

UAS :

Buat essay mengenai unstructure data , menjelaskan pa itu unstructure data? apa perbedaan dengan structure data? apa saja contohnya ?

Buat dalam 800 kata (diluar daftar pustaka) dalam format *.docx, doc, pdf

Jangan lupa Daftar pustaka

Terimakasih

ESSAY

MATA KULIAH DATA BASE

Struktur Data dan Unstruktur Data



Disusun

OLEH

Nanda S. Prawira

192420056

MAGISTER TEKNIK INFORMATIKA

UNIVERSITAS BINA DARMA

1. Pendahuluan

Kuantitas data telah berkembang dengan sangat pesat setiap harinya baik dari dalam perusahaan maupun luar perusahaan. Data – data tersebut sangat lah penting untuk dimanfaatkan dan di analisa sehingga dapat digunakan dalam inteligensi bisnis maupun dalam setiap pengambilan keputusan yang strategis dan taktis.

Data diklasifikasikan dalam 2 jenis, yaitu data terstruktur dan data tidak terstruktur. Data terstruktur direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya. Sedangkan data tidak terstruktur di representasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun diintegrasikan dengan sumber data lain.

Besarnya data yang tersimpan di dalam sebuah perusahaan berkembang sangat cepat tiap harinya. Kemampuan untuk mengakses dan menganalisa data tersebut dalam pembuatan keputusan yang cepat dan cerdas menjadi kunci kesuksesan sebuah perusahaan. Banyak perusahaan yang terus berkembang seiring dengan berputarnya waktu, sehingga menghasilkan informasi yang heterogen dari data yang terdistribusi di berbagai sumber. Data tersebut disimpan dalam lokasi, sistem, format dan skema yang berbeda dan memberikan tantangan dalam penggunaan maupun integrasinya. Dalam pengambilan keputusan yang efektif dan taktis, diperlukan kumpulan metodologi, proses, arsitektur dan teknologi yang mengubah data mentah menjadi informasi yang bermakna yang disebut Inteligensi Bisnis.

Sering kali kita mendengar kata data terstruktur (*structured*) dan juga Data Tidak Terstruktur (*Unstructured*). Namun, bagi sebagian orang ada yang belum mengerti apa arti sebenarnya dari kedua data tersebut. Data terstruktur adalah data yang berada dalam satu tempat baik berbentuk sebuah file termasuk data yang berada dalam database ataupun spreadsheet. Data terstruktur adalah yang membuat model data. Contohnya adalah data CRM, Industry Research Data dan lain sebagainya.

Sedangkan Data tidak terstruktur adalah data yang tidak mudah diklasifikasi dan dimasukkan kedalam sebuah kotak dengan rapi. Contohnya adalah foto, gambar grafis, streaming instrument data, webpages, pdf, Power Point presentations, konten blog dan lain sebagainya.

2. Struktur Data Dan Unstruktur Data

a. Struktur data

Merupakan Data yang berada dalam suatu field dari suatu record atau file, termasuk data yang berada di dalam database relasional. Structured data dapat berupa: structured data : numeric, character, date, dan time. Struktur data direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya, Data terstruktur tersedia sebesar 20% dari seluruh data, serta direpresentasikan dalam bentuk relasi yang mudah dipetakan dan disimpan dalam database relasional. Data terstruktur biasanya disimpan dengan skema yang terdefinisi sehingga mudah untuk dilakukan *query*, dianalisa, dan diintegrasikan dengan data terstruktur lainnya. (Afifanto, 2015)

Menurut (Napitupulu, 2010) Struktur Data bermakna pengaturan atau hubungan data di dalam suatu system. Struktur data mempelajari bagaimana pengaturan atau hubungan data di susun di memori utama computer agar *space* di memori dapat dilakukan secara optimal, proses pencarian dan pengambilan data Kembali bisa cepat, dan dapat di olah sesuai tujuan. Sehingga struktur data sangat erat kaitannya dengan membuat algoritmaynag tentunya dilakukan oleh perancangan program. Makna utama yang didapatkan adalah melatih logika bagaimana memecahkan masalah yang rumit dengan bantuan program.

b. Unstructured Data / Data Tidak Terstruktur

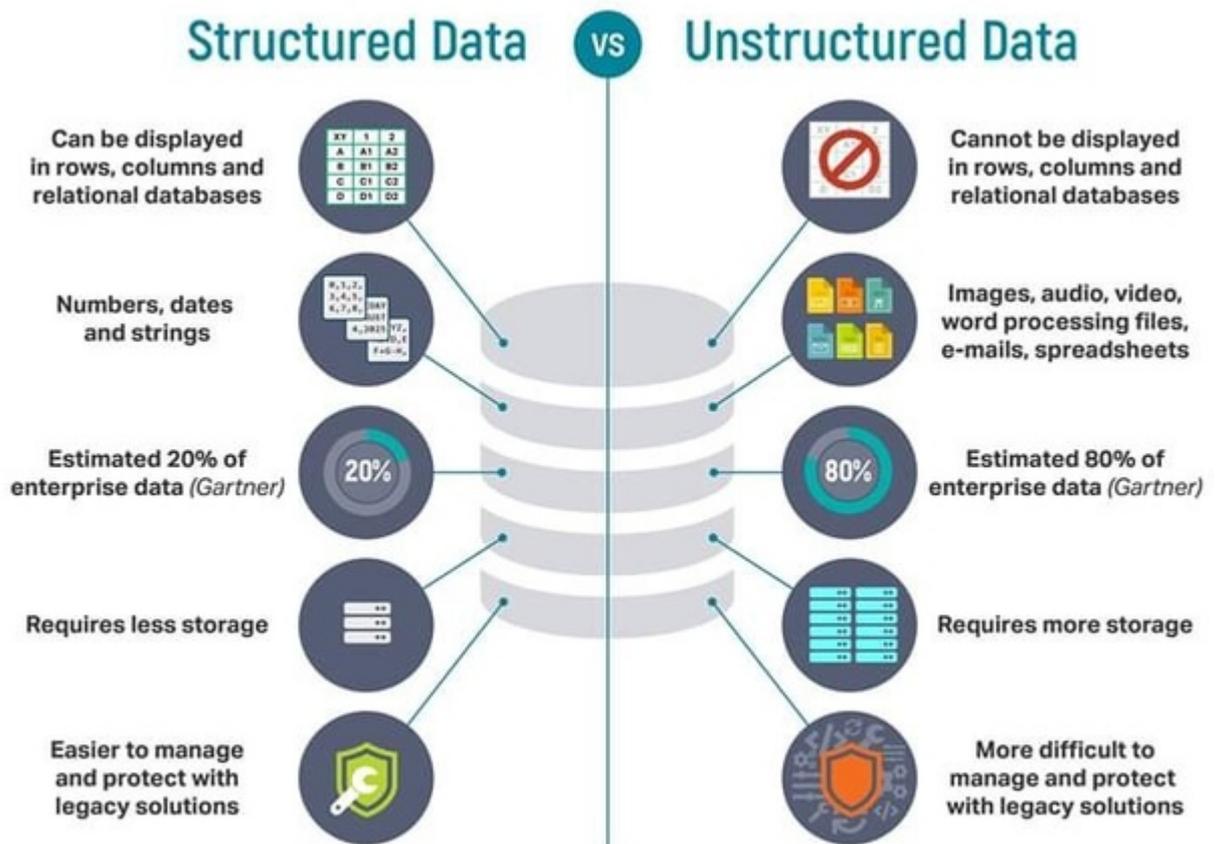
Merupakan informasi yang tidak memiliki model data yang telah ditentukan atau tidak terorganisir dengan cara yang telah ditentukan. Unstructured data biasanya berupa text document, dan mungkin juga berisi data seperti tanggal, angka, dan fakta lainnya. Unstructured data dapat berupa: text dokumen, email, video, audio, dan data dari sosial media. Data tidak terstruktur direpresentasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun diintegrasikan dengan sumber data lain, data tidak terstruktur tersedia sebanyak 80% dari seluruh data yang direpresentasikan. Namun di balik itu semua, informasi tersembunyi yang tersimpan dalam data tidak terstruktur bisa sangat berperan dalam pengambilan keputusan. Sehingga apabila dalam pengambilan keputusan menggunakan integrasi data terstruktur dan data tidak

terstruktur bisa menambah nilai yang signifikan serta kebenaran yang sejati bagi perusahaan/organisasi.

c. Perbedaan Struktur Data dan Unstruktur Data

Data terstruktur (structured) adalah data yang berada dalam satu tempat baik berbentuk sebuah file termasuk data yang berada dalam database maupun spreadsheet. Data terstruktur adalah yang membuat model data. Contohnya adalah data CRM, Industry Research Data dan lain sebagainya. Data terstruktur biasanya dijalankan untuk mengakses database yang disebut dengan Structured Query Language atau yang lebih dikenal dengan SQL. Sedangkan data tidak terstruktur (unstructured) adalah data yang tidak mudah diklasifikasi dan dimasukkan kedalam sebuah kotak dengan rapi. Contohnya adalah foto, gambar grafis, streaming instrument data, webpages, PDF, Power Point presentations, konten blog dan lain sebagainya.

Yang paling terlihat jelas adalah bahwa data terstruktur itu bisa ditampilkan dalam bentuk tabel sementara data tidak terstruktur tidak bisa. Untuk lebih detail perbedaan bisa lihat ini:



C. Contoh Data Terstruktur dan data tidak terstruktur

studi kasus tentang mendapatkan informasi yang lebih aktual dan berguna dari data terstruktur dari BPS dengan data tidak terstruktur dari media massa online, SINDONEWS, tentang informasi ekspor dan impor nonmigas.

Prosedur Pengambilan *Data* terstruktur BPS tentang nilai ekspor nonmigas bulan Mei 2015 yang berupa tabel di-*capture* dan dicatat nilainya, sehingga didapat angka mentah nilai ekspor nonmigas.

Sedangkan untuk data tidak terstruktur yang diambil dari salah satu berita pada koran SINDONEWS yaitu berita berjudul *Wow! Butuh Tujuh Tahun Pisang RI tembus Jepang*^[6]. Artikel berita tersebut berisikan tentang informasi yang lebih detail mengenai ekspor pisang Indonesia ke negara Jepang, seperti total ekspor pisang dari Indonesia ke Jepang, total konsumsi buah impor Jepang, supermarket yang menjual buah-buah impor, dsb.

Analisis Data tidak terstruktur yang berupa artikel berita dari SINDONEWS pertama kali diubah formatnya terlebih dahulu menjadi data terstruktur dengan teknik *text tagging* dan *annotation*. Sedangkan data terstruktur berupa data mentah ekspor nonmigas Indonesia bulan Mei 2015. Kemudian kedua jenis data, terstruktur dan tidak terstruktur, dilakukan proses ekstrak, transform maupun muat dengan alat ETL agar bisa diintegrasikan dan kemudian disimpan ke dalam CDW berupa informasi aktual.

Hasil yang tersimpan dalam ETL tersebut digunakan sebagai pembangunan inteligensi bisnis yang dapat digunakan sebagai landasan pendukung dalam pengambilan keputusan yang strategis dan taktis. Selain itu, inteligensi bisnis dapat diterapkan untuk tujuan bisnis seperti perkiraan, analitis, pelaporan perusahaan, kolaborasi serta manajemen pengetahuan.

Hasil Informasi yang bersifat *real-time* tentang suatu produk kita maupun kompetitor sangat krusial jika kita tidak bisa menganalisisnya secara bijak untuk kepentingan perusahaan kita. Seorang pengambil keputusan harus bisa menyerap dan menganalisa informasi yang tersedia dalam jumlah yang sangat besar yang muncul setiap saat. Untuk menjaga persaingan antar perusahaan, sebuah perusahaan harus sadar dan waspada akan perubahan trend pasar, kebijakan kompetitor, produk terbaru kompetitor, perubahan manajemen, penggabungan dan akuisisi perusahaan yang diterbitkan dalam media cetak/online seperti koran, majalah, dan website.

Artikel berita harian berupa kumpulan teks yang tersusun rapi dalam beberapa paragraf. Untuk mengumpulkan data yang tidak terstruktur tersebut tidaklah mudah dan butuh waktu cukup lama untuk melakukan review dan analisa. Kualitas dari sebuah keputusan yang strategis dan taktis terletak pada kualitas masukan informasi tersebut. Sehingga sangatlah penting untuk menganalisa kualitas informasi seb sebaik mungkin dalam rentang waktu yang terbatas.

Kesipulannya Menghilangkan pembatas antara data terstruktur dan tidak terstruktur berdampak pada cara perusahaan dalam memperlakukan dan mengolah datanya. Memang secara alami, data tidak terstruktur tersebut sangat sulit untuk diekstraks dan diintegrasikan dengan data terstruktur. Namun output dari integrasi data terstruktur dan tidak terstruktur tersebut memberikan manfaat yang besar bagi perusahaan/organisasi.

Seperti yang telah diperlihatkan pada pembahasan, teknologi ini bisa membantu perusahaan dalam mengambil keputusan yang strategis dan taktis dalam inteligensi bisnis dengan memanfaatkan data dari berbagai sumber baik itu data terstruktur maupun tidak terstruktur.

Daftar Pustaka

Afifanto, C. (2015). Integrasi Data Terstruktur dan Tidak Terstruktur dalam Sistem Inteligensi Bisnis. *Jurnal Komputasi Statistik*.

Napitupulu, N. (2010). *Strukter data dan algoritma*. Medan: USU Pres.

<https://blog.wowrack.co.id/2015/04/apa-itu-structured-data-dan.html>

<http://edomuhardosyafri.blogspot.com/2014/12/apakah-big-data-what-is-big-data.html>

<https://id.quora.com/Apa-perbedaan-antara-data-terstruktur-dan-tidak-terstruktur>

UNSTRUCTURE DATA
MENGENAL DETAIL UNSTRUCTURE DATA

Essay

Rani Okta Felani

192420048



PROGRAM PASCASARJANA
MEGISTER TEHNIK INFORMATIKA
UNIVERSITAS BINA DARMA PALEMBANG
TAHUN 2020

UNSTRUCTURE DATA

I. Pendahuluan

Kuantitas data telah berkembang dengan sangat pesat setiap harinya baik dari dalam perusahaan maupun luar perusahaan. Data-data tersebut sangatlah penting untuk dimanfaatkan dan di analisa sehingga dapat digunakan dalam inteligensi bisnis maupun dalam setiap pengambilan keputusan yang strategis dan taktis.

Data diklasifikasikan dalam 2 jenis, yaitu data terstruktur dan data tidak terstruktur. Data terstruktur direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya. Sedangkan data tidak terstruktur direpresentasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun diintegrasikan dengan sumber data lain. Data terstruktur dan tidak terstruktur apabila diintegrasikan dalam menganalisa suatu permasalahan akan memberikan pemahaman dan solusi yang lebih lengkap dan tepat sasaran. Namun, bukanlah hal yang mudah untuk mengintegrasikan kedua jenis data tersebut.

II. Pembahasan

Unstructure data atau data tidak terstruktur merupakan kelompok data yang tidak memiliki struktur atau aturan sesuai dengan format khusus, sehingga untuk menjadikan data terstruktur membutuhkan usaha, tools, dan waktu yang lebih maksimal. data tersebut di dapat dari URL log, media sosial, blog, video, audio, dan email. Jika 20 % tersedia data terstruktur di perusahaan maka tersedia 80 % data tidak terstruktur, karena masih sangat banyak data yang kita temui saat ini merupakan data tidak terstruktur dan teknologi akan menganalisanya secara manual.

