengan 3 centroid acak

NO	Aplikasi	User	Tweet	Dits C1	Dist C2	Dist C3	Cluster
1	Website	433	650	0	76501	338816	1
2	2d game	254	381	104402	2165	67063	2
3	Augmented Reality	315	473	45253	4078	136420	2
4	Training IT	252	378	106745	2513	65209	2
5	Virtual Reality	298	447	59434	1076	114439	2
6	3d game	122	183	314810	80936	441	3
7	E-budgeting	48	72	482309	174638	12635	3
8	E-planning	36	54	512825	193187	17967	3
9	Sistem Informasi Manajemen	153	230	254800	52071	5976	3
10	Photo VR	112	168	335365	91518	9	3
11	E-ticketing	145	218	269568	58861	3954	3
12	Game simulator	156	234	249785	49817	6772	3

b. Menghitung jarak rata-rata objek dalam cluster yang sama

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C, i \neq j} d(i,j)$$

Tabel 4. 4 Tabel hasil perhitungan a(i)

No	Aplikasi	User	Tweet	a(i)
1	Website	433	650	781,02
2	2d game	254	381	64,44
3	Augmented Reality	315	473	85,15
4	Training IT	252	378	66,84
5	Virtual Reality	298	447	64,44
6	3d game	122	183	77,66
7	E-budgeting	48	72	138,35
8	E-planning	36	54	156,38
9	Sistem Informasi Manajemen	153	230	91,85
10	Photo VR	112	168	80,66
11	E-ticketing	145	218	84,64
12	Game simulator	156	234	96,01

c. Menghitung jarak rata-rata minimum antara objek di cluster manapun

$$b(i) = min \frac{1}{|C_k|} \sum_{j \in Ck} d(i,j)$$

Tabel 4. 5 Tabel hasil perhitungan b(i)

No	Aplikasi	User	Tweet	d(i,1)	d(i,2)	d(i,3)	b(i)
1	Website	433	650	-	276,6	582,1	276,6
2	2d game	254	381	323,1	-	259,0	259,0
3	Augmented Reality	315	473	212,7	-	369,4	212,7
4	Training IT	252	378	326,7	-	255,4	255,4
5	Virtual Reality	298	447	243,8	-	338,3	243,8
6	3d game	122	183	561,1	284,5	-	284,5
7	E-budgeting	48	72	694,5	417,9	-	417,9
8	E-planning	36	54	716,1	439,5	-	439,5
9	Sistem Informasi Manajemen	153	230	504,8	228,2	-	228,2
10	Photo VR	112	168	579,1	302,5	-	302,5
11	E-ticketing	145	218	519,2	242,6	-	242,6
12	Game simulator	156	234	499,8	223,2	-	223,2

d. Menghitung nilai silhoutte

$$S(i) = \frac{b(i) - a(i)}{Max\{a(i), b(i)}$$

Tabel 4. 6 Perhitungan nilai indeks S(i)

No	Aplikasi	User	Tweet	a(i)	b(i)	S(i)
1	Website	433	650	781,02	276,6	-1,8237645
2	2d game	254	381	64,44	259,0	0,7511726
3	Augmented Reality	315	473	85,15	212,7	0,59973623
4	Training IT	252	378	66,84	255,4	0,73824632
5	Virtual Reality	298	447	64,44	243,8	0,73568297
6	3d game	122	183	77,66	284,5	0,72702809
7	E-budgeting	48	72	138,35	417,9	0,66893409
8	E-planning	36	54	156,38	439,5	0,64421294
9	Sistem Informasi Manajemen	153	230	91,85	228,2	0,59748157
10	Photo VR	112	168	80,66	302,5	0,73336328

Tabel 4. 6 Perhitungan nilai indeks S(i) (Lanjutan)

11	E-ticketing	145	218	84,64	242,6	0,6511359
12	Game simulator	156	234	96,01	223,2	0,56983854

Dari perhitungan dengan menggunakan 3 cluster, diperoleh nilai rata-rata sebesar S(i) = 0,56425 yang menunjukkan nilai coefficeint stuktur sedang. Semakin dekat S(i) dengan 1, maka semakin baik pengelompokan objek.

