

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 *Business Intelligence*

Business Intelligence adalah kumpulan proses yang terdiri atas ekstraksi data operasional perusahaan, lalu mengumpulkannya ke dalam sebuah data warehouse untuk selanjutnya diproses menggunakan berbagai macam proses analisis statistik dan proses *data mining*, untuk mendapatkan berbagai kecenderungan *pattern* (pola) dari data tersebut (Ralph Kimball & Joe Caserta 2004). *Business Intelligence* bukanlah sebuah produk atau sistem namun merupakan sebuah teknik.

Business Intelligence roadmap merupakan panduan siklus hidup proyek untuk mengembangkan aplikasi pendukung keputusan BI menggunakan data terstruktur yang secara khusus membahas aplikasi dan database pendukung keputusan. *Business Intelligence* menggabungkan arsitektur, basis data atau data warehouse, tool analisis dan aplikasi (Turban 2007).

BI digunakan untuk aplikasi dan teknologi dalam mengumpulkan, menyimpan, menganalisis, dan menyediakan akses pada data sehingga dapat membantu pengguna dari kalangan perusahaan atau organisasi untuk mengambil keputusan dengan lebih baik dan tepat.

2.1.2 *Dashboard System*

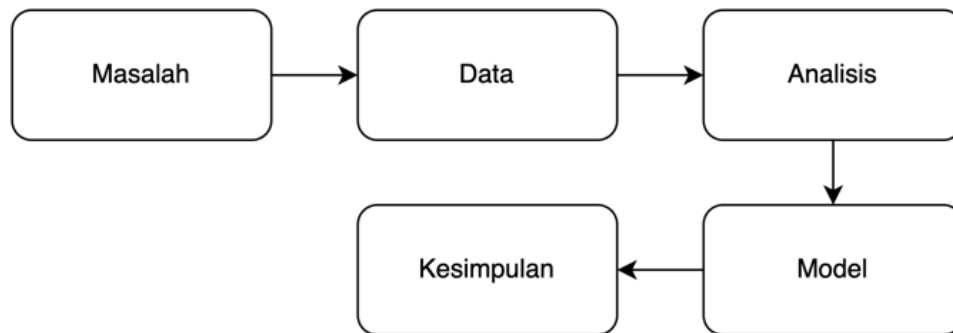
Dashboard merupakan sebuah aplikasi yang berfungsi untuk menampilkan informasi terkait kinerja bagi para manajer perusahaan. Konsep *dashboard* sudah ada selama bertahun-tahun dan telah diadopsi oleh banyak perusahaan di dunia (Eckersone, 2006). *Dashboard* adalah representasi visual berisi informasi penting yang dibutuhkan untuk mencapai tujuan dan dapat diatur pada satu layar sehingga akan memudahkan pengguna dalam memantaunya. Sementara itu, *information dashboard* adalah tampilan visual berisi informasi penting yang dibutuhkan untuk mencapai tujuan dengan mengatur informasi dalam satu layar sehingga kinerja organisasi dapat dimonitor (Few, 2006).

2.1.3 *Exploratory Data Analysis*

Exploratory Data Analysis (EDA) adalah proses menganalisis dan menampilkan data bertujuan mendapatkan pemahaman yang lebih baik tentang wawasan dari data (Andreas Chandra 2019). Ada berbagai langkah yang dilakukan saat melakukan EDA, berikut ini adalah langkah-langkah umum yang dapat diambil dalam melakukan analisis EDA data:

- a. Memaksimalkan wawasan ke dalam kumpulan data.
- b. Mengungkap struktur data.
- c. Ekstrak variabel yang penting.
- d. Mendeteksi outlier dan anomali.
- e. Melakukan uji asumsi.
- f. Mengembangkan model.
- g. Menentukan faktor yang optimal.

Exploratory Data Analysis (EDA) memiliki 5 tahapan yang digambarkan oleh Gambar 2. 1 Tahapan EDA (Dessy Aryanti & Johan Setiawan 2018) :



Gambar 2. 1 Tahapan EDA

Pada metode EDA, tidak diikuti dengan penerapan model, melainkan diikuti dengan tujuan yang menyimpulkan model apa yang cocok untuk digunakan.