Data tidak terstruktur merupakan domain penelitian yang menarik untuk diteliti. Banyaknya sumber data yang dapat diambil dengan mudah menjadi salah satu penyebabnya. Salah satu contoh data tidak terstruktur adalah berita online. Tahapan pre-processing diperlukan untuk menyiapkan data tidak terstruktur agar dapat diolah pada proses klasifikasi. Klasifikasi Rocchio adalah algoritma klasifikasi yang menggabungkan antara TF-IDF dan cosine similarity. Rekeyasa

web yang menggunakan klasifikasi occhio pada data tidak terstruktur dilakukan dengan menggunakan bahasa pemrograman PHP dan basis data MySQL. Kinerja hasil web diukur dengan menggunakan metode Black Box Testing. Pengujian dilakukan dengan menggunakan lima kriteria tes berdasarkan fungsionalitas web. Hasil pengujian menunjukkan bahwa web yang dihasilkan berfungsi sesuai dengan spesifikasi yang telah ditentukan. Adapun pengujian keakuratan klasifikasi Rocchio dilakukan dengan menggunakan perhitungan precision dan recall. Hasil perhitungan menunjukkan nilai precision sebesar 72,01%, dan recall sebesar 67,73%.

Dalam melakukan penelitian ini diperlukan perencanaan agar penelitian yang dilakukan dapat berjalan dengan baik, sistematis, dan efektif. a. Data tidak terstruktur (unstructured data) Data adalah serangkaian fakta atau symbol yang menerangkan sesuatu benda maupun kejadian yang terjadi dalam kehidupan sehari-hari (Hu & Liu, 2012). Data dapat dikelompokkan menjadi data terstruktur dan data tidak terstruktur. Data terstruktur (structured data) adalah data yang direpresentasikan ke dalam bentuk tabel atau relasi yang biasanya ditempatkan dalam suatu basis data. Adapun data tidak terstruktur (unstructured data) adalah data yang tidak dapat dengan mudah diinterpretasikan sebelum diolah terlebih dahulu. Contoh data tidak terstruktur adalah dokumen teks, multimedia, dan sebagainya. Data tidak terstruktur jika diolah sedemikian rupa bisa menjadi data terstruktur sesuai dengan kebutuhan (Aggarwal, 2012). b. Document Text Mining Document text mining juga merupakan analisis teks yang didefinisikan sebagai teknik pemodelan mesin pembelajaran untuk menemukan pengetahuan yang tersembunyi dari sebuah sumber data tekstual (Aggarwal, 2012). Gambar 1. Proses Penggalian Data Pada Dokumen Teks Pengetahuan yang didapatkan dapat digunakan untuk keperluan intelijen bisnis, analisis data, penelitian dan penyelidikan. Penggalian data pada dokumen teks dibutuhkan untuk mengetahui polapola yang terkandung di dalamnya, sehingga didapatkan informasi yang dibutuhkan.

Ada berbagai topik dalam penelitian penggalian data teks, di antaranya klasifikasi teks atau kategorisasi teks, clustering teks, summarization, dan sistem rekomendasi (Hu & Liu, 2012). Gambar 1 menunjukkan proses kerja penggalian

data pada metode document text mining, yang dimulai dari pengolahan artikel dengan menghilangkan kata-kata yang tidak penting, kemudian dilakukan pembobotan kata-kata yang penting. Setelah itu, data teks bisa digunakan untuk proses klasifikasi. c. Klasifikasi Rocchio Klasifikasi adalah salah satu metode pembelajaran data mining kategori supervised learning. Artinya, penentuan golongan sudah dilakukan terlebih dahulu. Klasifikasi adalah penentuan sebuah record data baru untuk dikelompokkan ke salah satu dari beberapa kategori atau kelas yang telah ditentukan sebelumnya (Manning, Raghavan, & Schuetze, 2009). Proses klasifikasi adalah proses pembelajaran suatu fungsi tujuan atau (f) yang memetakan tiap himpunan

Data Tidak Terstruktur, Data Semi Terstruktur dan Data Terstruktur

Data dapat dibagi menjadi 3 (tiga) kategori yaitu data tidak terstruktur, data semi terstruktur dan data terstruktur. Data tidak terstruktur berisi obyek atau dokumen yang tidak mempunyai format khusus baik ukuran maupun tipenya bebas. Berbeda dengan data terstruktur yang ada di database, setiap elemen data harus mengacu pada format yang telah ditetapkan.

Data tidak terstruktur

Data tidak terstruktur biasanya ada pada teks, file, video, email, laporan (report), presentasi power point, pesan suara (voice mail), memo, dan citra. Data dapat berbentuk dalam tipe apapun tanpa perlu mengikuti format, aturan atau alur tertentu. Sebagai contoh data yang ditampilkan pada halaman web termasuk data tidak terstruktur dan tidak beraturan, dan data pencarian dari Google Search merupakan contoh data tidak terstruktur yang tidak mempunyai format khusus.

Data Semi terstruktur

Data semi terstruktur adalah data yang memiliki struktur namun belum sepenuhnya terstruktur. Tipe data ini berisi komponen tidak terstruktur yang tersusun pada data terstruktur.

Data semi terstruktur diatur dalam bentuk entitas. Entitas yang sama dijadikan satu kelompok, namun dapat berbeda atributnya. Contoh dari data semi terstruktur adalah daftar riwayat hidup (CV). Individu pertama dapat memiliki

pengalaman kerja , berbeda dengan individu kedua yang memiliki pengalaman dalam melakukan penelitian dan berbeda pula dengan orang ketiga yang memiliki pengalaman dalam hal pengajaran. Sering dijumpai pula perbedaan dalam penulisan CV lebih dari satu macam.

Data Terstruktur

Data terstruktur mendefinisikan tipe data atribut, yang ada pada record maupun tuple. Setiap record memiliki field yang sama. Data ada dalam entitas dan entitas yang sama dikelompokkan menjadi satu menggunakan relasi atau kelas. Entitas dalam kelompok yang sama memiliki atribut yang sama pula. Deskripsi untuk semua entitas dalam skema memiliki kesamaan format.

Penggunaan data terstruktur semakin banyak dilakukan , sejak basisdata relasional ada untuk mengatur data dengan ukuran sangat besar. Sistem yang ada saat ini, seperti CRM (Customer Relationship Management), ERP (Enterprise Resource Planning) dan CMS (Content Managemnt System) merupakan data terstruktur untuk pemodelan data.

III. Kesimpulan

Data tidak terstruktur adalah kelompok data yang tidak memiliki struktur atau aturan sesuai dengan format khusus, sehingga untuk menjadikan data terstruktur membutuhkan usaha, tools, dan waktu yang lebih maksimal .contoh data tidak terstruktur adalah data yang ditampilkan pada halaman web dan data yang ada pada Google Search.

Data terstruktur adalah Data yang dapat disimpan, diakses dan diproses dalam bentuk format tetap disebut sebagai data ‘terstruktur’. selama periode waktu tertentu. Contohnya, Data yang tersimpan dalam RDBMS (*Relational Database Management System*)

DAFTAR PUSTAKA

Buneman et al, 1996, *A Query language and optimization techniques for unstructure data*, pdf sigmood record, Vol. 25 No.2, Pages. 505-516

C. Yusainy, A. Chawa, S.Kholifah, 2017. *Big data dan pemanfaatannya dalam berbagai sektor*, Media Informatika, Vol.16, No.2, Hal. 107-112

M.Abidin, S.Adji, 2012, *Analisa perfomance propeller B. Series dengan pendekatan Structure dan unstructure meshing*, Jurnal Teknik ITS, Vol.1, No.1 Hal. G241-G246

Maryanto budi, 2017, *Big Data dan pemanfaatannya dalam berbagai sektor*, Media Informatika, Vol.16 , No.2 , Hal.107-112

Wahyunita laili, 2019, *Rekayasa Web Klasifikasi pada Data Tidak Terstruktur*, Jurnal Komunika, Vol.8 No.2 ISSN : 2579-5899, Hal. 88-95

<https://www.dummies.com/programming/big-data/engineering/unstructured-data-in-a-big-data-environment/>

<https://www.researchgate.net/publication/279480483>

<https://wahyudisetiawan.wordpress.com/2009/08/03/data-tidak-terstruktur-semite-rstruktur-dan-terstruktur/>

<http://edomuhardosyafri.blogspot.com/2014/12/apakah-big-data-what-is-big-data.html>

UNSTRUCTURE DATA
MENGENAL DETAIL UNSTRUCTURE DATA

Essay

Rani Okta Felani

192420048



PROGRAM PASCASARJANA
MEGISTER TEHNIK INFORMATIKA
UNIVERSITAS BINA DARMA PALEMBANG
TAHUN 2020

UNSTRUCTURE DATA

I. Pendahuluan

Kuantitas data telah berkembang dengan sangat pesat setiap harinya baik dari dalam perusahaan maupun luar perusahaan. Data-data tersebut sangatlah penting untuk dimanfaatkan dan di analisa sehingga dapat digunakan dalam inteligensi bisnis maupun dalam setiap pengambilan keputusan yang strategis dan taktis.

Data diklasifikasikan dalam 2 jenis, yaitu data terstruktur dan data tidak terstruktur. Data terstruktur direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya. Sedangkan data tidak terstruktur direpresentasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun diintegrasikan dengan sumber data lain. Data terstruktur dan tidak terstruktur apabila diintegrasikan dalam menganalisa suatu permasalahan akan memberikan pemahaman dan solusi yang lebih lengkap dan tepat sasaran. Namun, bukanlah hal yang mudah untuk mengintegrasikan kedua jenis data tersebut.

II. Pembahasan

Unstructure data atau data tidak terstruktur merupakan kelompok data yang tidak memiliki struktur atau aturan sesuai dengan format khusus, sehingga untuk menjadikan data terstruktur membutuhkan usaha, tools, dan waktu yang lebih maksimal . data tersebut di dapat dari URL log, media sosial, blog, video, audio, dan email. Jika 20 % tersedia data terstruktur di perusahaan maka tersedia 80 % data tidak terstruktur, karena masih sangat banyak data yang kita temui saat ini merupakan data tidak terstruktur dan teknologi akan menganalisanya secara manual.

Data tidak terstruktur merupakan domain penelitian yang menarik untuk diteliti. Banyaknya sumber data yang dapat diambil dengan mudah menjadi salah satu penyebabnya. Salah satu contoh data tidak terstruktur adalah berita online. Tahapan pre-processing diperlukan untuk menyiapkan data tidak terstruktur agar dapat diolah pada proses klasifikasi. Klasifikasi Rocchio adalah algoritma klasifikasi yang menggabungkan antara TF-IDF dan cosine similarity. Rekayasa web yang menggunakan klasifikasi occhio pada data tidak terstruktur dilakukan dengan menggunakan bahasa pemrograman PHP dan basis data MySQL. Kinerja hasil web diukur dengan menggunakan metode Black Box Testing. Pengujian dilakukan dengan menggunakan lima kriteria tes berdasarkan fungsionalitas web. Hasil pengujian menunjukkan bahwa web yang dihasilkan berfungsi sesuai dengan spesifikasi yang telah ditentukan. Adapun pengujian keakuratan klasifikasi Rocchio dilakukan dengan menggunakan perhitungan precision dan recall. Hasil perhitungan menunjukkan nilai precision sebesar 72,01%, dan recall sebesar 67,73%,.

Dalam melakukan penelitian ini diperlukan perencanaan agar penelitian yang dilakukan dapat berjalan dengan baik, sistematis, dan efektif. a. Data tidak terstruktur (unstructured data) Data adalah serangkaian fakta atau symbol yang menerangkan sesuatu benda maupun kejadian yang terjadi dalam kehidupan sehari-hari (Hu & Liu, 2012). Data dapat dikelompokkan menjadi data terstruktur dan data tidak terstruktur. Data terstruktur (structured data) adalah data yang direpresentasikan ke dalam bentuk tabel atau relasi yang biasanya ditempatkan dalam suatu basis data. Adapun data tidak terstruktur (unstructured data) adalah

data yang tidak dapat dengan mudah diinterpretasikan sebelum diolah terlebih dahulu. Contoh data tidak terstruktur adalah dokumen teks, multimedia, dan sebagainya. Data tidak terstruktur jika diolah sedemikian rupa bisa menjadi data terstruktur sesuai dengan kebutuhan (Aggarwal, 2012).

b. Document Text Mining

Document text mining juga merupakan analisis teks yang didefinisikan sebagai teknik pemodelan mesin pembelajaran untuk menemukan pengetahuan yang tersembunyi dari sebuah sumber data tekstual (Aggarwal, 2012). Gambar 1. Proses Penggalian Data Pada Dokumen Teks Pengetahuan yang didapatkan dapat digunakan untuk keperluan intelijen bisnis, analisis data, penelitian dan penyelidikan. Penggalian data pada dokumen teks dibutuhkan untuk mengetahui polapola yang terkandung di dalamnya, sehingga didapatkan informasi yang dibutuhkan.

Ada berbagai topik dalam penelitian penggalian data teks, di antaranya klasifikasi teks atau kategorisasi teks, clustering teks, summarization, dan sistem rekomendasi (Hu & Liu, 2012). Gambar 1 menunjukkan proses kerja penggalian data pada metode document text mining, yang dimulai dari pengolahan artikel dengan menghilangkan kata-kata yang tidak penting, kemudian dilakukan pembobotan kata-kata yang penting. Setelah itu, data teks bisa digunakan untuk proses klasifikasi.

c. Klasifikasi Rocchio

Klasifikasi adalah salah satu metode pembelajaran data mining kategori supervised learning. Artinya, penentuan golongan sudah dilakukan terlebih dahulu. Klasifikasi adalah penentuan sebuah record data baru untuk dikelompokkan ke salah satu dari beberapa kategori atau kelas yang telah ditentukan sebelumnya (Manning, Raghavan, & Schuetze, 2009). Proses klasifikasi adalah proses pembelajaran suatu fungsi tujuan atau (f) yang memetakan tiap himpunan

Data Tidak Terstruktur, Data Semi Terstruktur dan Data Terstruktur

Data dapat dibagi menjadi 3 (tiga) kategori yaitu data tidak terstruktur, data semi terstruktur dan data terstruktur. Data tidak terstruktur berisi obyek atau dokumen yang tidak mempunyai format khusus baik ukuran maupun tipenya bebas. Berbeda dengan data terstruktur yang ada di database, setiap elemen data harus mengacu pada format yang telah ditetapkan.

Data tidak terstruktur

Data tidak terstruktur biasanya ada pada teks, file, video, email, laporan (report), presentasi power point, pesan suara (voice mail), memo, dan citra. Data dapat berbentuk dalam tipe apapun tanpa perlu mengikuti format, aturan atau alur tertentu. Sebagai contoh data yang ditampilkan pada halaman web termasuk data tidak terstruktur dan tidak beraturan, dan data pencarian dari Google Search merupakan contoh data tidak terstruktur yang tidak mempunyai format khusus.

Data Semi terstruktur

Data semi terstruktur adalah data yang memiliki struktur namun belum sepenuhnya terstruktur. Tipe data ini berisi komponen tidak terstruktur yang tersusun pada data terstruktur.

Data semi terstruktur diatur dalam bentuk entitas. Entitas yang sama dijadikan satu kelompok, namun dapat berbeda atributnya. Contoh dari data semi terstruktur adalah daftar riwayat hidup (CV). Individu pertama dapat memiliki pengalaman kerja, berbeda dengan individu kedua yang memiliki pengalaman dalam melakukan penelitian dan berbeda pula dengan orang ketiga yang memiliki pengalaman dalam hal pengajaran. Sering dijumpai pula perbedaan dalam penulisan CV lebih dari satu macam.

Data Terstruktur

Data terstruktur mendefinisikan tipe data atribut, yang ada pada record maupun tuple. Setiap record memiliki field yang sama. Data ada dalam entitas dan entitas yang sama dikelompokkan menjadi satu menggunakan relasi atau kelas. Entitas dalam kelompok yang sama memiliki atribut yang sama pula. Deskripsi untuk semua entitas dalam skema memiliki kesamaan format.

Penggunaan data terstruktur semakin banyak dilakukan, sejak basis data relasional ada untuk mengatur data dengan ukuran sangat besar. Sistem yang ada saat ini, seperti CRM (Customer Relationship Management), ERP (Enterprise Resource Planning) dan CMS (Content Management System) merupakan data terstruktur untuk pemodelan data.