2. Menentukan jumlah tingkat dan jumlah cluster yang diinginkan :

   C = Nilai cluster (centroid)

   C1 = Frekuensi user dan tweet tertinggi

   C2 = Frekuensi user dan tweet sedang

   C3 = Frekuensi user dan tweet rendah

3. Menentukan pusat awal cluster

   - Data ke-6 diambil sebagai pusat Cluster ke-1
   - Data ke-8 diambil sebagai pusat Cluster ke-2
   - Data ke-5 diambil sebagai pusat Cluster ke-3

Tabel 4.7 Pusat Awal Clustering

Pusat awal Cluster ke-1	433	650
Pusat awal Cluster ke-2	252	378
Pusat awal Cluster ke-3	36	54

4. Menghitung Jarak Pusat Cluster menggunakan rumus Euclidean dengan hasil sebagai berikut :

Tabel 4.8 Perhitungan Jarak Terpendek Iterasi 1

NO	Produk	User	Tweet	C1	C2	C3	Jarak Terpendek
1	2d game	254	381	323.1129833	3.605551275	393.005089	3.605551275
2	3d game	122	183	561.079317	234.3608329	155.0387048	155.0387048

Tabel 4.8 Perhitungan Jarak Terpendek Iterasi 1 (Lanjutan)

3	Augmented Reality	315	473	212.7275253	113.9912277	503.3905045	113.9912277
4	E-budgeting	48	72	694.4847011	367.7662301	21.63330765	21.63330765
5	E-planning	36	54	716.118007	389.3995378	0	0
6	Website	433	650	0	326.7185333	716.118007	0
7	Sistem Informasi Manajemen	153	230	504.7771786	178.058979	211.3409567	178.058979
8	Training IT	252	378	326.7185333	0	389.3995378	0
9	Photo VR	112	168	579.1070713	252.3885893	137.0109485	137.0109485
10	E-ticketing	145	218	519.1993837	192.4811679	196.918765	192.4811679
11	Game simulator	156	234	499.7849538	173.0664612	216.3330765	173.0664612
12	Virtual Reality	298	447	243.790894	82.92767934	472.3272171	82.92767934

5. Melakukan pengelompokan data kedalam masing-masing cluster yang telah ditentukan sebelumnya dari hasil sebelumnya (Langkah no 3)

Tabel 4.9 Hasil Pengelompokan Iterasi 1

No.	C1	C2	C3
1		1	
2			1
3		1	
4			1
5			1
6	1		
7		1	
8		1	
9			1
10		1	
11		1	
12		1	

Keterangan : angka 1 = jarak terpendek yang terletak pada Cn

6. Melakukan perhitungan pusat awal cluster menggunakan hasil iterasi pertama untuk mendapatkan cluster baru dengan menghitung rata-rata dari jumlah data pada masing-masing cluster.

Tabel 4.10 Perhitungan Cluster Baru Iterasi 2

NO	Produk	User	Tweet	Cluster Baru		
				C1	C2	C3
1	2d game	254	381	433	224.7142857	79.5
2	3d game	122	183	650	337.2857143	119.25
3	Augmented Reality	315	473			
4	E-budgeting	48	72			
5	E-planning	36	54			
6	Website	433	650			
7	Sistem Informasi Manajemen	153	230			
8	Training IT	252	378			
9	Photo VR	112	168			
10	E-ticketing	145	218			
11	Game simulator	156	234			
12	Virtual Reality	298	447			

7. Menentukan Pusat Awal Cluster baru dari Iterasi 1

Tabel 4.11 Hasil Cluster Baru Iterasi 2

Cluster baru yang ke-1	433	650
Cluster baru yang ke-2	224.7142857	337.2857143
Cluster baru yang ke-3	79.5	119.25

8. Melakukan perhitungan Jarak Pusat Cluster Baru Itersi 2

Tabel 4.12 Perhitungan Jarak Terpendek

NO	Produk	User	Tweet	C1	C2	C3	Jarak Terpendek
1	2d game	254	381	323.11298	52.61741	314.58435	52.61741
2	3d game	122	183	561.07932	185.34915	76.61796	76.61796
3	Augmented Reality	315	473	212.72753	163.00269	424.96978	163.00269
4	E-budgeting	48	72	694.48470	318.75453	56.78743	56.78743
5	E-planning	36	54	716.11801	340.38783	78.42074	78.42074
6	Website	433	650	0.00000	375.73017	637.69727	0.00000

Tabel 4.12 Perhitungan Jarak Terpendek (Lanjutan)

7	Sistem Informasi Manajemen	153	230	504.77718	129.04714	132.92032	129.04714
8	Training IT	252	378	326.71853	49.01187	310.97880	49.01187
9	Photo VR	112	168	579.10707	203.37690	58.59021	58.59021
10	E-ticketing	145	218	519.19938	143.46933	118.49815	118.49815
11	Game simulator	156	234	499.78495	124.05479	137.91234	124.05479
12	Virtual Reality	298	447	243.79089	131.93946	393.90648	131.93946