2.1.4 *Fuzzy String Matching*

Algoritma pencarian string atau dapat disebut juga algoritma pencocokan string yaitu algoritma untuk dapat melakukan pencarian semua kemunculan teks/string pendek dan panjang, untuk string pendek yang disebut pola/*pattern* dan string yang lebih panjang yang disebut teks. string pendek adalah $pattern [0..n-1]$ string panjang adalah $teks [0..m-1]$. Algoritma yang digunakan dalam pencarian *string* ini dapat diklasifikasikan ke 3 bagian menurut arah pencariannya . Arah yang paling alami yaitu dimulai dari kiri ke kanan, kemudian yang merupakan arah untuk membaca (Noor Fauzy 2019).

Fuzzy String Matching merupakan salah satu metode pencarian string dengan menggunakan proses pendekatan pada pola dari string yang akan dicari.

Dalam melakukan proses pencarian pada string yang sama dan juga string mendekati string lainnya akan terkumpul dalam sebuah kamus.

Kunci konsep pencarian seperti ini ialah bagaimana memutuskan bahwa sebuah string yang akan dicari dapat memiliki kesamaan dengan string di kamus, meskipun tidak sama persis dalam susunan karakternya. Untuk memutuskan kesesuaian dipergunakan sebuah fungsi yang disebut sebagai *similarity function* (Noor Fauzy 2019). Fungsi ini bertugas dalam memutuskan string hasil pencarian jika ada ditemukan string hasil pendekatan (aproksimasi). *Approximate String Matching* adalah pencocokan string dengan dasar kemiripan dari segi penulisannya, tingkat kemiripan ditentukan dengan jauh tidaknya beda penulisan dengan dua buah string dibandingkan itu.

Ada banyak variasi dari *Fuzzy String Matching* ini, beberapa contohnya adalah Levensthein, Sellers, Damerau-Levenshtein, dan Hamming. Keempat algoritma ini sebenarnya hanya menggunakan cara yang berbeda untuk melakukan hal yang sama (Tregoat 2018).

2.1.5 *Approximate String Matching*

Approximate String Matching merupakan pencocokan string dengan dasar kemiripan dari segi penulisannya (jumlah karakter dan susunan karakter), tingkat kemiripan ditentukan dengan jauh tidaknya beda penulisan dua buah string yang dibandingkan tersebut (Hariyanto 2004).

Operasi mengubah string tersebut dapat berupa mengubah satu huruf ke huruf yang lain, menghapus satu huruf dari string, atau memasukkan satu huruf ke

dalam string. Operasi- operasi ini digunakan untuk menghitung jumlah perbedaan yang diperlukan untuk pertimbangan kecocokan suatu string dengan string sumber, jumlah perbedaan tersebut diperoleh dari penjumlahan semua perubahan yang terjadi dari masing-masing operasi (Isbad Amar Gurning et al. 2016).

Operasi untuk mengubah string tersebut ada 3 operasi (Madaharsa & Adiwidya 2009), yaitu:

1. Operasi Penghapusan

Operasi ini dilakukan dengan menghapus karakter pada suatu kata agar menyesuaikan string sumber (S) dengan string target (T).

2. Operasi Penyisipan

Operasi ini dilakukan dengan menyisipkan karakter pada indeks tertentu untuk menyamakan string sumber (S) dengan string target (T).

3. Operasi Penukaran

Operasi ini dilakukan dengan menukar suatu karakter untuk menyamakan string sumber (S) dengan string target (T)

2.1.6 *Algoritma Levenhstein Distance*

Algoritma *Levenshtein distance* ditemukan oleh ilmuwan asal Rusia bernama Vladimir Levenshtein pada tahun 1963, algoritma ini juga disebut dengan algoritma Edit Distance. Perhitungan edit distance didapatkan dari matriks yang digunakan untuk menghitung jumlah perbedaan string antara dua string, sebagai contoh hasil penggunaan algoritma ini, string “komputer” dan “*computer*” memiliki distance 1 karena hanya perlu dilakukan satu operasi saja untuk mengubah satu

string ke string yang lain. Dalam kasus dua string di atas, string “*computer*” dapat menjadi “*komputer*” hanya dengan melakukan satu penukaran karakter “*c*” menjadi “*k*” (Andhika, 2010).