III. Kesimpulan

Data tidak terstruktur adalah kelompok data yang tidak memiliki struktur atau aturan sesuai dengan format khusus, sehingga untuk menjadikan data terstruktur membutuhkan usaha, tools, dan waktu yang lebih maksimal. Contoh data tidak terstruktur adalah data yang ditampilkan pada halaman web dan data yang ada pada Google Search.

Data terstruktur adalah Data yang dapat disimpan, diakses dan diproses dalam bentuk format tetap disebut sebagai data 'terstruktur'. selama periode waktu tertentu. Contohnya, Data yang tersimpan dalam RDBMS (*Relational Database Management System*)

DAFTAR PUSTAKA

Buneman et al, 1996, *A Query language and optimization techniques for unstructure data*, pdf sigmood record, Vol. 25 No.2, Pages. 505-516

C. Yusainy, A. Chawa, S.Kholifah, 2017. *Big data dan pemanfaatannya dalam berbagai sektor*, Media Informatika, Vol.16, No.2, Hal. 107-112

M. Abidin, S.Adji, 2012, *Analisa perfomance propeller B. Series dengan pendekatan Structure dan unstructure meshing*, Jurnal Teknik ITS, Vol.1, No.1 Hal. G241-G246

Maryanto budi, 2017, *Big Data dan pemanfaatannya dalam berbagai sektor*,
Media Informatika, Vol.16 , No.2 , Hal.107-112

Wahyunita laili, 2019, *Rekayasa Web Klasifikasi pada Data Tidak Terstruktur*,
Jurnal Komunika, Vol.8 No.2 ISSN : 2579-5899, Hal. 88-95

<https://www.dummies.com/programming/big-data/engineering/unstructured-data-in-a-big-data-environment/>

<https://www.researchgate.net/publication/279480483>

<https://wahyudisetiawan.wordpress.com/2009/08/03/data-tidak-terstruktur-semite-rstruktur-dan-terstruktur/>

<http://edomuhardosyafri.blogspot.com/2014/12/apakah-big-data-what-is-big-data.html>

UNSTRUCTURE DATA
MENGENAL DETAIL UNSTRUCTURE DATA

Essay

Rani Okta Felani

192420048



PROGRAM PASCASARJANA
MEGISTER TEHNIK INFORMATIKA
UNIVERSITAS BINA DARMA PALEMBANG
TAHUN 2020

UNSTRUCTURE DATA

I. Pendahuluan

Kuantitas data telah berkembang dengan sangat pesat setiap harinya baik dari dalam perusahaan maupun luar perusahaan. Data-data tersebut sangatlah penting untuk dimanfaatkan dan di analisa sehingga dapat digunakan dalam inteligensi bisnis maupun dalam setiap pengambilan keputusan yang strategis dan taktis.

Data diklasifikasikan dalam 2 jenis, yaitu data terstruktur dan data tidak terstruktur. Data terstruktur direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya. Sedangkan data tidak terstruktur direpresentasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun diintegrasikan dengan sumber data lain. Data terstruktur dan tidak terstruktur apabila diintegrasikan dalam menganalisa suatu permasalahan akan memberikan pemahaman dan solusi yang lebih lengkap dan tepat sasaran. Namun, bukanlah hal yang mudah untuk mengintegrasikan kedua jenis data tersebut.

II. Pembahasan

Unstructure data atau data tidak terstruktur merupakan kelompok data yang tidak memiliki struktur atau aturan sesuai dengan format khusus, sehingga untuk menjadikan data terstruktur membutuhkan usaha, tools, dan waktu yang lebih maksimal. data tersebut di dapat dari URL log, media sosial, blog, video, audio, dan email. Jika 20 % tersedia data terstruktur di perusahaan maka tersedia 80 % data tidak terstruktur, karena masih sangat banyak data yang kita temui saat ini merupakan data tidak terstruktur dan teknologi akan menganalisanya secara manual.

Data tidak terstruktur merupakan domain penelitian yang menarik untuk diteliti. Banyaknya sumber data yang dapat diambil dengan mudah menjadi salah satu penyebabnya. Salah satu contoh data tidak terstruktur adalah berita online. Tahapan pre-processing diperlukan untuk menyiapkan data tidak terstruktur agar dapat diolah pada proses klasifikasi. Klasifikasi Rocchio adalah algoritma klasifikasi yang menggabungkan antara TF-IDF dan cosine similarity. Rekeyasa

web yang menggunakan klasifikasi occhio pada data tidak terstruktur dilakukan dengan menggunakan bahasa pemrograman PHP dan basis data MySQL. Kinerja hasil web diukur dengan menggunakan metode Black Box Testing. Pengujian dilakukan dengan menggunakan lima kriteria tes berdasarkan fungsionalitas web. Hasil pengujian menunjukkan bahwa web yang dihasilkan berfungsi sesuai dengan spesifikasi yang telah ditentukan. Adapun pengujian keakuratan klasifikasi Rocchio dilakukan dengan menggunakan perhitungan precision dan recall. Hasil perhitungan menunjukkan nilai precision sebesar 72,01%, dan recall sebesar 67,73%,.

Dalam melakukan penelitian ini diperlukan perencanaan agar penelitian yang dilakukan dapat berjalan dengan baik, sistematis, dan efektif. a. Data tidak terstruktur (unstructured data) Data adalah serangkaian fakta atau symbol yang menerangkan sesuatu benda maupun kejadian yang terjadi dalam kehidupan sehari-hari (Hu & Liu, 2012). Data dapat dikelompokkan menjadi data terstruktur dan data tidak terstruktur. Data terstruktur (structured data) adalah data yang direpresentasikan ke dalam bentuk tabel atau relasi yang biasanya ditempatkan dalam suatu basis data. Adapun data tidak terstruktur (unstructured data) adalah data yang tidak dapat dengan mudah diinterpretasikan sebelum diolah terlebih dahulu. Contoh data tidak terstruktur adalah dokumen teks, multimedia, dan sebagainya. Data tidak terstruktur jika diolah sedemikian rupa bisa menjadi data terstruktur sesuai dengan kebutuhan (Aggarwal, 2012). b. Document Text Mining Document text mining juga merupakan analisis teks yang didefinisikan sebagai teknik pemodelan mesin pembelajaran untuk menemukan pengetahuan yang tersembunyi dari sebuah sumber data tekstual (Aggarwal, 2012). Gambar 1. Proses Penggalian Data Pada Dokumen Teks Pengetahuan yang didapatkan dapat digunakan untuk keperluan intelijen bisnis, analisis data, penelitian dan penyelidikan. Penggalian data pada dokumen teks dibutuhkan untuk mengetahui polapola yang terkandung di dalamnya, sehingga didapatkan informasi yang dibutuhkan.

Ada berbagai topik dalam penelitian penggalian data teks, di antaranya klasifikasi teks atau kategorisasi teks, clustering teks, summarization, dan sistem rekomendasi (Hu & Liu, 2012). Gambar 1 menunjukkan proses kerja penggalian

data pada metode document text mining, yang dimulai dari pengolahan artikel dengan menghilangkan kata-kata yang tidak penting, kemudian dilakukan pembobotan kata-kata yang penting. Setelah itu, data teks bisa digunakan untuk proses klasifikasi. c. Klasifikasi Rocchio Klasifikasi adalah salah satu metode pembelajaran data mining kategori supervised learning. Artinya, penentuan golongan sudah dilakukan terlebih dahulu. Klasifikasi adalah penentuan sebuah record data baru untuk dikelompokkan ke salah satu dari beberapa kategori atau kelas yang telah ditentukan sebelumnya (Manning, Raghavan, & Schuetze, 2009). Proses klasifikasi adalah proses pembelajaran suatu fungsi tujuan atau (f) yang memetakan tiap himpunan

Data Tidak Terstruktur, Data Semi Terstruktur dan Data Terstruktur

Data dapat dibagi menjadi 3 (tiga) kategori yaitu data tidak terstruktur, data semi terstruktur dan data terstruktur. Data tidak terstruktur berisi obyek atau dokumen yang tidak mempunyai format khusus baik ukuran maupun tipenya bebas. Berbeda dengan data terstruktur yang ada di database, setiap elemen data harus mengacu pada format yang telah ditetapkan.

Data tidak terstruktur

Data tidak terstruktur biasanya ada pada teks, file, video, email, laporan (report), presentasi power point, pesan suara (voice mail), memo, dan citra. Data dapat berbentuk dalam tipe apapun tanpa perlu mengikuti format, aturan atau alur tertentu. Sebagai contoh data yang ditampilkan pada halaman web termasuk data tidak terstruktur dan tidak beraturan, dan data pencarian dari Google Search merupakan contoh data tidak terstruktur yang tidak mempunyai format khusus.

Data Semi terstruktur

Data semi terstruktur adalah data yang memiliki struktur namun belum sepenuhnya terstruktur. Tipe data ini berisi komponen tidak terstruktur yang tersusun pada data terstruktur.

Data semi terstruktur diatur dalam bentuk entitas. Entitas yang sama dijadikan satu kelompok, namun dapat berbeda atributnya. Contoh dari data semi terstruktur adalah daftar riwayat hidup (CV). Individu pertama dapat memiliki

pengalaman kerja , berbeda dengan individu kedua yang memiliki pengalaman dalam melakukan penelitian dan berbeda pula dengan orang ketiga yang memiliki pengalaman dalam hal pengajaran. Sering dijumpai pula perbedaan dalam penulisan CV lebih dari satu macam.

Data Terstruktur

Data terstruktur mendefinisikan tipe data atribut, yang ada pada record maupun tuple. Setiap record memiliki field yang sama. Data ada dalam entitas dan entitas yang sama dikelompokkan menjadi satu menggunakan relasi atau kelas. Entitas dalam kelompok yang sama memiliki atribut yang sama pula. Deskripsi untuk semua entitas dalam skema memiliki kesamaan format.

Penggunaan data terstruktur semakin banyak dilakukan , sejak basisdata relasional ada untuk mengatur data dengan ukuran sangat besar. Sistem yang ada saat ini, seperti CRM (Customer Relationship Management), ERP (Enterprise Resource Planning) dan CMS (Content Managemnt System) merupakan data terstruktur untuk pemodelan data.

III. Kesimpulan

Data tidak terstruktur adalah kelompok data yang tidak memiliki struktur atau aturan sesuai dengan format khusus, sehingga untuk menjadikan data terstruktur membutuhkan usaha, tools, dan waktu yang lebih maksimal .contoh data tidak terstruktur adalah data yang ditampilkan pada halaman web dan data yang ada pada Google Search.

Data terstruktur adalah Data yang dapat disimpan, diakses dan diproses dalam bentuk format tetap disebut sebagai data ‘terstruktur’. selama periode waktu tertentu. Contohnya, Data yang tersimpan dalam RDBMS (*Relational Database Management System*)

DAFTAR PUSTAKA

Buneman et al, 1996, *A Query language and optimization techniques for unstructure data, pdf sigmood record*, Vol. 25 No.2, Pages. 505-516

C. Yusainy, A. Chawa, S.Kholifah, 2017. *Big data dan pemanfaatannya dalam berbagai sektor*, Media Informatika, Vol.16, No.2, Hal. 107-112

M.Abidin, S.Adji, 2012, *Analisa perfomance propeller B. Series dengan pendekatan Structure dan unstructure meshing*, Jurnal Teknik ITS, Vol.1, No.1 Hal. G241-G246

Maryanto budi, 2017, *Big Data dan pemanfaatannya dalam berbagai sektor*, Media Informatika, Vol.16 , No.2 , Hal.107-112

Wahyunita laili, 2019, *Rekayasa Web Klasifikasi pada Data Tidak Terstruktur*, Jurnal Komunika, Vol.8 No.2 ISSN : 2579-5899, Hal. 88-95

<https://www.dummies.com/programming/big-data/engineering/unstructured-data-in-a-big-data-environment/>

<https://www.researchgate.net/publication/279480483>

<https://wahyudisetiawan.wordpress.com/2009/08/03/data-tidak-terstruktur-semite-rstruktur-dan-terstruktur/>

<http://edomuhardosyafri.blogspot.com/2014/12/apakah-big-data-what-is-big-data.html>

ESSAY

MATA KULIAH DATA BASE

Struktur Data dan Unstruktur Data



Disusun

OLEH :

SUWANI

192420049

MAGISTER TEKNIK INFORMATIKA

UNIVERSITAS BINA DARMA

Tahun 2020

I. PENDAHULUAN

Kuantitas data telah berkembang dengan sangat pesat setiap harinya baik dari dalam perusahaan maupun luar perusahaan. Data – data tersebut sangat lah penting untuk dimanfaatkan dan di analisa sehingga dapat digunakan dalam inteligensi bisnis maupun dalam setiap pengambilan keputusan yang strategis dan taktis. Data diklasifikasikan dalam 2 jenis, yaitu data terstruktur dan data tidak terstruktur. Data terstruktur direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya. Sedangkan data tidak terstruktur di representasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun diintegrasikan dengan sumber data lain.

Besarnya data yang tersimpan di dalam sebuah perusahaan berkembang sangat cepat tiap harinya. Kemampuan untuk mengakses dan menganalisa data tersebut dalam pembuatan keputusan yang cepat dan cerdas menjadi kunci kesuksesan sebuah perusahaan. Banyak perusahaan yang terus berkembang seiring dengan berputarnya waktu, sehingga menghasilkan informasi yang heterogen dari data yang terdistribusi di berbagai sumber. Data tersebut disimpan dalam lokasi, sistem, format dan skema yang berbeda dan memberikan tantangan dalam penggunaan maupun integrasinya. Dalam pengambilan keputusan yang efektif dan taktis, diperlukan kumpulan metodologi, proses, arsitektur dan teknologi yang mengubah data mentah menjadi informasi yang bermakna yang disebut Inteligensi Bisnis.

Sering kali kita mendengar kata data terstruktur (*structured*) dan juga Data Tidak Terstruktur (*Unstructured*). Namun, bagi sebagian orang ada yang belum mengerti apa arti sebenarnya dari kedua data tersebut. Data terstruktur adalah data yang berada dalam satu tempat baik berbentuk sebuah file termasuk data yang berada dalam database ataupun spreadsheet. Data terstruktur adalah yang membuat model data. Contohnya adalah data CRM, Industry Research Data dan lain sebagainya.

Sedangkan Data tidak terstruktur adalah data yang tidak mudah diklasifikasi dan dimasukkan kedalam sebuah kotak dengan rapi. Contohnya adalah

foto, gambar grafis, streaming instrument data, webpages, pdf, Power Point presentations, konten blog dan lain sebagainya.