9. Melakukan pengelompokan data kedalam masing-masing cluster dari perhitungan jarak pusat cluster.

Tabel 4.13 Hasil Pengelompokan Iterasi 2

No.	C1	C2	C3
1		1	
2			1
3		1	
4			1
5			1
6	1		
7		1	
8		1	
9			1
10			1
11		1	
12		1	

10. Melakukan perhitungan pusat awal cluster menggunakan hasil iterasi kedua untuk mendapatkan cluster baru dengan menghitung rata-rata dari jumlah data pada masing-masing cluster.

Tabel 4.14 Perhitungan Cluster Baru Iterasi 2

NO	Produk	User	Tweet	Cluster Baru		
				C1	C2	C3
1	2d game	254	381	433	238	92.6
2	3d game	122	183	650	357.2	139

Tabel 4.14 Perhitungan Cluster Baru Iterasi 2 (Lanjutan)

3	Augmented Reality	315	473			
4	E-budgeting	48	72			
5	E-planning	36	54			
6	Website	433	650			
7	Sistem Informasi Manajemen	153	230			
8	Training IT	252	378			
9	Photo VR	112	168			
10	E-ticketing	145	218			
11	Game simulator	156	234			
12	Virtual Reality	298	447			

11. Menentukan Pusat Awal Cluster baru dari Iterasi 2

Tabel 4.15 Hasil Cluster Baru Iterasi 3

Cluster baru yang ke-1	433	650
Cluster baru yang ke-2	238	357.2
Cluster baru yang ke-3	92.6	139

12. Melakukan perhitungan Jarak Pusat Cluster Baru Itersi 3

NO	Produk	User	Tweet	C1	C2	C3	Jarak Terpendek
1	2d game	254	381	323.1129833	28.70588403	290.8847882	28.70588403
2	3d game	122	183	561.079317	209.2606694	52.91842779	52.91842779
3	Augmented Reality	315	473	212.7275253	139.0911971	401.2701833	139.0911971
4	E-budgeting	48	72	694.4847011	342.6660587	80.48701759	80.48701759
5	E-planning	36	54	716.118007	364.2993656	102.1203212	102.1203212
6	Website	433	650	0	351.818648	613.9976873	0
7	Sistem Informasi Manajemen	153	230	504.7771786	152.9586909	109.220694	109.220694
8	Training IT	252	378	326.7185333	25.10035414	287.279237	25.10035414
9	Photo VR	112	168	579.1070713	227.2884242	34.89068644	34.89068644
10	E-ticketing	145	218	519.1993837	167.3808863	94.7985232	94.7985232
11	Game simulator	156	234	499.7849538	147.9663062	114.2127839	114.2127839
12	Virtual Reality	298	447	243.790894	108.0279028	370.2069151	108.0279028

13. Melakukan pengelompokan data kedalam masing-masing cluster dari perhitungan jarak pusat cluster.

Tabel 4.16 Hasil Pengelompokan Iterasi 3

No.	C1	C2	C3
1		1	
2			1
3		1	
4			1
5			1
6	1		
7			1
8		1	
9			1
10			1
11			1
12		1	

14. Melakukan perhitungan pusat awal cluster menggunakan hasil iterasi kedua untuk mendapatkan cluster baru dengan menghitung rata-rata dari jumlah data pada masing-masing cluster.

Tabel 4. 17 Perhitungan Cluster Baru Iterasi 4

NO	Produk	User	Tweet	Cluster Baru		
				C1	C2	C3
1	2d game	254	381	433	279.75	110.2857
2	3d game	122	183	650	419.75	165.5714
3	Augmented Reality	315	473			
4	E-budgeting	48	72			
5	E-planning	36	54			
6	Website	433	650			
7	Sistem Informasi Manajemen	153	230			
8	Training IT	252	378			
9	Photo VR	112	168			

Tabel 4. 17 Perhitungan Cluster Baru Iterasi 4 (Lanjutan)

10	E-ticketing	145	218			
11	Game simulator	156	234			
12	Virtual Reality	298	447			

15. Menentukan Pusat Awal Cluster baru dari Iterasi 3

Tabel 4. 18 Hasil Cluster Baru Iterasi 4

Cluster baru yang ke-1	433	650
Cluster baru yang ke-2	279.75	419.75
Cluster baru yang ke-3	110.285	165.571