Algoritma *Levenshtein Distance* bekerja dengan menghitung jumlah minimum pentransformasian suatu string menjadi string lain yang meliputi penghapusan, penyisipan, dan penukaran.

Selisih perbedaan antar string dapat diperoleh dengan memeriksa apakah suatu string sumber sesuai dengan string target. Nilai selisih perbedaan ini disebut juga *edit distance* atau jarak Levenshtein. Jarak Levenshtein antar string “*s*” dan string “*t*” tersebut adalah fungsi *D* yang memetakan (*s,t*) ke suatu bilangan real nonnegatif, sebagai contoh diberikan dua buah *string* $s = s(1)s(2),s(3),\dots,s(m)$ dan $t = t(1),t(2),t(3),\dots,t(n)$ dengan $|s| = m$ dan $|t| = n$ sepanjang alfabet *V* berukuran *r* sehingga “*s*” dan “*t*” anggota dari V^* . $s(j)$ adalah karakter pada posisi ke-*j* pada *string* “*s*” dan $t(i)$ adalah karakter pada posisi ke-*i* pada *string* “*t*”. Sehingga jarak Levenshtein dapat didefinisikan sebagai (Harahap, 2013)

$$D(s, t) = d(s_1, t_1) + d(s_2, t_2) + \dots + d(s_l, t_l)$$

$$D(s, t) = \sum_{i=1}^l d(s_i, t_i)$$

dimana $s_i, t_i \in V$ *untuk* $i = 1, 2, 3, \dots, l$
 $d(s_i, t_i) = 0$ *jika* $s_i = t_i$ *dan* $d(s_i, t_i) = 1$ *jika* $s_i \neq t_i$

D(s,t) adalah banyaknya operasi minimum dari operasi penghapusan, penyisipan dan penukaran untuk menyamakan string *s* dan *t*. Pada implementasi pencocokan antar string, ketiga operasi tersebut dapat dilakukan sekaligus untuk menyamakan string sumber dengan string target.

2.1.7 Python Programming Language

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. *Python* juga didukung oleh komunitas yang besar (Nur Syahrudin & Kurniawan 2018). Kepopuleran penggunaan *Python* menempatkannya menjadi bahasa pemrograman yang mulai banyak dipelajari oleh mahasiswa terutama mahasiswa di kampus yang berbasis IT, guna menyelesaikan tugas kuliah, tugas akhir maupun tugas penelitian, untuk dapat menyelesaikan berbagai tugas pemrograman, seseorang perlu memahami algoritma, karena pada dasarnya program komputer adalah implementasi dari algoritma.

2.1.8 Tableau

Tableau adalah aplikasi platform BI. *Tableau* didesain untuk penggunaan individu tetapi memiliki skala perusahaan. *Tableau* memiliki empat jenis, yaitu *Tableau Prep*, *Tableau Desktop*, *Tableau Online*, dan *Tableau Server*. *Tableau Prep* digunakan untuk menyiapkan data. *Tableau Desktop* yang digunakan untuk menganalisis data. *Tableau Online* merupakan aplikasi yang memiliki fungsi yang sama dengan *Tableau Desktop* tetapi bisa digunakan secara online tanpa perlu memiliki server yang sama tetapi memiliki fitur yang lebih sedikit daripada *Tableau Desktop*. *Tableau Server* dikhususkan untuk penggunaan skala perusahaan untuk menganalisis data yang memiliki server yang sama.