II. PEMBAHASAN

a. Struktur data

Merupakan Data yang berada dalam suatu field dari suatu record atau file, termasuk data yang berada di dalam database relasional. Structured data dapat berupa: structured data : numeric, character, date, dan time. Struktur data direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya, Data terstruktur tersedia sebesar 20% dari seluruh data, serta direpresentasikan dalam bentuk relasi yang mudah dipetakan dan disimpan dalam database relasional. Data terstruktur biasanya disimpan dengan skema yang terdefinisi sehingga mudah untuk dilakukan *query*, dianalisa, dan diintegrasikan dengan data terstruktur lainnya.(Afifanto, 2015)

Data terstruktur biasanya dijalankan untuk mengakses database yang disebut dengan Structured Query Language atau yang lebih dikenal dengan SQL. Berikut adalah ilustrasi tentang SQL. Anda ingin melihat daftar buku yang memiliki harga diatas 1000 dan di urutkan sesuai dengan judul, Anda bisa menggunakan :

```
SELECT * FROM buku WHERE harga > 1000 ORDER BY judul;
```

Dari situ Anda akan mendapatkan data dari database sesuai permintaan Anda yaitu memunculkan buku yang memiliki harga diatas dan otomatis diurutkan sesuai judul.

b. Unstructured Data / Data Tidak Terstruktur

Merupakan informasi yang tidak memiliki model data yang telah ditentukan atau tidak terorganisir dengan cara yang telah ditentukan. Unstructured data biasanya berupa text document, dan mungkin juga berisi data seperti tanggal, angka, dan fakta lainnya. Unstructured data dapat berupa: text dokumen, email, video, audio, dan data dari sosial media. Data tidak terstruktur direpresentasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun

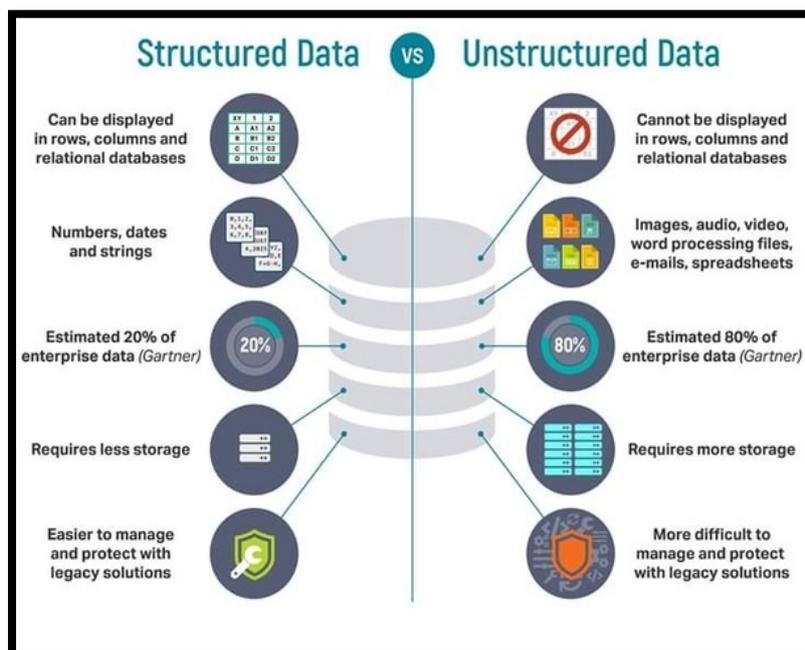
diintegrasikan dengan sumber data lain, data tidak terstruktur tersedia sebanyak 80% dari seluruh data yang direpresentasikan. Namun di balik itu semua, informasi tersembunyi yang tersimpan dalam data tidak terstruktur bias sangat berperan dalam pengambilan keputusan. Sehingga apabila dalam pengambilan keputusan menggunakan integrasi data terstruktur dan data tidak terstruktur bias menambah nilai yang signifikan serta kebenaran yang sejati bagi perusahaan /organisasi.

c. Perbedaan Struktur Data dan Unstruktur Data

Perbedaan Antara stuktur data dan unstruktur data adalah :

Struktur data merupakan Kelompok data yang memiliki tipe data, format, dan struktur yang telah terdefinisi. Sumber datanya dapat berupa data transaksional, OLAP data, tradisional RDBMS, file CSV, spreadsheets, sedangkan unstruktur data Kelompok data tekstual dengan format tidak menentu atau tidak memiliki struktur melekat, sehingga untuk menjadikannya data terstruktur membutuhkan usaha, tools, dan waktu yang lebih. Data ini dihasilkan oleh aplikasi – aplikasi internet, seperti data URL log, media sosial, e-mail, blog, video, dan audio.

Yang paling terlihat jelas adalah bahwa data terstruktur itu bias ditampilkan dalam bentuk table sementara data tidak terstruktur tidak bisa. Untuk lebih detail perbedaan dapat di lihat pada gambar di bawah ini.



d. Contoh Unstruktur Data

Contoh Unstruktur data yaitu penyimpanan pada google drive.

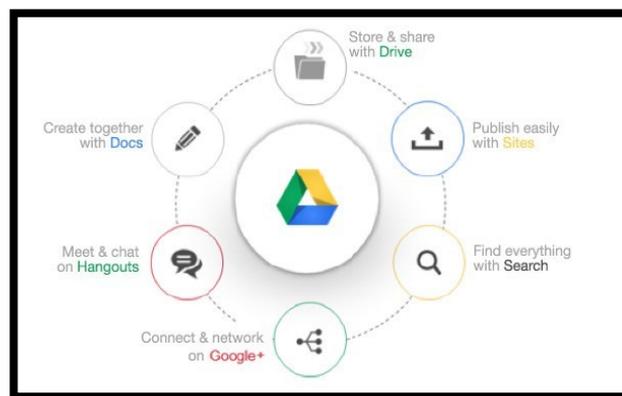
Google drive adalah media penyimpanan data online (*daring*) berbasis cloud atau Internet yang dikembangkan dan diperkenalkan oleh Google pada tanggal 24 April 2012. Pada dasarnya layanan Google Drive sama seperti *cloud storage* lain semacam dropbox atau One Drive.

Aplikasi ini tersedia dalam bentuk desktop ataupun smartphone. Pada beberapa jenis ponsel Google Drive atau GDrive menjadi aplikasi bawaan yang terinstal secara otomatis di sistem.

Di google drive kita dapat menyimpan file – file berupa dokumen, gambar, audio ataupun video dengan kapasitas gratis sebesar 15 Gigabyte. Jika ingin lebih dari itu maka kita harus *upgrade account*.

Sementaritu, kita dapat menyimpan, mengedit dan mengambil file dari google drive dengan catatan perangkat computer atau smartphone harus terhubung dengan jaringan internet.

Fungsi Google Drive



boxcryptor.com

Google drive mempunyai beberapa fungsi utama untuk para penggunanya. Beberapa fungsi tersebut diantaranya:

1. Untuk Menyimpan File

Fungsi utama Google Drive adalah sebagai media penyimpanan data dengan Google Drive kita dapat menyimpan data dalam bentuk dokumen gambar atau apapun pada media penyimpanan online tanpa harus menyediakan *storage* atau media penyimpanan *offline* seperti ram ataupun hardisk.

2. Berbagi File

Dengan menggunakan Google Drive kita dapat berbagi file secara langsung tanpa perlu melakukan konfigurasi konfigurasi yang rumit kita dapat berbagi file secara langsung dengan mengundang pengguna lain melalui Link yang kita berikan atau pun kita undang melalui alamat email.

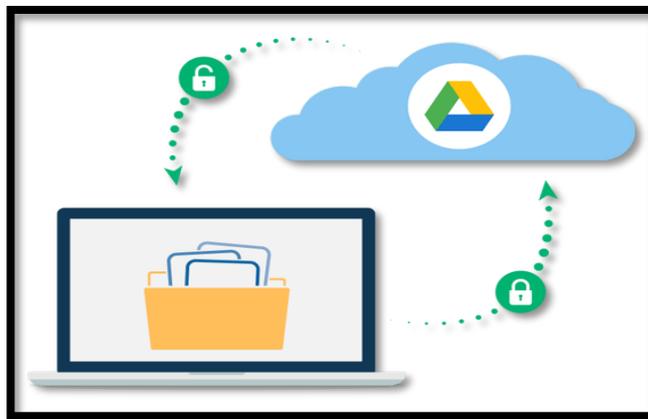
3. Edit File

Kita dapat melakukan *editing* data langsung tanpa harus mengunduhnya. GDrive mendukung fitur edit data teks yang terhubung dengan google spreadsheet. Aplikasi tersebut berfungsi mirip seperti ms word atau excel hanya saja berbentuk online tanpa harus melakukan instalasi software.

4. Melakukan Back Up Data

Sebetulnya ini sama seperti poin pertama, Google drive digunakan untuk memback up data baik itu data yang ada di computer ataupun perangkat lunak lainnya.

Manfaat Google Drive



boxcryptor.com

1. Data Tersimpan Aman

Menyimpan data di google drive lebih aman dibandingkan kita menyimpan data di komputer ataupun telepon pintar, karena alat semacam itu rawan mengalami kerusakan ataupun hilang karena pencurian. Walaupun seperti itu, tetap saja kita perlu berhati – hati karena pencurian data pun bias saja terjadi pada perangkat *cloud storage* semacam GDrive ini.

2. Data Dapat Digunakan Kapan pun dan Dimanapun

Kita dapat menggunakan GDrive dimanapun dan kapanpun sesuai keinginan kita sepanjang *device* terhubung ke Internet. Ini akan memudahkan kita. Misalnya, ketika akan melakukan presentasi tiba – tiba laptop tertinggal di rumah, sedangkan data kita ada di laptop tersebut. Apabila kita menyimpan data tersebut di GDrive kita tinggal meminjam laptop kerabat kemudian masuk ke akun GDrive kita dan datapun bias langsung kita buka atau unduh untuk kemudian digunakan sebagai materi presentasi.

III. KESIMPULAN

Dari pembahasan di atas dapat disimpulkan bahwa data terstruktur adalah Data terstruktur biasanya dijalankan untuk mengakses database yang disebut dengan Structured Query Language atau yang lebih dikenal dengan SQL. Berikut adalah ilustrasi tentang SQL. Anda ingin melihat daftar buku yang memiliki harga diatas 1000 dan di urutkan sesuai dengan judul.

Sedangkan Data tidak terstruktur adalah data yang tidak mudah diklasifikasi dan dimasukkan kedalam sebuah kotak dengan rapi. Contohnya adalah foto, gambar grafis, streaming instrument data, webpages, pdf, Power Point presentations, konten blog dan lain sebagainya.

Data tidak terstruktur dapat di contohkan pada penyimpanan google drive / Gdrive.

DAFTAR PUSTAKA

Afifanto, C. Integrasi Data Terstruktur dan Tidak Terstruktur dalam Sistem Integrasi Bisnis. Jurnal Komputasi Statistik. Sekolah Tinggi Ilmu Statistik. Jakarta : 2015

Maryanto, budi. 2017. Jurnal Big Data Dan Pemanfaatannya Dalam Berbagai Sektor. Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI: Bandung. Media Informatika Vol.16 No.2

<https://salamadian.com/pengertian-google-drive/>

<https://blog.wowrack.co.id/2015/04/apa-itu-structured-data-dan.html>

<http://edomuhardosyafri.blogspot.com/2014/12/apakah-big-data-what-is-big-data.html>

<https://id.quora.com/Apa-perbedaan-antara-data-terstruktur-dan-tidak-terstruktur>

ESSAY

MATA KULIAH ADVANCED DATABASE



UNSTRUCTURE DATA

NAMA : THEO VHALDINO

NIM : 192420058

MAGISTER TEKNIK INFORMATIKA

UNIVERSITAS BINA DARMA

PALEMBANG

I. Pendahuluan

Di dunia digital modern sekarang ini, penggunaan data yang tepat dapat memainkan peran besar dalam kesuksesan bisnis. Kumpulan data meledak dengan kecepatan yang semakin cepat, sehingga mengumpulkan dan menganalisis data hingga efek maksimum sangat penting. Perusahaan dan bisnis banyak berfokus pada pengumpulan data untuk memastikan mereka mendapatkan wawasan yang berharga. Memahami struktur data adalah kunci untuk membuka nilainya.

"Struktur" data mengacu pada cara tertentu mengatur dan menyimpannya dalam database atau gudang sehingga dapat diakses dan dianalisis. Berbagai jenis informasi lebih cocok untuk disimpan dalam format terstruktur atau tidak terstruktur.

1. Data Terstruktur (Structure Data)

Seperti kata "terstruktur" menyarankan, ini adalah data yang sangat terorganisir dan diformat dengan rapi. Data terstruktur diatur dalam format tabular (mis. Baris dan kolom) dan ada hubungan antara baris dan kolom yang berbeda. Karena itu, sangat terorganisir dan diformat serta mudah untuk menyimpan, memproses, dan mengakses. Ini dapat bekerja dengan mudah dengan sebagian besar model analitik standar. Sebagian besar alat BI juga tahu cara bekerja dengannya, memungkinkan pengguna untuk mengoptimalkan sumber daya dari perspektif teknis. Ini juga membutuhkan lebih sedikit ruang penyimpanan. Beberapa contoh data terstruktur adalah file Excel, Google Sheets, dan Sistem Manajemen DataBase tradisional (DBMS).

2. Data Tidak Terstruktur (Unstructured Data)

Data tidak terstruktur adalah data yang tidak diatur dengan cara apa pun yang telah ditentukan. Ini bisa berupa teks, angka, tanggal, atau BLOB (Binary Large

Objects). Penyimpangan dan disorganisasi dalam data yang tidak terstruktur membuatnya sulit untuk ditangani dan dipahami.

Beberapa poin penting yang perlu diketahui tentang data tidak terstruktur:

- 1) Sekitar 80% dari data di seluruh dunia tidak terstruktur
- 2) sulit untuk memproses dan mengatur
- 3) Cenderung kompleks di alam
- 4) Memberikan lebih banyak kebebasan untuk analisis
- 5) Membutuhkan lebih banyak penyimpanan
- 6) Jenis media yang kaya (gambar, video, audio) juga dapat dianalisis dengan teknologi canggih
- 7) Beberapa contoh termasuk data teks, komentar media sosial, dokumen, transkripsi panggilan telepon, berbagai file log seperti log server, log sensor, gambar, audio, video, dll.

3. Data Semi Terstruktur (Semi-Structured Data)

Data semi-terstruktur adalah gabungan dari data terstruktur dan tidak terstruktur. Ini memiliki beberapa kerangka kerja organisasi tetapi tidak memiliki struktur lengkap yang diperlukan untuk masuk dalam database relasional. Data semi-terstruktur memiliki struktur yang menggambarkan diri sendiri yang berisi tag atau atribut untuk memisahkan berbagai entitas dalam data.

Poin-poin penting yang perlu diingat tentang data semi-terstruktur:

- 1) Jatuh di bawah tajuk data tidak terstruktur, tetapi memiliki beberapa organisasi tingkat rendah (masih kurang dari database relasional)
- 2) Dapat dipaksa ke dalam format tabel yang bermanfaat dan mudah digunakan
- 3) Contoh data semi-terstruktur termasuk XML, JSON, Email, DB NoSQL, pelacakan acara, dan halaman web

4. Ilustrasi Perbedaan Data Terstruktur, Data Tidak Terstruktur, dan Data Semi Terstruktur

Untuk dengan mudah memahami perbedaan antara klasifikasi data, mari gunakan analogi ini untuk mengilustrasikan. Saat mewawancarai suatu pekerjaan, katakanlah ada tiga klasifikasi wawancara yang berbeda: terstruktur, semi-terstruktur dan tidak terstruktur.

Dalam wawancara terstruktur, pewawancara mengikuti naskah ketat yang didefinisikan oleh departemen sumber daya manusia dan diikuti untuk setiap kandidat. Bentuk lain dari wawancara adalah wawancara tidak terstruktur. Dalam wawancara tidak terstruktur, sepenuhnya tergantung pada pewawancara untuk menentukan pertanyaan dan urutan pertanyaan yang akan ditanyakan (atau bahkan jika mereka akan diminta) untuk setiap kandidat. Wawancara semi-terstruktur mengambil elemen-elemen dari klasifikasi wawancara terstruktur dan tidak terstruktur. Ia menggunakan konsistensi dan elemen kuantitatif yang diperbolehkan dengan wawancara terstruktur tetapi menawarkan kebebasan untuk menyesuaikan berdasarkan keadaan yang lebih sejalan dengan wawancara tidak terstruktur.

Jadi, untuk data, data terstruktur mudah diatur dan mengikuti format yang kaku; tidak terstruktur adalah informasi yang kompleks dan seringkali kualitatif yang tidak mungkin direduksi atau disusun dalam database relasional dan data semi-terstruktur memiliki elemen keduanya.