16. Melakukan perhitungan Jarak Pusat Cluster Baru Itersi 4

Tabel 4. 19 Melakukan perhitungan Jarak Pusat Cluster Baru Itersi 3

NO	Produk	User	Tweet	C1	C2	C3	Jarak Terpendek
1	2d game	254	381	323.1129833	46.52553063	258.9657609	46.52553063
2	3d game	122	183	561.079317	284.4918716	20.99951409	20.99951409
3	Augmented Reality	315	473	212.7275253	63.86019887	369.3511409	63.86019887
4	E-budgeting	48	72	694.4847011	417.8972661	112.4060606	112.4060606
5	E-planning	36	54	716.118007	439.5305735	134.0393637	134.0393637
6	Website	433	650	0	276.5874636	582.0786468	0
7	Sistem Informasi Manajemen	153	230	504.7771786	228.189888	77.30168834	77.30168834
8	Training IT	252	378	326.7185333	50.13107819	255.3602098	50.13107819
9	Photo VR	112	168	579.1070713	302.5196275	2.972664578	2.972664578
10	E-ticketing	145	218	519.1993837	242.6120875	62.87954146	62.87954146
11	Game simulator	156	234	499.7849538	223.1975022	82.29377441	82.29377441
12	Virtual Reality	298	447	243.790894	32.7967224	338.2878861	32.7967224

17. Melakukan pengelompokan data kedalam masing-masing cluster dari perhitungan jarak pusat cluster.

No.	C1	C2	C3
1		1	
2			1
3		1	
4			1
5			1
6	1		
7			1
8		1	
9			1
10			1
11			1
12		1	

Perhitungan *clustering* berhenti jika telah didapatkan hasil dari perhitungan pengelompokkan antar iterasi yang sama. Dari hasil diatas perhitungan berhenti di iterasi ke-4 yang mempunyai hasil sama dengan pengelompokan cluster pada iterasi ke-3. Hasil perhitungan pengelompokkan produk menggunakan algoritma *K-Means* sebagai berikut :

Tabel 4.20 Pengelompokan Item dalam Cluster

No.	Produk	C1	C2	C3
1	2d game		1	
2	3d game			1
3	Augmented Reality		1	
4	E-budgeting			1
5	E-planning			1
6	Website	1		
7	Sistem Informasi Manajemen			1
8	Training IT		1	
9	Photo VR			1
10	E-ticketing			1
11	Game simulator			1
12	Virtual Reality		1	

Dari tabel 4.16 didapatkan hasil bahwa produk Website termasuk kedalam pengelompokan tertinggi yang menjadi perbincangan *user Twitter*. Sehingga pada penelitian ini akan dilakukan perangkingan dengan menggunakan semua cluster 2 menggunakan metode *Simple Additive Weighting* (SAW) .

4.2.2 Pengujian Data Metode *Davies-Bouldin Index* (DBI)

1. Mengehitung *Sum of square within cluster*

Tabel 4. 21 Hasil Centroid

Cluster	User (x)	Tweet (y)
C1	433	650
C2	279.75	419.75
C3	110.286	165.571

Tabel 4. 22 Data Berdasarkan Cluster

NO	Aplikasi	Cluster	User	Tweet
1	2d game	2	254	381
2	3d game	3	122	183
3	Augmented Reality	2	315	473
4	E-budgeting	3	48	72
5	E-planning	3	36	54
6	Website	1	433	650
7	Sistem Informasi Manajemen	3	153	230
8	Training IT	2	252	378
9	Photo VR	3	112	168
10	E-ticketing	3	145	218
11	Game simulator	3	156	234
12	Virtual Reality	2	298	447

Tabel 4. 23 Rincian Hasil SSW

NO	Aplikasi	Cluster	User	Tweet	SSW1	SSW2	SSW3
1	2d game	2	254	381		46.526	
2	3d game	3	122	183			21.000
3	Augmented Reality	2	315	473		63.860	
4	E-budgeting	3	48	72			112.406
5	E-planning	3	36	54			134.039
6	Website	1	433	650	0		
7	Sistem Informasi Manajemen	3	153	230			77.302
8	Training IT	2	252	378		50.131	
9	Photo VR	3	112	168			2.973
10	E-ticketing	3	145	218			62.880
11	Game simulator	3	156	234			82.294
12	Virtual Reality	2	298	447		32.797	

Dengan total rata-rata Setiap SSW adalah sebagai berikut :

$SSW_1 = 0$

$SSW_2 = 48.33$

$SSW_3 = 70.41$

SSW merupakan indikator kohesi atau keterikatan antar anggota adalam satu kluster (homogenitas). Semakin kecil SSW dibandingkan dengan nilai centroid, maka nilai SSW/keterikatannya semakin baik. Dari hasil perhitungan diatas menunjukkan nilai keterikatan yang cukup kuat serperti SSW1yang menandakan karateristiknya mirip atau bahkan sama, seperti halnya juga di SSW2 dan SSW3.