2.2 Penelitian Terkait

Tabel 2. 1 Tabel Penelitian Terkait

No	Judul	Penulis dan Tahun	Hasil Penelitian
1	Exploratory Data Analysis pada Kasus COVID-19 di Indonesia Menggunakan HiveQL dan Hadoop Environment	(Fahrudin et al. 2020)	Hasil analisis menunjukkan bahwa teknik statistik deskriptif menemukan total jumlah kasus COVID-19 tiap kategori paling sedikit terjadi pada bulan Maret sedangkan paling besar terjadi di bulan September 2020. Rata-rata harian kasus COVID-19 tiap bulannya selalu bertambah, dari 50 kasus per hari di bulan Maret hingga melonjak menjadi 3.740 per hari di bulan September.
2	Exploratory Data Analysis terhadap Kepadatan Penumpang Kereta Rel Listrik	(Samosir et al. 2021)	Jadwal kereta memiliki korelasi kuat (linier) dan sangat kuat (eksponensial), terhadap kepadatan penumpang pada setiap kereta KRL
3	Analisis Big Data dengan Metode Exploratory Data Analysis (EDA) dan Metode Visualisasi Menggunakan Jupyter Notebook	(Radhi et al. 2021)	Pada penelitian ini dapat ditarik kesimpulan bahwa dalam penjualan barang elektronik dari bulan januari 2019 – desember 2019 yang paling banyak terjual adalah baterai AAA (4 packs) sedangkan barang yang paling sedikit terjual adalah LG Dryer. Dan untuk produk elektronik yang paling mahal terjual sepanjang bulan januari – desember adalah Macbook Pro Laptop dengan kisaran harga sebesar 1750 dolar

No	Judul	Penulis dan Tahun	Hasil Penelitian
4	Algoritma Fuzzy String Matching untuk Pencocokan String	(Aditya Dirgandhavi 2018)	Levensthein Distance sejauh ini menutupi kesalahan yang dapat diperbuat oleh algoritma KMP, BM, dan RegEx pada banyak kasus. Namun tidak menutup kemungkinan jika menggunakan Levensthein Distance saja dapat membuat sebuah kesalahan yang mungkin bisa ditutupi jika menggunakan KMP, BM, dan RegEx
5	Chatbot Menggunakan Metode Fuzzy String Matching Sebagai Virtual Assistant Pada Pusat Layanan Informasi Akademik	(Noor Fauzy 2019)	Penerapan metode Fuzzy String Matching dalam pembuatan aplikasi chatbot ini sangat membantu terlebih pada pencarian kata-kata yang memiliki penulisan berbeda namun memiliki makna yang sama.
6	Analisa Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis Pada PT. Telkominfra	(Wibowo, Faisah & Devianto 2022)	Sistem informasi yang dikembangkan dalam penelitian ini dapat melakukan pre-processing data, analisis data dengan EDA hingga menampilkan hasil visualisasi dalam bentuk grafik, sehingga dapat memberikan informasi kepada manajemen terkait kontrak mana yang perlu di evaluasi.
7	Deteksi Kemiripan Artikel Melalui Keywords dengan Metode Fuzzy String Matching dalam Natural Language Processing	(Rumapea 2021)	Dari pengujian data yang dilakukan diperoleh hasil kemiripan yang bervariasi karena kata yang benar sama ditambahkan dengan kata yang mirip dengan kata kunci pada huruf awal dan akhir serta panjang kata tersebut

No	Judul	Penulis dan Tahun	Hasil Penelitian
8	Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining	(Wahyuni, Arifiyanti & Kustyani 2019)	EDA dapat dipergunakan untuk membantu tahapan pra proses dalam data mining (klasifikasi) dengan cara menampilkan keberadaan missing value dan juga outlier. EDA juga dapat mengoptimalkan pengetahuan mengenai data, yang dapat digunakan untuk memperkaya pemahaman atas analisis data
9	Visualisasi Data Penjualan dan Produksi PT Nitto Alam Indonesia Periode 2014-2018	(Dessy A & Johan S 2018)	Dengan menggunakan metode Visual Data Mining (VDM) dan metode Exploratory Data Analysis (EDA) penelitian ini telah berhasil menghasilkan dashboard.
10	Implementasi Algoritma Levenshtein Distance Dalam Mendeteksi Plagiarisme	(Adawiyah & Saragih 2022)	Algoritma jarak Levenshtein memberikan rumus untuk menentukan nilai kemiripan yang membantu pengguna untuk memastikan dokumen mana yang merupakan plagiarisme.