Properties	Structured Data	Unstructured Data
Characteristics	Predefined format Easy to search	Raw/native format Difficult to search
Data Types	Defined data types	Many varied data types
Store in	Relational databases Data warehouses	NoSQL databases Data warehouses Data lakes
Generated by	Humans or machines	Humans or machines
Flexibility	Not flexible; schema-dependent	Flexible; not schema-dependent
Data percentage	Estimated 20% of data	Estimated 80% of data
Examples	Excel, Google Sheets, SQL, customer data, phone records, transaction history	Text data, social media comments, phone calls transcriptions, various logs files, images, audio, video

Tabel Perbedaan Data Terstruktur dan Data Tidak Terstruktur

II. Pembahasan

Salah satu studi kasus yang menggunakan data tidak terstruktur dilakukan penelitian oleh Laili Wahyunita tentang Rekayasa Web Klasifikasi pada Data Tidak Terstruktur. Proses penelitian yang digunakan dalam pengolahan data tidak terstruktur pertama Pembersihan data dilakukan terlebih dahulu dengan tujuan untuk menghilangkan isi yang dianggap tidak perlu untuk diolah dan mengubah isi dokumen menjadi huruf kecil semua. Contoh isi yang dihapus adalah label-label HTML. Proses pembersihan data juga bertujuan untuk menghilangkan tanda baca, misalnya titik (.) dan menyeragamkan huruf menjadi huruf kecil semua. Setelah data dibersihkan, proses selanjutnya adalah prapemrosesan data (pre-processing data).

Setelah dilakukan Tahapan pre-processing diperlukan untuk menyiapkan data tidak terstruktur agar dapat diolah pada proses klasifikasi. Klasifikasi Rocchio adalah algoritma klasifikasi yang menggabungkan antara TF-IDF dan cosine similarity.

Rekayasa web yang menggunakan klasifikasi occhio pada data tidak terstruktur dilakukan dengan menggunakan bahasa pemrograman PHP dan basis data MySQL. Kinerja hasil web diukur dengan menggunakan metode Black Box Testing. Pengujian dilakukan dengan menggunakan lima kriteria tes berdasarkan fungsionalitas web. Hasil pengujian menunjukkan bahwa web yang dihasilkan berfungsi sesuai dengan spesifikasi yang telah ditentukan. Adapun pengujian keakuratan klasifikasi Rocchio dilakukan dengan menggunakan perhitungan precision dan recall. Hasil perhitungan menunjukkan nilai precision sebesar 72,01%, dan recall sebesar 67,73%.

Penelitian ini membuat rekayasa web yang menghasilkan klasifikasi Rocchio pada data tidak terstruktur. Data yang digunakan berupa artikel berita online. Proses pre-processing berhasil membuat data artikel berita menjadi token, yang kemudian dijadikan term. Selanjutnya, dilakukan proses klasifikasi dengan menggunakan algoritma Rocchio. Dari hasil pengujian terhadap fungsionalitas sistem dengan menggunakan metode Black Box Testing tampak bahwa web yang telah direkayasa berfungsi sesuai dengan spesifikasi kebutuhan. Pengujian terhadap performa klasifikasi Rocchio yang dibuat dengan menggunakan perhitungan recall dan precision menunjukkan nilai yang cukup baik, yaitu di atas 50 %. Dengan demikian dapat disimpulkan bahwa algoritma Rocchio juga dapat diterapkan pada proses klasifikasi dokumen untuk data tidak terstruktur.

III. Kesimpulan

Pada pembahasan diatas, dapat disimpulkan bahwa pemanfaatan Data Tidak Terstruktur dapat menghasilkan data yang akurat dengan metode klasifikasi yang dilakukan pada penelitian diatas. Data yang tidak terstruktur tumbuh lebih cepat daripada data terstruktur dan perannya sangat penting dalam memberikan wawasan dan bisa juga menjadi nilai bisnis.

DAFTAR PUSTAKA

<https://blog.westerndigital.com/examples-of-unstructured-data/>

<https://www.sisense.com/blog/understanding-structured-and-unstructured-data/>

<https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/>

Wahyunita, Laili. (2019). Rekayasa Web Klasifikasi Rocchio pada Data Tidak Terstruktur. *Jurnal Komunika : Jurnal Komunikasi, Media dan Informatika*. 8. 88. 10.31504/komunika.v8i2.2016.

ESSAY



UNSTRUCTURE DATA

NAMA : YAYAN CANDRA SUBIDIN
NIM : 192420054
ESSAY : STRUCTURE DATA DAN
UNSTRUCTURE DATA

MAGISTER TEKNIK INFORMATIKA
UNIVERITAS BINA DARMA
PALEMBANG

I. Pendahuluan

Kuantitas data berkembang sangat pesat tiap harinya baik dari dalam maupun luar perusahaan. Sangatlah penting untuk memanfaatkan data-data tersebut untuk dianalisa sehingga bisa digunakan dalam inteligensi bisnis maupun pengambilan keputusan yang strategis dan taktis. Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Data yang dikumpulkan dari setiap variable ditentukan oleh defenisi operasional variable. Metode pengumpulan data dapat dilakukan dengan cara kusioner, observasi, wawancara.

Data diklasifikasikan dalam 2 (dua) jenis, yaitu terstruktur dan tidak terstruktur. Data terstruktur direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya. Sedangkan data tidak terstruktur direpresentasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun diintegrasikan dengan sumber data lain.

Data terstruktur dan tidak terstruktur apabila diintegrasikan dalam menganalisa suatu permasalahan akan memberikan pemahaman dan solusi yang lebih lengkap dan tepat sasaran. Namun, bukanlah hal yang mudah untuk mengintegrasikan kedua jenis data tersebut. Dibutuhkan teknik maupun arsitektur yang tepat untuk mengatasi permasalahan tersebut. Beberapa teknik yang bisa digunakan adalah text tagging dan annotation. Teknik tersebut merupakan teknik yang populer dalam natural processing techniques untuk preprocessing data tidak terstruktur agar bisa dengan mudah diintegrasikan dengan data terstruktur

Paper ini akan menjelaskan unstructure data, perbedaan dengan structure data, dan contohnya unstructured dan structure data agar dapat dipahami sifat, jenis dan manfaat data tersebut.

II. Penjelasan dan Pembahasan

a. Pengertian dan Defenisi data

Menurut bahasa, data merupakan bentuk jamak dari kata *datum* (bahasa latin) yang berarti *sesuatu yang diberikan*. Menurut istilah, pengertian data adalah kumpulan informasi atau keterangan-keterangan yang diperoleh dari pengamatan, informasi itu bisa berupa angka, lambang atau sifat. Dalam kehidupan sehari-hari data berarti suatu pernyataan yang diterima secara apa adanya.

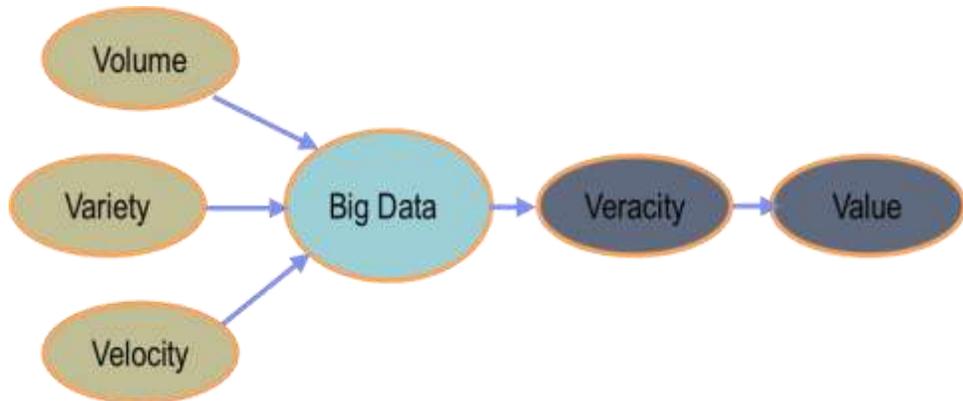
Pengertian data menurut "Nuzulla Agustina" definisi data adalah keterangan mengenai suatu hal yang sudah sering terjadi dan berupa himpunan fakta, angka, grafik tabel, gambar, lambang, kata, huruf, yang menyatakan sesuatu pemikiran, objek, serta kondisi dan situasi.

Menurut "Arikunto (2002)", data merupakan segala fakta dan angka yang dapat dijadikan bahan untuk menyusun suatu informasi, sedangkan informasi adalah hasil pengolahan data yang dipakai untuk suatu keperluan.

Menurut "Slamet Riyadi", data adalah kumpulan informasi yang diperoleh dari hasil suatu pengamatan di mana data dapat berupa angka atau lambang.

b. Jenis data

Seperti yang kita tahu, Big data memiliki beberapa karakteristik, salah satunya adalah *variety* (Variasi). *Variety* merupakan permasalahan yang disebabkan karena adanya keberagaman data, baik itu dari format filenya, maupun struktur isi data tersebut. Data yang diproduksi/dihasilkan kini tidak hanya dalam bentuk data terstruktur, ada yang dalam data semi-terstruktur, dan tidak terstruktur. **Apa bedanya?**



Contoh : 1. Big Data Characteristics

Ilmuwan data membagi data menjadi tiga kelompok besar : data terstruktur (structure data), data semi-terstruktur (middle structure data), dan data tidak terstruktur (unstructured data) dimana pengertiannya yaitu :

1. Data terstruktur (structure data)

Data terstruktur merupakan data yang dapat diproses, disimpan, dan diambil dalam format tetap. Normalnya kita mengenal sebagai data terstruktur yang ada dalam excel atau spreadsheet. Sebagai contoh, data penjualan pada suatu perusahaan.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
530385	85123A	WHITE HANGING HEART T-LIGHT HD	8	01/12/2010 08:26	2,55	17850	United Kingdom
530385	73053	WHITE METAL LANTERN	8	01/12/2010 08:26	8,39	17850	United Kingdom
530385	84406B	CREAM CUPID HEARTS COAT HANGE	8	01/12/2010 08:26	2,75	17850	United Kingdom
530385	84029G	KNITTED UNION FLAG HOT WATER B	8	01/12/2010 08:26	3,39	17850	United Kingdom
530385	84029E	RED WOOLLY HOTTIE WHITE HEART.	8	01/12/2010 08:26	3,39	17850	United Kingdom
530385	22752	SET 7 BABUSHKA NESTING BOXES	2	01/12/2010 08:26	7,85	17850	United Kingdom
530385	22799	GLASS STAR FROSTED T-LIGHT HOLD	8	01/12/2010 08:26	4,25	17850	United Kingdom
530386	22833	HAND WARMER UNION JACK	8	01/12/2010 08:28	1,85	17850	United Kingdom
530386	22852	HAND WARMER RED POLKA DOT	8	01/12/2010 08:28	1,85	17850	United Kingdom
530387	84879	ASSORTED COLOUR BIRD ORNAMENT	32	01/12/2010 08:34	1,69	13047	United Kingdom
530387	22345	POPPY'S PLAYHOUSE BEDROOM	6	01/12/2010 08:34	2,1	13047	United Kingdom
530387	22348	POPPY'S PLAYHOUSE KITCHEN	6	01/12/2010 08:34	2,1	13047	United Kingdom
530387	22349	TELCRAFT PRINCESS CHARLOTTE DC	8	01/12/2010 08:34	5,75	13047	United Kingdom
530387	22338	IVORY KNITTED MUG COSY	8	01/12/2010 08:34	1,85	13047	United Kingdom
530387	84989	BOX OF 8 ASSORTED COLOUR TEASP	4	01/12/2010 08:34	4,25	13047	United Kingdom
530387	22673	BOX OF VINTAGE ROSAW BLOCKS	3	01/12/2010 08:34	4,95	13047	United Kingdom

Contoh : 2. Data Penjualan pada perusahaan

2. Data Semi-Terstruktur

Data semi terstruktur merupakan bentuk data yang berisi format yang disebutkan diatas, yaitu data terstruktur dan tidak terstruktur. Lebih tepatnya, Data Semi-Terstruktur merujuk pada data yang belum diklasifikasikan dibawah repositori tertentu (database), namun mengandung informasi penting yang memisahkan elemen individu dalam data. Contohnya, data dalam bentuk file .json atau .xml.

```
{
  "text": "RT @sengineand: My Single Best... ",
  "created_at": "Fri Apr 15 23:37:26 +0000 2011",
  "retweet_count": 0,
  "id_str": "59037647649259521",
  "entities": {
    "user_mentions": [
      {
        "screen_name": "sengineand",
        "id_str": "1059801",
        "name": "Search Engine Land",
      }
    ],
    "hashtags": [],
    "urls": [
      {
        "url": "http://\sengineand.com/e2QPS1",
        "expanded_url": null
      }
    ]
  },
  "user": {
    "created_at": "Sat Jan 22 18:39:46 +0000 2011",
    "friends_count": 63,
    "id_str": "241622902",
    ...
  },
  "retweeted_status": {
    "text": "My Single Best... ",
    "created_at": "Fri Apr 15 21:40:10 +0000 2011",
    "id_str": "59008136320786432",
    ...
  }
}
```

Contoh :3. Data semi-terstruktur

3. Data tidak terstruktur (unstructured data)

Data tidak terstruktur pada dasarnya adalah segalanya yang tidak memiliki bentuk atau struktur khusus. Data yang tidak terstruktur memiliki struktur internal tetapi tidak terstruktur melalui model atau skema data yang ditentukan sebelumnya. Ini mungkin tekstual atau non-tekstual, dan dihasilkan manusia atau mesin. Data tersebut dapat dikelompokkan sebagai berikut :

- a. Data tidak terstruktur yang dihasilkan manusia meliputi:
 - ✓ **File teks:** Pengolah kata, spreadsheet, presentasi, email, log.
 - ✓ **Email:** Email memiliki beberapa struktur internal berkat metadata-nya, dan kami terkadang menyebutnya sebagai semi-terstruktur . Namun, bidang pesannya tidak terstruktur dan alat analisis tradisional tidak dapat menguraikannya.
 - ✓ **Media Sosial:** Data dari Facebook, Twitter, LinkedIn.
 - ✓ **Situs web:** YouTube, Instagram, situs berbagi foto.
 - ✓ **Data seluler:** Pesan teks, lokasi.
 - ✓ **Komunikasi:** Obrolan, IM, rekaman telepon, perangkat lunak kolaborasi.
 - ✓ **Media:** MP3, foto digital, file audio dan video.
 - ✓ **Aplikasi bisnis:** dokumen MS Office, aplikasi produktivitas.
- b. Data tidak terstruktur khas yang dihasilkan mesin meliputi:
 - ✓ **Citra satelit:** Data cuaca, bentuk tanah, gerakan militer.
 - ✓ **Data ilmiah:** Eksplorasi minyak dan gas, eksplorasi ruang angkasa, citra seismik, data atmosfer.
 - ✓ **Digital surveillance:** Foto dan video pengawasan.
 - ✓ **Data sensor:** Lalu lintas, cuaca, sensor oseanografi.



Contoh :4. Data tidak terstruktur

4. Perbedaan data terstruktur (structure data) dan data tidak terstruktur (unstructured data)

Selain perbedaan yang jelas antara menyimpan dalam satu database dan menyimpan banyak jenis database, perbedaan terbesar adalah kemudahan menganalisis data terstruktur dibanding data tidak terstruktur. Analisis yang matang ada pada data terstruktur, tetapi analisis untuk menghasilkan banyak data tidak terstruktur dapat lebih berkembang.

Pengguna dapat menjalankan pencarian konten sederhana di seluruh data tekstual yang tidak terstruktur. Tetapi kurangnya struktur internal yang tertib mengalahkan tujuan dari alat penambangan data tradisional, dan perusahaan mendapatkan sedikit nilai dari sumber data yang berpotensi berharga seperti media yang kaya, jaringan atau weblog, interaksi pelanggan, dan data media sosial. Meskipun alat analisis data yang tidak terstruktur ada di pasar, tidak ada vendor atau perangkat yang jelas pemenangnya. Dan banyak pelanggan yang enggan berinvestasi dalam alat analitik dengan peta jalan pengembangan yang tidak pasti.