2. Menghitung *Sum of square between cluster* (SSB)

$SSB_{i,j} = d(c_i, c_j)$

$SSB_{1,2} = \sqrt{(433 - 279.75)^2 + (650 - 419.75)^2} = 276.59$

$$SSB_{1,3} = \sqrt{(433 - 110.286)^2 + (650 - 165.571)^2} = 582.08$$

$$SBB_{2,3} = \sqrt{(279.75 - 110.286)^2 + (419.75 - 165.571)^2} = 305.49$$

Tabel 4. 24 Hasil Matrik SSB

SSB	1	2	3
1	**0**	276.59	582.08
2	276.59	**0**	305.49
3	582.08	305.49	**0**

3. Menghitung Nilai Rasio

$$Ri,j = \frac{SSWi + SSWj}{SSBi,j}$$

$$R1,2 = \frac{0 + 48.33}{276.59} = 0.175$$

$$R1,3 = \frac{0 + 70.41}{582.08} = 0.121$$

$$R2,3 = \frac{48.33 + 70.41}{305.49}$$

Tabel 4. 25 Hasil Perhitungan Rasio

R	1	2	3	RMax
1	**0**	0.175	0.121	0.175
2	0.175	**0**	0.389	0.389
3	0.121	0.389	**0**	0.389

4. Perhitungan *Davies-Bouldin Index* (DBI)

$$DBI = 1/k \sum_{i=1}^{k} i \neq j^{Max}(\mathrm{Ri,j})$$

$$DBI = \frac{0.175 + 0.389 + 0.389}{3} = 0.10$$

Dari hasil evaluasi pengujian *clustering* yang telah dilakukan, diperoleh nilai sebesar **0.10** yang dapat diartikan *cluster* yang dilakukan sudah cukup baik karena semakin kecil nilai *Davies-Bouldin Index* (DBI) maka semakin baik dan karena mendekati angka 0 (non-negatif>=0).

4.2.3 Perhitungan Metode *Simple Additive Weighting* (SAW)

Perangkingan data dilakukan menggunakan metode *Simple Additive Weighting* (SAW). Konsep dasar metode SAW adalah mencari penjumlahan terbobot dari rating kinerja pada setiap alternatif dari semua atribut (Fishburn, 1967). Kelebihan dari model *Simple Additive Weighting* (SAW) dibandingkan dengan model pengambilan keputusan yang lain terletak pada kemampuannya untuk melakukan penilaian secara lebih tepat karena didasarkan pada nilai kriteria dan bobot preferensi yang sudah ditentukan, selain itu SAW juga dapat menyeleksi alternatif terbaik dari sejumlah alternatif yang ada karena adanya proses perangkingan setelah menentukan nilai bobot untuk setiap atribut.

Gambar 4.12 Flowchart *Simple Additive Weighting*

4.2.3.1 Kriteria Penilaian

Data yang digunakan dalam perhitungan *Simple Additive Weighting* adalah 6 data produk dari Cluster 2 dengan ditambahkan 1 anggota pada Cluster 1 yang telah dilakukan perhitungan sebelumnya menggunakan K-Means Clustering.

Tabel 4.26 Data Cluster II

Kode	Produk	User	Tweet
1	2d game	254	381
2	Augmented Reality	315	473
3	Training IT	252	378
4	Virtual Reality	298	447
5	Website	433	650

4.2.3.2 Menentukan Nilai Bobot & Kriteria

Setelah mendapatkan data yang akan di rangkingkan, maka selanjutnya menentukan nilai bobot setiap atribut dan kriteria masing-masing data.

Tabel 4.27 Daftar Bobot Atribut

Kode	Keterangan	Bobot
K1	User	0.40
K2	Tweet	0.60

Tabel 4.29 Daftar Kriteria Tweet

Nilai	Keterangan	Range
1	Sangat Rendah	<= 200
2	Rendah	201-400
3	Cukup	401-600
4	Tinggi	601-800
5	Sangat Tinggi	> 800

Tabel 4. 29 Daftar Kriteria User

Nilai	Keterangan	Range
1	Sangat Rendah	<= 50
2	Rendah	51-149
3	Cukup	150-299
4	Tinggi	300-449
5	Sangat Tinggi	> 450

4.2.3.3 Mengkonversi Data

Setelah mengetahui standart/kriteria masing-masing data, maka dilakukan pengkonversian pada algorithma SAW.