Beberapa penelitian sebelumnya yang telah dituangkan dalam Tabel 2. 1 Tabel Penelitian Terkait memperlihatkan bagaimana manfaat penggunaan metode analisis yang menghasilkan kajian terhadap data untuk pengambilan keputusan. Penelitian yang dilakukan (Wibowo, Faisah & Devianto 2022) menyatakan bahwa sistem informasi yang dikembangkan dalam penelitian ini dapat melakukan pre-processing data, analisis data dengan EDA hingga menampilkan hasil visualisasi dalam bentuk grafik, sehingga dapat memberikan informasi kepada manajemen terkait kontrak mana yang perlu di evaluasi. Ditambah penelitian (Dessy A & Johan S 2018) yang

menggunakan metode Visual Data Mining (VDM) dan metode Exploratory Data Analysis (EDA) telah berhasil menghasilkan dashboard. Hal tersebut selaras dengan penelitian ini, berfokus pada Exploratory Data Analysis dan *Fuzzy String Matching* serta pembuatan *dashboard* untuk visualisasi dengan tujuan untuk memperoleh informasi/*insight* dari hasil eksplorasi data dengan menggunakan data dari aplikasi *chatbot*. Berdasarkan hal tersebut, dilakukan pengumpulan dan pengolahan data perilaku penghuni perumahan pada masa pandemi penghuni perumahan bumi patra untuk diproses menggunakan Exploratory Data Analysis (EDA), dan dilanjutkan dengan pencocokan string oleh Fuzzy String Matching, diakhir membuat dashboard menggunakan tableau yang diharapkan dapat berguna sesuai dengan kebutuhan tim satgas CMT dalam menjaga keamanan dan kesehatan masyarakatnya.

2.3 Keterbaruan Penelitian

Tabel 2. 2 Tabel Kebaruan Penelitian

No	Judul	Penulis dan Tahun	Metode		Algoritma	Visualisasi
			Exploratory Data Analysis (EDA)	Fuzzy String Matching	Levenhstein Distance	
1	Exploratory Data Analysis pada Kasus COVID-19 di Indonesia Menggunakan HiveQL dan Hadoop Environment	(Fahrudin et al. 2020)	√			

No	Judul	Penulis dan Tahun	Metode		Algoritma	Visualisasi
			Exploratory Data Analysis (EDA)	Fuzzy String Matching	Levenhstein Distance	
2	Exploratory Data Analysis terhadap Kepadatan Penumpang Kereta Rel Listrik	(Samosir et al. 2021)	√			
3	Analisis Big Data dengan Metode Exploratory Data Analysis (EDA) dan Metode Visualisasi Menggunakan Jupyter Notebook	(Radhi et al. 2021)	√			
4	Algoritma Fuzzy String Matching untuk Pencocokan String	(Aditya Dirgandhavi 2018)		√	√	
5	Chatbot Menggunakan Metode Fuzzy String Matching Sebagai Virtual Assistant Pada Pusat Layanan Informasi Akademik	(Noor Fauzy 2019)		√	√	
6	Analisa Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis Pada PT. Telkominfra	(Wibowo, Faisah & Devianto 2022)	√			
7	Deteksi Kemiripan Artikel Melalui Keywords dengan Metode Fuzzy String Matching dalam Natural Language Processing	(Rumapea 2021)		√	√	
8	Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining	(Wahyuni, Arifiyanti & Kustyani 2019)	√			

No	Judul	Penulis dan Tahun	Metode		Algoritma	Visualisasi
			Exploratory Data Analysis (EDA)	Fuzzy String Matching	Levenhstein Distance	
9	Visualisasi Data Penjualan dan Produksi PT Nitto Alam Indonesia Periode 2014-2018	(Dessy A & Johan S 2018)	√			√
10	Implementasi Algoritma Levenshtein Distance Dalam Mendeteksi Plagiarisme	(Adawiyah & Saragih 2022)		√		
11	Penerapan Metode Analisa Fuzzy String Matching Terhadap Data Perilaku Penghuni Perumahan Pada Masa Pandemi Dalam Upaya Memvisualisasi Data Menggunakan Tableau. (Studi Kasus: Perumahan Bumi Patra)	(Raisya, 2023)	√	√	√	√

Berdasarkan Tabel 2. 2 Tabel Kebaruan Penelitian, belum ada yang menerapkan *Fuzzy String Matching* pada data mining khususnya sebelumnya menggunakan metode *Exploratory Data Analysis*. Oleh karena itu, penelitian ini akan membangun *Exploratory Data Analysis* yang kemudian nantinya akan menggunakan metode *Fuzzy String Matching* untuk melakukan proses pencocokan string agar membuat datanya lebih mudah dipahami dan lebih mudah untuk dilakukan visualisasi, dan diharapkan dapat bermanfaat bagi seluruh pihak yang terkait.