Selain itu, ada jauh lebih banyak data tidak terstruktur daripada terstruktur. Data yang tidak terstruktur membentuk 80% dan lebih banyak data perusahaan, dan tumbuh pada tingkat 55% dan 65% per tahun. Dan tanpa alat untuk menganalisis data yang sangat besar ini, organisasi meninggalkan banyak sekali data berharga di atas meja intelijen bisnis.

	Structured Data	Unstructured Data
Characteristics	<ul style="list-style-type: none">• Pre-defined data models• Usually text only• Easy to search	<ul style="list-style-type: none">• No pre-defined data model• May be text, images, sound, video or other formats• Difficult to search
Resides In	<ul style="list-style-type: none">• Relational databases• Data warehouses	<ul style="list-style-type: none">• Applications• NoSQL databases• Data warehouses• Data lakes
Generated by	Humans or machines	Humans or machines
Typical applications	<ul style="list-style-type: none">• Airline reservation systems• Inventory control• CRM systems• ERP systems	<ul style="list-style-type: none">• Word processing• Presentation software• Email clients• Tools for viewing or editing media
Examples	<ul style="list-style-type: none">• Dates• Phone numbers• Social security numbers• Credit card numbers• Customer names• Addresses• Product names and numbers• Transaction information	<ul style="list-style-type: none">• Text files• Reports• Email messages• Audio files• Video files• Images• Surveillance imagery

Contoh : 5. Perbedaan data terstruktur dan tidak terstruktur

III. Kesimpulan

Data terstruktur dan Data tidak terstruktur : data *terstruktur* terdiri dari tipe data yang terdefinisi jelas yang polanya membuatnya mudah dicari; sementara *data tidak terstruktur* - "segalanya" - terdiri dari data yang biasanya tidak mudah dicari, termasuk format seperti audio, video, dan posting media sosial.

Data terstruktur dan tidak terstruktur apabila diintegrasikan dalam menganalisa suatu permasalahan akan memberikan pemahaman dan solusi yang lebih lengkap dan tepat sasaran. Namun, bukanlah hal yang mudah untuk mengintegrasikan kedua jenis data tersebut. Dibutuhkan teknik maupun arsitektur yang tepat untuk mengatasi permasalahan tersebut. Beberapa teknik yang bisa digunakan adalah text tagging dan annotation. Teknik tersebut merupakan teknik yang populer dalam natural processing techniques untuk preprocessing data tidak terstruktur agar bisa dengan mudah diintegrasikan dengan data terstruktur

Namun, ada ketegangan yang berkembang antara kemudahan analisis pada data terstruktur dibanding analisis yang lebih kerumitan pada data tidak terstruktur. Analisis data terstruktur adalah proses dan teknologi yang matang. Analisis yang matang ada pada data terstruktur, tetapi analisis untuk menghasilkan banyak data tidak terstruktur dapat lebih berkembang, tetapi bukan teknologi yang matang.

Daftar pustaka

1. <https://www.dosenpendidikan.co.id/data-adalah/>
2. https://www.researchgate.net/publication/279480483_Integrasi_Data_Terstruktur_dan_Tidak_Terstruktur_dalam_Sistem_Inteligensi_Bisnis
3. <https://www.maxmanroe.com/vid/teknologi/pengertian-data.html>
4. <https://medium.com/@apriandito/data-terstruktur-semi-struktur-dan-tidak-terstruktur-fc5227cf233b>
5. <https://www.datamation.com/big-data/structured-vs-unstructured-data.html>

ESSAY



UNSTRUCTURE DATA

AL ADRI NOFA GUSANDI

192420053

**MAGISTER TEKNIK INFORMATIKA
UNIVERITAS BINA DARMA
PALEMBANG**

I. PENDAHULUAN

Data tidak terstruktur biasanya ada pada teks, file, video, email, laporan (report), presentasi power point, pesan suara (voice mail), memo, dan citra. Data dapat berbentuk dalam tipe apapun tanpa perlu mengikuti format, aturan atau alur tertentu. Sebagai contoh data yang ditampilkan pada halaman web termasuk data tidak terstruktur dan tidak beraturan.

Data Tidak Terstruktur merupakan data yang tidak memiliki bentuk atau struktur khusus.

Data terstruktur dan tidak terstruktur apabila diintegrasikan dalam menganalisa suatu permasalahan akan memberikan pemahaman dan solusi yang lebih lengkap dan tepat sasaran. Namun, bukanlah hal yang mudah untuk mengintegrasikan kedua jenis data tersebut. Dibutuhkan teknik maupun arsitektur yang tepat untuk mengatasi permasalahan tersebut. Beberapa teknik yang bisa digunakan adalah text tagging dan annotation. Teknik tersebut merupakan teknik yang populer dalam natural processing techniques untuk preprocessing data tidak terstruktur agar bisa dengan mudah diintegrasikan dengan data terstruktur

II. PEMBAHASAN

Ilmuwan data membagi data menjadi tiga kelompok besar : data terstruktur (structure data), data semi-terstruktur (middle structure data), dan data tidak terstruktur (unstructure data) dimana pengertiannya yaitu:

Data terstruktur merupakan data yang yang dapat diproses, disimpan, dan diambil dalam format tetap. Normalnya kita mengenal sebagai data terstruktur yang ada dalam excel atau spreadsheet. Sebagai contoh, data penjualan pada suatu perusahaan

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HO	6	01/12/2010 08:26	2,55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	01/12/2010 08:26	3,39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGE	8	01/12/2010 08:26	2,75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER B	6	01/12/2010 08:26	3,39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01/12/2010 08:26	3,39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	01/12/2010 08:26	7,65	17850	United Kingdom
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDI	6	01/12/2010 08:26	4,25	17850	United Kingdom
536366	22633	HAND WARMER UNION JACK	6	01/12/2010 08:28	1,85	17850	United Kingdom
536366	22632	HAND WARMER RED POLKA DOT	6	01/12/2010 08:28	1,85	17850	United Kingdom
536367	84879	ASSORTED COLOUR BIRD ORNAMEN	32	01/12/2010 08:34	1,69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	01/12/2010 08:34	2,1	13047	United Kingdom
536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	01/12/2010 08:34	2,1	13047	United Kingdom
536367	22749	FELTCRAFT PRINCESS CHARLOTTE DC	8	01/12/2010 08:34	3,75	13047	United Kingdom
536367	22310	IVORY KNITTED MUG COSY	6	01/12/2010 08:34	1,65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTED COLOUR TEASP	6	01/12/2010 08:34	4,25	13047	United Kingdom
536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	01/12/2010 08:34	4,95	13047	United Kingdom

Contoh data terstruktur

Data Tidak Terstruktur merupakan data yang tidak memiliki bentuk atau struktur khusus . Contohnya adalah data yang berformat foto/gambar, video, atau suara.

Karakteristik Data Tidak Terstruktur:

- a) Data tidak sesuai dengan model data atau memiliki struktur apa pun.
- b) Data tidak dapat disimpan dalam bentuk baris dan kolom seperti pada Database
- c) Data tidak mengikuti semantik atau aturan apa pun
- d) Data tidak memiliki format atau urutan tertentu
- e) Data tidak memiliki struktur yang mudah diidentifikasi
- f) Karena kurangnya struktur yang dapat diidentifikasi, itu tidak dapat digunakan oleh program komputer dengan mudah



contoh data tidak terstruktur.

Sementara, Data Semi-Terstruktur merupakan bentuk data yang berisi format yang disebutkan di atas, yaitu data terstruktur dan tidak terstruktur. Lebih tepatnya, Data Semi-Terstruktur merujuk pada data yang belum diklasifikasikan di bawah repositori tertentu (database), namun mengandung informasi penting yang memisahkan elemen individu dalam data. Contohnya, data dalam bentuk file .json atau .xml.

```

"text": "RT @sengineland: My Single Best... ",
"created_at": "Fri Apr 15 23:37:26 +0000 2011",
"retweet_count": 0,
"id_str": "59037647649259521",
"entities": {
  "user_mentions": [{
    "screen_name": "sengineland",
    "id_str": "1059801",
    "name": "Search Engine Land",
  }],
  "hashtags": [],
  "urls": [{
    "url": "http://se.lnd.com/e2QP81",
    "expanded_url": null
  }]
},
"user": {
  "created_at": "Sat Jan 22 18:39:46 +0000 2011",
  "friends_count": 63,
  "id_str": "241622902",
  ...
},
"retweeted_status": {
  "text": "My Single Best... ",
  "created_at": "Fri Apr 15 21:40:10 +0000 2011",
  "id_str": "59008136320786432",
  ...
},

```

Contoh data semi-terstruktur

III. KESIMPULAN

Data terstruktur biasanya dijalankan untuk mengakses database yang disebut dengan Structured Query Language atau yang lebih dikenal dengan SQL. Sedangkan data tidak terstruktur (unstructured) adalah data yang tidak mudah diklasifikasi dan dimasukkan kedalam sebuah kotak dengan rapi. Data tidak terstruktur biasanya ada pada teks, file, video, email, laporan (report), presentasi power point, pesan suara (voice mail), memo, dan citra. Data dapat berbentuk dalam tipe apapun tanpa perlu mengikuti format, aturan atau alur tertentu. Data terstruktur dan tidak terstruktur apabila diintegrasikan dalam menganalisa suatu permasalahan akan memberikan pemahaman dan solusi yang lebih lengkap dan tepat sasaran. Namun, bukanlah hal yang mudah untuk mengintegrasikan kedua jenis data tersebut. Dibutuhkan teknik maupun arsitektur yang tepat untuk mengatasi permasalahan tersebut. Beberapa teknik yang bisa digunakan adalah text tagging dan annotation. Teknik tersebut merupakan teknik yang populer dalam natural processing techniques untuk preprocessing

data tidak terstruktur agar bisa dengan mudah diintegrasikan dengan data terstruktur Namun, ada ketegangan yang berkembang antara kemudahan analisis pada data terstruktur disbanding analisis yang lebih kerumitan pada data tidak terstruktur.

DAFTAR PUSTAKA

https://web.facebook.com/WowrackIndonesia/posts/apa-sebenarnya-perbedaan-structured-dan-unstructured-datadata-terstruktur-struct/10154454119702330/?_rdc=1&_rdr#:~:text=Data%20terstruktur%20biasanya%20dijalankan%20untuk,kedalam%20sebuah%20kotak%20dengan%20rapi.

<https://medium.com/@apriandito/data-terstruktur-semi-struktur-dan-tidak-terstruktur-fc5227cf233b>

<https://blog.wowrack.co.id/2015/04/apa-itu-structured-data-dan.html>

ESAY
UNSTRUCTURE DATA
INTEGRASI DATA TERSTRUKTUR
DAN TIDAK TERSTRUKTUR



OLEH :

ARPA PAUZIAH

192420055

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

UNIVERSITAS BINA DARMA

2020/2021

I. Pendahuluan

Perkembangan dunia teknologi saat ini sangat cepat, sehingga manusia dituntut untuk mengikuti perkembangan. Untuk itu dibutuhkan sumber daya manusia dalam bidang IT yang memiliki kemampuan beradaptasi secara cepat. Dalam dunia teknologi informasi saat ini kebutuhan terhadap data adalah sebuah kunci yang wajib dimiliki pada setiap lini bisnis. Meskipun bisnisnya bahkan tidak berhubungan dengan dunia teknologi, namun data sangat diperlukan dalam melakukan analisis yang nantinya akan digunakan untuk pengambilan keputusan.

Besarnya data yang tersimpan di dalam sebuah perusahaan berkembang sangat cepat tiap harinya. Kemampuan untuk mengakses dan menganalisa data tersebut dalam pembuatan keputusan yang cepat dan cerdas menjadi kunci kesuksesan sebuah perusahaan. Banyak perusahaan yang terus berkembang seiring dengan berputarnya waktu, sehingga menghasilkan informasi yang heterogen dari data yang terdistribusi di berbagai sumber. Data tersebut disimpan dalam lokasi, sistem, format dan skema yang berbeda dan memberikan tantangan dalam penggunaan maupun integrasinya.

Secara umum, ada dua kategori data yaitu data terstruktur dan tidak terstruktur. Data terstruktur tersedia sebesar 20% dari seluruh data, serta direpresentasikan dalam bentuk relasi yang mudah dipetakan dan disimpan dalam database relasional.

Sedangkan data tidak terstruktur tersedia sebanyak 80% dari seluruh data yang direpresentasikan dalam berbagai bentuk dokumen seperti laporan, artikel berita, e-mail, dan konten web.

Data terstruktur biasanya disimpan dengan skema yang terdefinisi sehingga mudah untuk dilakukan query, dianalisa, dan diintegrasikan dengan data terstruktur

lainnya. Berbeda dengan data tidak terstruktur, yang secara alami susah untuk dilakukan query, dianalisa, maupun diintegrasikan dengan sumber data lain. Namun di balik itu semua, informasi tersembunyi yang tersimpan dalam data tidak terstruktur bisa sangat berperan dalam pengambilan keputusan. Sehingga apabila dalam pengambilan keputusan menggunakan integrasi data terstruktur dan data tidak terstruktur bisa menambah nilai yang signifikan serta kebenaran yang sejati bagi perusahaan/organisasi. Inilah yang menjadi tantangan sekaligus manfaat dalam integrasi data terstruktur dan tidak terstruktur.

Pada tugas ini penulis akan membahas tentang apa itu unstructure data kemudian perbedaan unstructured data dan structure data serta contoh-contohnya.

II. ISI PEMBAHASAN

1. Pengertian Unstructure Data

Menurut Pane (2020), Unstructured data seperti mis, teks, gambar, video, dokumen, dll. Bias juga dalam bentuk keluhan pelanggan, kontrak, atau e-mail internal. Jenis data ini menyumbang sekitar 90% dari data yang dibuat pada abad ini. Faktanya, pertumbuhan vulkanik media social (mis. Facebook dan twitter), sejak pertengahan decade terakhir, bertanggung jawab atas bagian utama dari data tidak terstruktur yang kita miliki saat ini. Data yang tidak terstruktur tidak dapat disimpan menggunakan databaserelasional tradisional. Menyimpan data dengan variasi dan kompleksitas seperti itu membutuhkan penggunaan system yang memadai, yang biasa disebut basis sebagai data NoSQL, seperti mis. MongoDB dan CouchDB. Pentingnya data yang tidak terstruktur terletak pada hubungan timbal balik yang tertanam yang mungkin tidak ditemukan jika jenis data lain dipertimbangkan. Apa yang membuat data yang dihasilkan dimedia

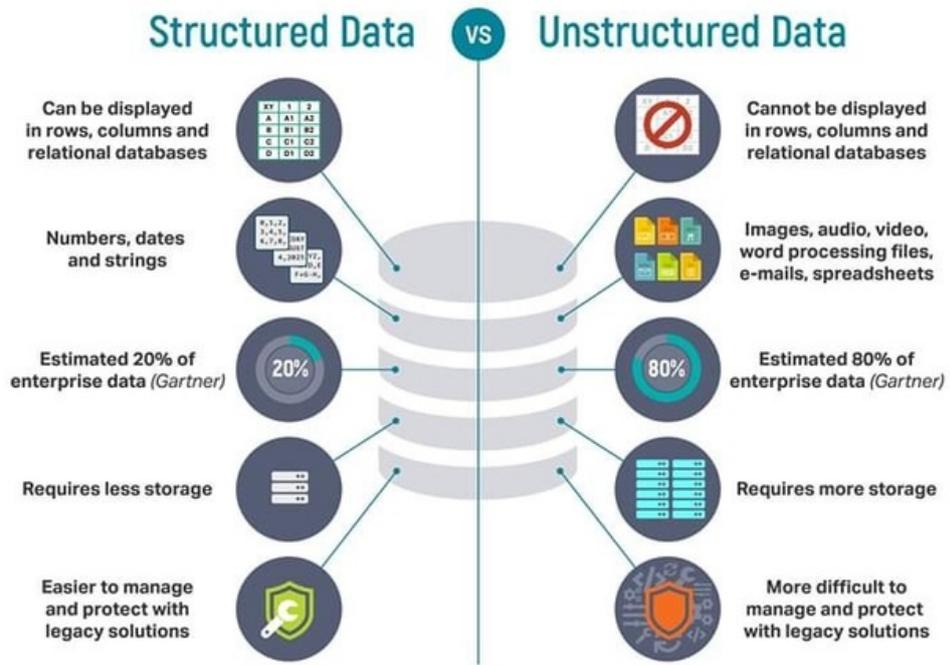
sosial berbeda dari tipe data lainnya adalah bahwa data dimedia social memiliki selera pribadi.