Tabel 4.30 Konversi Data

No	Nama Produk	Kode	K1	K2
1	2d game	A1	3	4
2	Augmented Reality	A2	4	5
3	Training IT	A3	3	4
4	Virtual Reality	A4	3	5
5	Website	A5	4	5

4.2.3.4 Normalisasi Terbobot

Proses normalisasi dilakukan dengan pengakalian kriteria dengan bobot yang sudah ditentukan. Dengan terlebih dahulu menghitung vektor atau nilai alternatif

Alternatif-1 (A1)
r1,1 0.429
r1,2 0.571
Alternatif-2 (A2)
r2,1 0.444
r2,2 0.556
Alternatif-3 (A3)
r3,1 0.429
r3,2 0.571

Alternatif-4 (A4)	
r4,1	0.375
r4,2	0.625
Alternatif-5 (A5)	
r5,1	0.444
r 5,2	0.556

Dari hasil perhitungan normalisasi terbobot didapatkan hasil seperti berikut

Tabel 4.31 Hasil Normalisasi Terbobot

Kode	K1	K2
A1	0.429	0.571
A2	0.444	0.556
A3	0.429	0.571
A4	0.375	0.625
A5	0.444	0.556

Dari hasil normalisasi terbobot didapatkan perangkingan berdasarkan kriteria yang telah ditentukan, yaitu sebagai berikut :

Tabel 4.32 Hasil Perangkingan dengan SAW

Kode	Produk	Bobot	Ranking
A4	Virtual Reality	0.525	1
A3	Training IT	0.514	2
A1	2D Game	0.514	
A2	Augmented Reality	0.511	3
A5	Website	0.511	

Dari dua tahap perhitungan yang dilakukan yaitu dengan K-Means Clustering dan Simple Additive Weighting didapatkan hasil rekomendasi dalam perhitungan akhirnya. 4 dari 12 data produk yang masuk ke dalam cluster 2 pada perhitungan K-Means menghasilkan perangkingan data yaitu produk **Virtual Reality** pada urutan pertama, **Training IT** dan **2D Game** urutan kedua, disusul **Augmented Reality** dan **Website** pada peringkat ketiga.

Dengan menggunakan algoritma perhitungan yang sama, peneliti juga menghitung kemungkinan lain dalam pengolahan data, yaitu dari proses pengelompokkan sampai perangkingan *Simple Additive Weighting* dengan 2 clustering. Perhitungan clustering dilakukan dengan perhitungan jarak terpendek dengan menggunakan 2 centroid awal.

Tabel 4. 33 Hasil Perhitungan Jarak Pusat Cluster

No	Aplikasi	User	Tweet	C1	C2	Jarak Terpendek
1	2d game	254	381	101,843	258,97	101,84
2	3d game	122	183	339,809	21	21,00
3	Augmented Reality	315	473	8,544	369,35	8,54
4	E-budgeting	48	72	473,215	112,41	112,41
5	E-planning	36	54	494,848	134,04	134,04
6	Website	433	650	221,27	582,08	221,27
7	Sistem Informasi Manajemen	153	230	283,507	77,302	77,30
8	Training IT	252	378	105,449	255,36	105,45
9	Photo VR	112	168	357,837	2,9727	2,97
10	E-ticketing	145	218	297,93	62,88	62,88
11	Game simulator	156	234	278,515	82,294	82,29
12	Virtual Reality	298	447	22,5211	338,29	22,52

Hasil dari clustering didapatkan 5 produk yang termasuk dalam cluster tertinggi atau cluster pertama yaitu **Virtual Reality, Augmented Reality, Training IT, 2D Game,** dan **Website**. Sedangkan 7 produk dapat dikelompokan ke dalam clustering kedua yaitu **3D Game, E-Budgeting, E-Planning, SIM, Photo VR, E-Ticketing, Game Simulator.**

Tabel 4. 34 Pengelompokan Data

No.	Aplikasi	C1	C2
1	2d game	1	
2	3d game		1
3	Augmented Reality	1	
4	E-budgeting		1
5	E-planning		1
6	Website	1	
7	Sistem Informasi Manajemen		1
8	Training IT	1	
9	Photo VR		1
10	E-ticketing		1
11	Game simulator		1
12	Virtual Reality	1	