Setiap data dengan bentuk atau struktur yang tidak dikenal diklasifikasikan sebagai data yang tidak terstruktur. Selain ukurannya yang besar, data yang tidak terstruktur menimbulkan banyak tantangan dalam hal pemrosesan untuk mendapatkan nilai darinya. Contoh khas data tidak terstruktur adalah sumber data heterogen yang berisi kombinasi file teks sederhana, gambar, video, dll. Sekarang organisasi saat ini memiliki banyak data yang tersedia dengan mereka tetapi sayangnya, mereka tidak tahu bagaimana mendapatkan nilai dari itu karena data ini dalam bentuk mentah atau format tidak terstruktur.

2. Perbedaan Antara Unstructure data dengan Structure data

Menurut Nengsih (2019) , Data terstruktur adalah data yang berada dalam satu tempat baik berbentuk file, termasuk data yang berada didalam batabase ataupun spreadsheet. Sedangkan data tidak terstruktur adalah data yang tidak mudah dikasifikasi. Data tidak terstruktur seperti teks, file, video, email, laporan (report), persentasi powerpoint, pesan suara (voice mail), memo, dan citra.

Menurut Rianto (2019), yang paling terlihat jelas adalah bahwa data terstruktur itu bisa ditampilkan dalam bentuk table sementara data tidak terstruktur tidak bias. Untuk lebih detail perbedaan dapat dilihat pada gambar dibawah ini:



3. Contoh Data Structure dan Unstructure

Dikutip dari halaman Wikipedia bahwa Teknik seperti penambangan data, pemrosesan bahasa alami (NLP), dan analisis teks memberikan metode berbeda untuk menemukan pola dalam, atau menafsirkan informasi ini. Teknik umum untuk penataan teks biasanya melibatkan penandaan manual dengan metadata atau penandaan sebagian kata untuk penataan lebih lanjut berbasis penambangan teks. Standar Arsitektur Manajemen Informasi Tidak Terstruktur (UIMA) menyediakan kerangka kerja umum untuk memproses informasi ini untuk mengekstrak makna dan membuat data terstruktur tentang informasi tersebut.

Perangkat lunak yang menciptakan struktur yang dapat diproses dengan mesin dapat memanfaatkan struktur linguistik, pendengaran, dan visual yang ada dalam semua bentuk komunikasi manusia. Algoritma dapat menyimpulkan

struktur inheren ini dari teks, misalnya, dengan memeriksa morfologi kata, sintaksis kalimat, dan pola skala kecil dan besar lainnya. Informasi yang tidak terstruktur kemudian dapat diperkaya dan ditandai untuk mengatasi ambiguitas dan teknik berbasis relevansi yang kemudian digunakan untuk memfasilitasi pencarian dan penemuan. Contoh "data tidak terstruktur" dapat mencakup buku, jurnal, dokumen, metadata, catatan kesehatan, audio, video, data analog, gambar, file, dan teks tidak terstruktur seperti isi pesan email, halaman Web, atau kata-dokumen pengolah. Sementara konten utama yang disampaikan tidak memiliki struktur yang ditentukan, umumnya dikemas dalam objek (misalnya dalam file atau dokumen, ...) yang memiliki struktur dan dengan demikian merupakan campuran dari data terstruktur dan tidak terstruktur, tetapi secara kolektif ini masih disebut sebagai "data tidak terstruktur". Misalnya, halaman web HTML diberi tag, tetapi mark-up HTML biasanya hanya berfungsi untuk rendering. Itu tidak menangkap arti atau fungsi elemen yang ditandai dengan cara yang mendukung pemrosesan otomatis konten informasi halaman. Penandaan XHTML memungkinkan pemrosesan elemen mesin, meskipun biasanya tidak menangkap atau menyampaikan makna semantik dari istilah yang ditandai.

Karena data tidak terstruktur biasanya terjadi dalam dokumen elektronik, penggunaan konten atau sistem manajemen dokumen yang dapat mengkategorikan seluruh dokumen sering lebih disukai daripada transfer data dan manipulasi dari dalam dokumen. Manajemen dokumen dengan demikian menyediakan sarana untuk menyampaikan struktur ke koleksi dokumen.

III. Kesimpulan

Data terstruktur dan data tidak terstruktur apabila di integrasikan dalam menganalisa suatu permasalahan akan memberikan solusi yang lebih lengkap dan tepat sasaran. Akan tetapi untuk mengintegrasikan kedua jenis data tersebut bukanlah menjadi hal yang mudah karena membutuhkan teknik bahkan arsitektur yang tepat untuk mengatasi permasalahannya. Adapun teknik yang bisa digunakan yaitu text tagging dan annotation dan merupakan teknik yang populer dalam natural processing techniques.

DAFTAR PUSTAKA

Nengsih Warnia. 2019 : Implementasi Data Mining

Pane, Syafrial Fachri. 2020 : Big data classification behavior menggunakan python

https://en.m.wikipedia.org/wiki/Unstructured_data

<https://id.quora.com/Apa-perbedaan-antara-data-terstruktur-dan-tidak-terstruktur>

ESSAY
MATA KULIAH
ADVANCED DATABASE
STRUKTUR DATA DAN UNSTRUKTUR DATA



DI SUSUN

OLEH :

ELPINA SARI

192420050

MAGISTER TEKNIK INFORMATIKA
UNIVERSITAS BINA DARMA
TAHUN 2019/2020

1. PENDAHULUAN

Besarnya data yang tersimpan di dalam sebuah perusahaan berkembang sangat cepat tiap harinya. Kemampuan untuk mengakses dan menganalisa data tersebut dalam pembuatan keputusan yang cepat dan cerdas menjadi kunci kesuksesan sebuah perusahaan. Banyak perusahaan yang terus berkembang seiring dengan berputarnya waktu, sehingga menghasilkan informasi yang heterogen dari data yang terdistribusi di berbagai sumber. Data tersebut disimpan dalam lokasi, sistem, format dan skema yang berbeda dan memberikan tantangan dalam penggunaan maupun integrasinya. Dalam pengambilan keputusan yang efektif dan taktis, diperlukan kumpulan metodologi, proses, arsitektur dan teknologi yang mengubah data mentah menjadi informasi yang bermakna yang disebut Inteligensi Bisnis.

Sering kali kita mendengar kata data terstruktur (*structured*) dan juga Data Tidak Terstruktur (*Unstructured*). Namun, bagi sebagian orang ada yang belum mengerti apa arti sebenarnya dari kedua data tersebut. Data terstruktur adalah data yang berada dalam satu tempat baik berbentuk sebuah file termasuk data yang berada dalam database ataupun spreadsheet. Data terstruktur adalah yang membuat model data. Contohnya adalah data CRM, Industry Research Data dan lain sebagainya. Sedangkan Data tidak terstruktur adalah data yang tidak mudah diklasifikasi dan dimasukkan kedalam sebuah kotak dengan rapi. Contohnya adalah foto, gambar grafis, streaming instrument data, webpages, pdf, Power Point presentations, konten blog dan lain sebagainya.

2. PEMBAHASAN

2.1 Struktur data

Data terstruktur mendefinisikan tipe data atribut, yang ada pada record maupun tuple. Setiap record memiliki field yang sama. Data ada dalam entitas dan entitas yang sama dikelompokkan menjadi satu menggunakan relasi atau kelas. Entitas dalam kelompok yang sama memiliki atribut yang sama pula. Deskripsi untuk semua entitas dalam skema memiliki kesamaan format.

Penggunaan data terstruktur semakin banyak dilakukan, sejak basisdata relational ada untuk mengatur data dengan ukuran sangat besar. Sistem yang

ada saat ini, seperti CRM (Customer Relationship Management), ERP (Enterprise Resource Planning) dan CMS (Content Management System) merupakan data terstruktur untuk pemodelan data.

Data terstruktur biasanya dijalankan untuk mengakses database yang disebut dengan Structured Query Language atau yang lebih dikenal dengan SQL. Berikut adalah ilustrasi tentang SQL. Anda ingin melihat daftar buku yang memiliki harga diatas 1000 dan di urutkan sesuai dengan judul, Anda bisa menggunakan :***SELECT * FROM buku WHERE harga > 1000 ORDER BY judul;***

Maka akan mendapatkan data dari database sesuai permintaan Anda yaitu memunculkan buku yang memiliki harga diatas dan otomatis diurutkan sesuai judul.

a. Unstructured Data / Data Tidak Terstruktur

Merupakan informasi yang tidak memiliki model data yang telah ditentukan atau tidak terorganisir dengan cara yang telah ditentukan. Unstructured data biasanya berupa text document, dan mungkin juga berisi data seperti tanggal, angka, dan fakta lainnya. Unstructured data dapat berupa: text dokumen, email, video, audio, dan data dari sosial media. Data tidak terstruktur direpresentasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun diintegrasikan dengan sumber data lain, data tidak terstruktur tersedia sebanyak 80% dari seluruh data yang direpresentasikan. Namun di balik itu semua, informasi tersembunyi yang tersimpan dalam data tidak terstruktur bias sangat berperan dalam pengambilan keputusan. Sehingga apabila dalam pengambilan keputusan menggunakan integrasi data terstruktur dan data tidak terstruktur bias menambah nilai yang signifikan serta kebenaran yang sejati bagi perusahaan /organisasi.

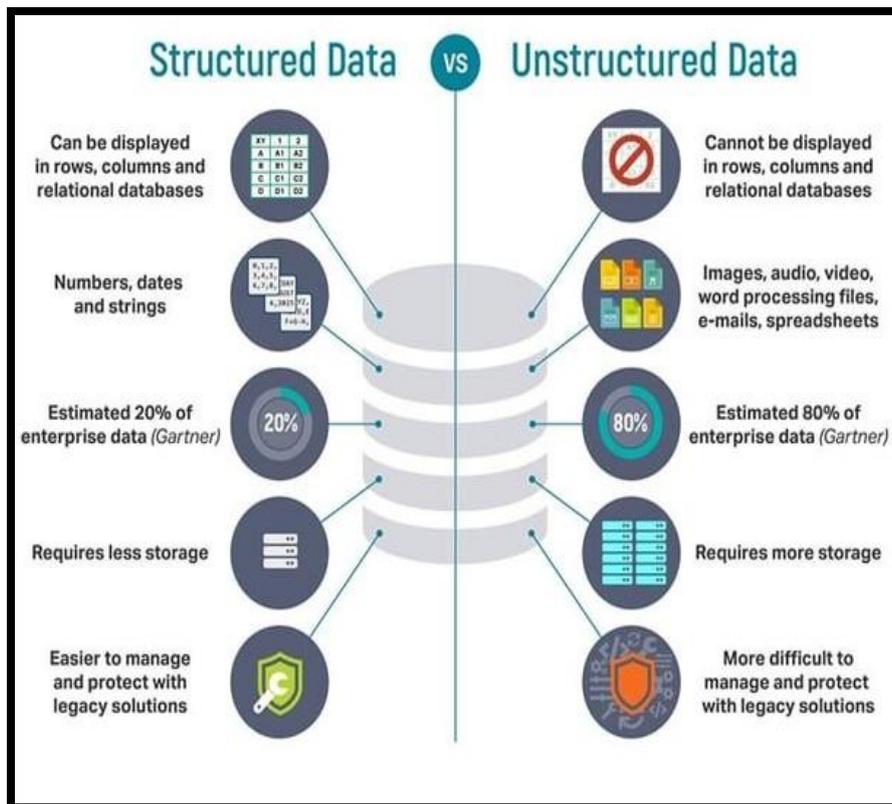
Data tidak terstruktur biasanya ada pada teks, file, video, email, laporan (report), presentasi power point, pesan suara (voice mail), memo, dan citra. Data dapat berbentuk dalam tipe apapun tanpa perlu mengikuti format, aturan atau alur tertentu. Sebagai contoh data yang ditampilkan pada halaman web termasuk data tidak terstruktur dan tidak beraturan, dan data pencarian dari

Google Search merupakan contoh data tidak terstruktur yang tidak mempunyai format khusus.

b. Perbedaan Struktur Data dan Unstruktur Data

Data tidak terstruktur adalah kelompok data yang tidak memiliki struktur atau aturan sesuai dengan format khusus, sehingga untuk menjadikan data terstruktur membutuhkan usaha, tools, dan waktu yang lebih maksimal .contoh data tidak terstruktur adalah data yang ditampilkan pada halaman web dan data yang ada pada Google Search.

Data terstruktur adalah Data yang dapat disimpan, diakses dan diproses dalam bentuk format tetap disebut sebagai data ‘terstruktur’. selama periode waktu tertentu. Contohnya, Data yang tersimpan dalam RDBMS (*Relational Database Management System*). Terlihat jelas bahwa data terstruktur itu bias ditampilkan dalam bentuk table sementara data tidak terstruktur tidak bisa. Untuk lebih detail perbedaan dapat di lihat pada gambar di bawah ini.



C. Contoh Unstruktur Data

Contoh Unstruktur data yaitu penyimpanan pada google drive. Dropbox adalah layanan penyimpanan cloud secara online yang berarti Anda dapat menyalin file Anda ke cloud dan mengaksesnya nanti atau kapan saja serta dimana saja, bahkan jika Anda mengaksesnya dari menggunakan perangkat yang berbeda sekalipun. Dropbox tidak akan secara otomatis menyalin semua file di komputer Anda jika Anda menggunakan paket pribadi, jadi Anda harus memilih dan memilih yang mana yang ingin Anda simpan.

Data dan informasi yang biasa disimpan di dalam Dropbox adalah berupa gambar, video, dokumen, aplikasi dan program. Setelah Anda selesai melakukan transfer data tersebut, Dropbox akan membuat dan menyimpan semua file-file itu di dalam cloud dan melakukan back up data, jadi Anda dapat membukanya kapan Anda membutuhkannya lagi. Aplikasi ini juga menyediakan layanan baik gratis maupun berbayar tentunya dengan keunggulan dan keuntungan yang bervariasi pula.

Pada tahun 2011 silam Dropbox meluncurkan Dropbox For Teams yang merupakan sebuah layanan dari Dropbox yang ditargetkan dan dikhususkan untuk mereka yang berada di kelompok bisnis atau dalam kelompok besar yang membutuhkan layanan untuk mengendalikan administrasi, tagihan yang terpusat dan keperluan lain yang cukup detail serta kompleks. Dropbox For Teams ini tersedia di laman web Dropbox dengan harga yang berbeda dari paket Dropbox lainnya

2.2 Dropbox

Bila dibandingkan dengan layanan serupa lainnya dan para pesaingnya, Dropbox berani menawarkan fitur dan akses yang relatif besar serta banyak dengan penggunaan sistem operasi yang tidak ditentukan dan bervariasi baik bagi pengguna mobile maupun pengguna desktop atau laptop. Dropbox juga menyediakan berbagai versi untuk beragam sistem operasi yang ada termasuk pengguna Microsoft Windows, Mac OS X, dan juga Linux. Menariknya, Dropbox akan memberi peningkatan kapasitas penyimpanan hingga 8 GB jika Anda berhasil menggaet atau menyarankan aplikasi ini kepada pengguna lain

yang belum pernah menggunakannya atau pengguna baru. Beberapa kompetitor dan aplikasi yang juga mempunyai layanan seperti Dropbox adalah FilesAnywhere, CloudMe, Egynye, iCloud, Mozy, Ubuntu One dan banyak lagi.