Dari hasil clustering tersebut dilakukan perangkingan menggunakan *Simple Additive Weighting* dengan mengambil data pada produk dalam cluster 1, nilai bobot dan kriteria perangkingan menggunakan nilai yang sama dengan perhitungan yang sebelumnya dilakukan. Hasil perhitungan SAW sebagai berikut :

Tabel 4. 35 Hasil Normalisasi dengan Algoritma SAW

Kode	Nama Aplikasi	K1	K2
A1	2d game	0,60	0,40
A2	Augmented Reality	0,57	0,43
A3	Training IT	0,60	0,40
A4	Virtual Reality	0,50	0,50
A5	Website	0,50	0,50

Tabel 4. 36 Hasil Perangkingan dengan SAW

Kode	Produk	Bobot	Ranking
A4	Virtual Reality	0.500	1
A3	Training IT	0.480	2
A1	2D Game	0.480	
A2	Augmented Reality	0.480	
A5	Website	0.300	3

Peneliti juga melakukan perhitungan untuk perbandingan dengan manual menggunakan menu toolbar Short dari microsoft excel. Hasil dari pengurutan manual (short) didapatkan hasil sebagai berikut :

Tabel 4. 37 Hasil Perhitungan Manual (Short ms. Excel)

Produk	Ranking
Website	1
Augmented Reality	2
Virtual Relality	3
2D Game	4
Training IT	5

Dengan melihat hasil perhitungan manual (short) dan perhitungan menggunakan clustering dan perangkingan *Simple Additive Weighting* terlihat perbedaan dari hasil perangkingan itu sendiri dan dari parameter yang digunakan. Jika menggunakan short perangkingan hanya menggunakan salah satu atribut yang dimiliki yaitu user atau tweet, sedangan jika dalam algoritma K-Menas Clustering kedua atribut dari data twitter tersebut dapat digunakan, begitu juga dengan algoritma perangkingan *Simple Additive Weighting* yang memperhitungan nilai bobot yang berbeda-beda dari masing-masing atribut baik jumlah user maupun tweet.

Gambar 4. 13 Bagan Penggunaan 2 Metode

Hasil perangkingan menggunakan metode SAW kemudian diuji menggunakan data para pakar, dimana data pada penelitian ini diperoleh dengan metode wawancara kepada *stake holder* PT GIT Solution yang berwenang sebagai Direktur Divisi Pengembangan *Software*. Hasil wawancara tersebut memberikan data perangkingan produk berdasarkan skala prioritas produk yang akan dikembangkan secara intens, seperti pada tabel 4.30 berikut :

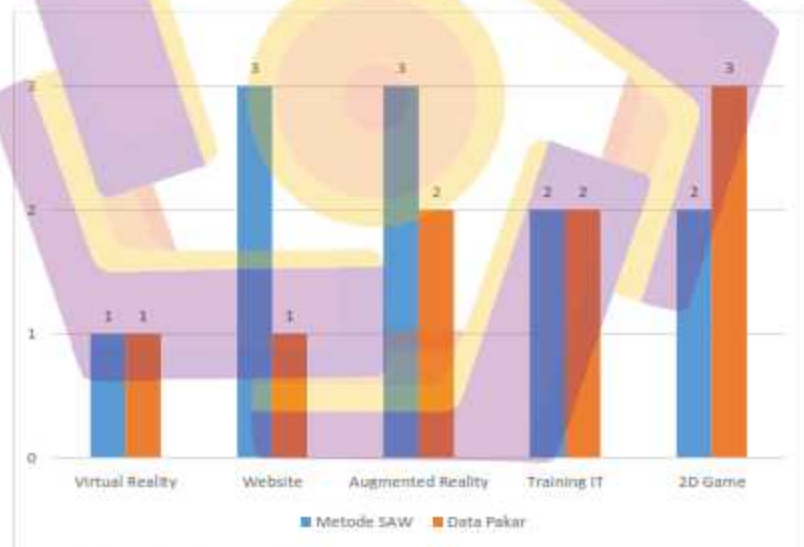
Tabel 4. 39 Skala Prioritas Pengembangan Produk

No	Keterangan	Nilai
1	Sangat Rendah	1
2	Rendah	2
3	Cukup	3
4	Tinggi	4
5	Sangat Tinggi	5

Tabel 4. 40 Data Pakar Pengembangan Produk

No	Kode	Nama Produk	Skala (1-5)	Prosentase
1	A1	2D Game	2	40%
2	A2	Augmented Reality	3	60%
3	A3	Training IT	3	60%
4	A4	Virtual Reality	4	80%
5	A5	Website	4	80%

Dari data tabel 4.30 dapat disimpulkan perangkingan menurut pakar, produk yang menduduki peringkat pertama yaitu **Virtual Relaity dan Website**, kedua **Augmented Reality** dan **Training IT** kemudian peringkat ketiga ialah **2D Game.** Hasil dari data pakar kemudian digunakan untuk menguji akurasi metode *Simple Additive Weighting* yang sudah didapatkan sebelumnya, dengan nilai kecocokan dan perbedaan sebesar 33,33%.



Gambar 4. 14 Grafik Hasil Perangkingan Metode SAW & Data Pakar

BAB V

PENUTUP

5.1. Kesimpulan

1. Hasil pengujian DBI dari perhitungan menggunakan algoritma K-Means Clustering didapatkan nilai akurasi pengujian yaitu sebesar 0.10 yang berarti *clustering* sudah cukup baik karena nilai mendekati angka 0 dengan jumlah pengelompokan yang dihasilkan sebanyak 3 cluster.
2. Dari hasil perangkingan menggunakan algorithma *Simple Additive Weighting* diperoleh nilai akurasi sebesar 33,33% dengan menggunakan data pakar, dengan urutan perangkingan produk peringkat pertama **Virtual Reality,** kedua **Training IT** dan **2D Game,** sedangkan peringkat ketiga yaitu **Augmented Reality & Website**.

5.2. Saran

1. Penggunaan database dari sumber lain seperti data keperintahan guna mendapatkan informasi beragam yang dapat dijadikan perbandingan.
2. Penggunaan *Sentiment Analyst* untuk mempertimbangkan *issue* positif & negatif terhadap produk perusahaan.
3. Perlunya penambahan pengujian metode seperti akurasi, recall dan presisi pada algoritma perangkingan SAW

DAFTAR PUSTAKA

PUSTAKA BUKU

Kusumadewi, S., & Pramono, H.(2010). Aplikasi Logika Fuzzy untuk Pendukung Keputusan Edisi , Yogyakarta, Graha Ilmu

Kusumadewi, S. Hatati, S. Harjoko, A. dan Wardoyo, R. (2006). Fuzzy Multi Attribute Decision Making (FUZZY MADM). Yogyakarta : Graha Ilmu

Marimin. (2004). Teknik dan Aplikasi Pengambilan Keputusan Kriteria. Jakarta : Penerbit Grasindo

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

Herda D.R, Ikhwan B S, Arief Setyanto. 2018 . *Food Trend Based on Social Media Big Data Analysis Using K-Mean Clustering Algorithm (A Case Study on Yogyakarta Culinary Industry)*. IEEE : International Conference on Information and Communications Technology (ICOIACT)

Gaojun Liu, Xingyu Wu. 2019. *Using Collaborative Filtering Algorithms Combined with Doc2Vec for Movie Recommendation*. IEEE : 3rd Information Technology,Networking,Electronic and Automation Control Conference (ITNEC 2019)

Eriko Otsuka, Scott A. Wallace, David Chiu. 2016. *A Hashtag Recommendation System For Twitter Data Streams*. Computational Social Network DOI 10.1186/s40649-016-0028-9

Lutfi Ali Muharom, Alfian Futuhul Hadi, Dian Anggraeni.(2016). *Rancang Bangun Data Warehouse dan R Studio Serta Pemanfaatanya dalam Peramalan Pola Konsumsi Masyarakat di Kabupaten Jember*. Jurnal Sistem dan Teknologi Informasi Indonesia

S.M.Shamimul Hasan, Drew Schmidt, Ramakrishnan Kannan, Neena Imam.(2016). *A Scable Graph Analytics Framework for Programming With Big Data in R (pbdR)*. IEEE : International Conference on Big Data (Big Data) IEEE 2029

Won-jo Lee, Kyo-Joong oh, Chae-Gyun Lim dan Ho-jin choi.(2014). *User Profile Extraction From Twitter for personalized News Recommendation*. IEEE : International Conference on Web Intelligence Workshops IEEE/WIC/ACM

PUSTAKA LAPORAN PENELITIAN

Jumadi Bernad D.S22 Januari 2022, 2018, Peningkatan Hasil Evaluasi *Clustering Davides-Boulding Index* Dengan Penentuan Titik Pusat Cluster Awal Algortima K-Means, Tesis, S2 Teknik Informatika, Universitas Sumatera Utara, Medan