2.3 Manfaat dan fungsi Dropbox

Adanya penyimpanan cloud online dan layanan seperti Dropbox tentu saja banyak membantu cara simpan konvensional dulu, inilah manfaat dan fungsi

1. Menyimpan sekaligus mengamankan file

Banyak hal bisa Anda eksplorasi di Dropbox, tetapi layanan ini lebih sering dimanfaatkan untuk menyimpan, berbagi dan juga mengamankan data-data penggunanya. Meskipun penyimpanan ini dalam bentuk cloud yang dioperasikan secara online, data dan informasi serta file Anda akan tetap terjamin karena untuk membukanya Anda memerlukan username dan password pribadi sehingga Anda dapat menghindari kerusakan jika Anda menyimpannya di perangkat keras.

2. Mampu membawa dan membuka data dimana saja

Penyimpanan model lama mengharuskan Anda menenteng harddisk maupun drive kecil yang saat ini tidak lagi relevan. Tidak hanya beresiko rusak, data yang corrupt, serta hilang, gaya penyimpanan lama ini juga mempunyai batas storage. Oleh karena itulah Dropbox hadir untuk meminimalisir kemungkinan buruk yang ditimbulkan oleh perangkat keras.

Belum lagi kemudahan untuk mengakses data dari koneksi internet dan smartphone yang Anda punya akan terasa lebih ringkas, praktis dan efisien mengingat Anda langsung dapat menyimpan dan mengakses file, mengunduh file baru, berbagi file dan menata file yang sudah ada.

3. Berbagi video dan file

File dalam format video biasanya mempunyai ukuran dan memerlukan storage yang besar, melalui Dropbox Anda dapat menggunakan layanan yang satu ini karena aplikasi tersebut memberi kemudahan dalam melakukan pengiriman dan pengunduhan berbagi video yang cukup berbeda dari layanan lainnya. Pengguna juga dapat memilih penerima yang hendak mereka bagikan, sehingga pembagian akses akan terjamin kerahasiaannya.

4. Kemampuan melakukan back up data

Fitur atau opsi ini adalah faktor yang paling menarik dari Dropbox dibandingkan dengan penyimpanan konvensional. Anda dapat memback up foto- foto, video, dokumen dan apa saja yang berharga dan penting bagi Anda. ditambah pula Anda juga dapat mengorganisir file Anda berdasarkan keinginan Anda.

5. Berbagi dan bekerja sama

Tidak hanya mampu menyimpan file dan data yang bersifat pribadi, Dropbox juga memiliki fitur serta opsi yang dapat memungkinkan dan mengakomodasi para penggunanya untuk mengerjakan suatu proyek atau dokumen bersama. Fitur yang canggih ini menghubungkan banyak pengguna untuk mengakses satu file yang sama disaat bersamaan tanpa takut terjadinya error atau penabrakan data.

6. Jaminan keamanan yang lebih baik

Karena menggunakan username dan password, maka Anda tidak perlu merasa khawatir akan pembobolan dan kehilangan data. Dengan program khusus yang dirancang untuk mengamankan jutaan file penggunanya dan jaminan keamanan akan terjadinya penyebaran malware diantara semua file yang ada.

3. KESIMPULAN

Dari pembahasan di atas dapat di simpulkan bahwa data terstruktur adalah Data terstruktur biasanya dijalankan untuk mengakses database yang disebut dengan Structured Query Language atau yang lebih dikenal dengan SQL. Berikut adalah ilustrasi tentang SQL. Anda ingin melihat daftar buku yang memiliki harga diatas 1000 dan di urutkan sesuai dengan judul. Sedangkan Data tidak terstruktur adalah data yang tidak mudah diklasifikasi dan dimasukkan kedalam sebuah kotak dengan rapi. Contohnya adalah foto, gambar grafis, streaming instrument data, webpages, pdf, Power Point presentations, konten blog dan lain sebagainya. Data tidak terstruktur dapat di contohkan pada penyimpanan Dropbox.

4. DAFTAR PUSTAKA

Afifanto, C. Integrasi Data Terstruktur dan Tidak Terstruktur dalam Sistem Integrasi Bisnis. Jurnal Komputasi. 2015

Maryanto, budi. 2017. JURNAL BIG DATA DAN PEMANFAATANNYA DALAM BERBAGAI SEKTOR. Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI: Bandung. Media Informatika Vol.16 No.2

<https://wahyudisetiawan.wordpress.com/2009/08/03/data-tidak-terstruktur-semiterstruktur-dan-terstruktur/>

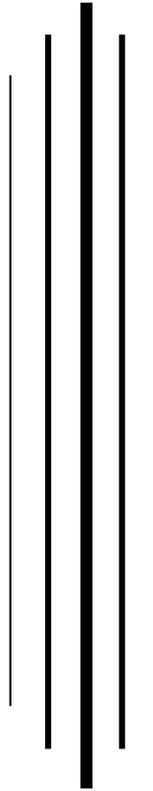
<https://idcloudhost.com/mengenal-apa-itu-dropbox-keunggulan-manfaat-fungsi-dan-cara-kerjanya/>

<https://blog.wowrack.co.id/2015/04/apa-itu-structured-data-dan.html>

<http://edomuhardosyafri.blogspot.com/2014/12/apakah-big-data-what-is-big-data.html>

<https://id.quora.com/Apa-perbedaan-antara-data-terstruktur-dan-tidak-terstruktur>

UAS ADVANCED DATABASE
KELAS MTI 22A



DOSEN PENGASUH

MUHAMMAD AKBAR, ST., M.IT

DISUSUN OLEH:

FADEL MUHAMMAD MADJID

192420052

PROGRAM PASCA SARJANA MAGISTER TEKNIK INFORMATIKA

UNIVERSITAS BINA DARMA

2020

Penggunaan Text Mining Untuk Klasifikasi Berita Bencana Hidrometeorologi Berdasarkan Data Web

Dibuat oleh: Fadel Muhammad Madjid

Perbedaan data terstruktur dan data tidak terstruktur

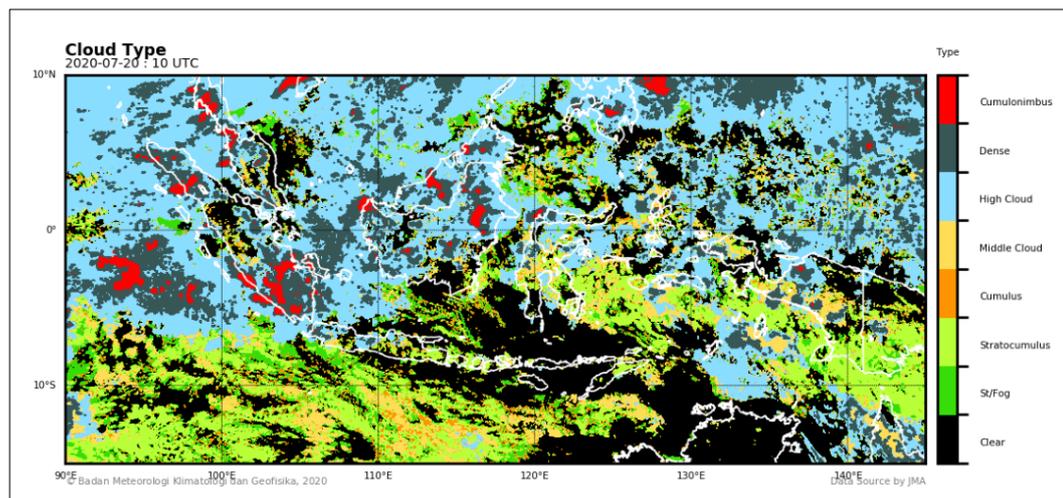
Data diklasifikasikan dalam 2 jenis, yaitu terstruktur dan tidak terstruktur. Data terstruktur direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya. Sedangkan data tidak terstruktur direpresentasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisa maupun diintegrasikan dengan sumber data lain.

Data terstruktur mendefinisikan tipe data atribut, yang ada pada record maupun tuple. Setiap record memiliki field yang sama. Data ada dalam entitas dan entitas yang sama dikelompokkan menjadi satu menggunakan relasi atau kelas. Entitas dalam kelompok yang sama memiliki atribut yang sama pula. Deskripsi untuk semua entitas dalam skema memiliki kesamaan format [1].

Penggunaan data terstruktur semakin banyak dilakukan, sejak basis data relasional ada untuk mengatur data dengan ukuran sangat besar. Sistem yang ada saat ini, seperti CRM (Customer Relationship Management), ERP (Enterprise Resource Planning) dan CMS (Content Management System) merupakan data terstruktur untuk pemodelan data. Sebagai contoh data terstruktur yang ada di institusi BMKG adalah data cuaca berupa numerik seperti suhu, kelembaban, arah dan kecepatan angin dll.

dramaga	RR	WS_AVG	WS_MAX	WD_AVG	TT_AIR_MAX	TT_AIR_ATT	AIR_M	RH_AVG	PP_AIR
2017-10-01 00:00:00	0	0.8265		60	183.07	34.466	25.972	0	78.536
2017-10-01 00:10:00	0	0.8265		60	183.07	34.466	25.972	0	78.537
2017-10-01 00:20:00	0	0.82645		60	183.07	34.466	25.972	0	78.538
2017-10-01 00:30:00	0	0.8264		60	183.06	34.466	25.972	0	78.538
2017-10-01 00:40:00	0	0.82635		60	183.07	34.466	25.972	0	78.538
2017-10-01 00:50:00	0	0.8263		60	183.07	34.466	25.972	0	78.538
2017-10-01 01:00:00	0	0.82625		60	183.08	34.466	25.972	0	78.538
2017-10-01 01:10:00	0	0.82619		60	183.08	34.466	25.972	0	78.539
2017-10-01 01:20:00	0	0.82615		60	183.08	34.466	25.972	0	78.539
2017-10-01 01:30:00	0	0.82611		60	183.08	34.466	25.972	0	78.539
2017-10-01 01:40:00	0	0.82606		60	183.07	34.466	25.972	0	78.538
2017-10-01 01:50:00	0	0.82601		60	183.07	34.466	25.972	0	78.538
2017-10-01 02:00:00	0	0.82599		60	183.06	34.466	25.972	0	78.538
2017-10-01 02:10:00	0	0.82597		60	183.06	34.466	25.972	0	78.537
2017-10-01 02:20:00	0	0.82604		60	183.05	34.466	25.972	0	78.537
2017-10-01 02:30:00	0	0.82615		60	183.05	34.466	25.973	0	78.536

Data tidak terstruktur biasanya ada pada teks, file, video, email, laporan (report), presentasi power point, pesan suara (voice mail), memo, dan citra [1]. Data dapat berbentuk dalam tipe apapun tanpa perlu mengikuti format, aturan atau alur tertentu. Sebagai contoh data yang ditampilkan pada halaman web termasuk data tidak terstruktur dan tidak beraturan. Sebagai contoh data tidak terstruktur yang ada di institusi BMKG adalah data citra satelit dan radar serta visualisasi prakiraan cuaca.



Karakteristik data tidak terstruktur adalah data tidak sesuai dengan model atau struktur apapun, data tidak data disimpan dalam bentuk baris dan kolom seperti pada database, tidak mengikuti semantik atau aturan apapun, sata tidak memiliki format dan urutan tertentu, identifikasi yang cukup sulit karena tidak memiliki struktur data.

Bencana hidrometeorologi

Dalam kehidupan sehari-hari tentunya sangat berkaitan erat dengan kondisi fisik lingkungan alam. Alam selalu menciptakan suatu kesetimbangan untuk bisa mempertahankan keseimbangan alam. Salah satu contoh bentuk keseimbangan alam adalah adanya bencana baik berupa geologis maupun hidrometeorologi.

Bencana hidrometeorologi merupakan sebuah fenomena alam yang terjadi berkaitan dengan lapisan atmosfer, hidrologi dan oceanografi yang berpotensi membahayakan, merusak, dan menyebabkan hilangnya nyawa penduduk [2]. Bencana hidrometeorologi ini adalah bencana yang termasuk banjir, tanah longsor, angin puting beliung, badai es atau di Indonesia sering terjadi hujan es, badai salju, bencana kekeringan, hujan yang sangat lebat, hujan salju yang sangat lebat dll.

Permasalahannya adalah basis data mengenai bencana hidrometeorologi masih minim. Di situs web BNPB sendiri masih belum memiliki basis data mengenai kejadian hujan es di seluruh Indonesia [3]. Sementara data mengenai kejadian hujan es masih dalam bentuk teks di koran-koran daring. Akibatnya dibutuhkan suatu sistem untuk bisa mencari dan mengklasifikasi ribuan bahkan jutaan konten-konten dalam daring untuk bisa menjadi data yang bisa diambil sesuai kebutuhan user itu sendiri.

Text mining

Salah satu cara untuk mengumpulkan informasi adalah dengan dilakukannya data mining. Data mining adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar [4]. Proses data mining seringkali menggunakan metode statistika, matematika, hingga memanfaatkan teknologi artificial intelligence. Dari beberapa metode data mining tersebut, salah satunya adalah text mining.

Text mining adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dimana, text mining merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar [5]. Selain klasifikasi, text mining juga digunakan untuk menangani masalah clustering, information extraction, dan information retrieval [6].

Perancangan program classifier dapat dibagi menjadi masing-masing kejadian bencana hidrometeorologi dengan menggunakan bahasa PHP dan database server MySQL. Sebagai contoh dalam melakukan pengklasifikasian data mining kejadian bencana hidrometeorologi seperti hujan es, puting beliung, banjir, tanah longsor serta kekeringan adalah dengan menggunakan algoritma naïve bayes classifier [7].

Kategori berita mengenai kejadian bencana dapat diambil dari media daring yang menggunakan Bahasa Indonesia. Oleh karena itu, pada tahap text mining tidak dilakukan tahap tagging karena tidak menangani teks yang berbahasa inggris. Berikut adalah langkah-langkah dalam melakukan text mining:

Langkah pertama adalah text processing yang merupakan tindakan untuk mengubah semua karakter huruf menjadi huruf kecil yaitu toLowerCase [8]. Setelah itu dilakukannya proses Tokenizing yaitu penguraian deskripsi yang semula berupa kalimat menjadi kata-kata dan menghilangkan delimiter seperti tanda (. - , #).

Langkah kedua adalah feature selection dimana tahap ini menghilangkan stopwords dan stemming terhadap kata berimbuhan seperti (di, ke an). Tujuan dari proses ini adalah menghilangkan imbuhan seperti sufiks, prefix, konfiks pada setiap kata. Oleh karena itu setiap kata diubah menjadi kata dasar sesuai dengan buku KBBI.

Algoritma Naïve Bayes Classifier

Naïve Bayes Classifier (NBC) merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Dalam algoritma NBC dokumen direpresentasikan dengan pasangan atribut “ $x_1, x_2, x_3 \dots x_n$ ” dimana x_1 adalah kata pertama dst. Sedangkan V adalah himpunan kategori berita. Pada saat klasifikasi algoritma akan mencari probabilitas yang tertinggi dari semua kategori dokumen uji. Persamaannya adalah sebagai berikut:

$$V_{map} = \underset{V_j \in V}{\operatorname{arg\,max}} \left(\frac{P(x_1, x_2 \dots x_n | V_j) P(V_j)}{P(x_1 x_2 \dots x_n)} \right)$$

Untuk $P(x_1, x_2 \dots x_n)$ nilainya konstan untuk semua kategori (V_j) sehingga persamaannya ditulis sebagai berikut:

$$V_{map} = \underset{V_j \in V}{\operatorname{argmax}} (P(x_1, x_2 \dots x_n | V_j) P(V_j))$$

Persamaan diatas dapat disederhanakan sebagai berikut:

$$V_{map} = \underset{V_j \in V}{\operatorname{argmax}} \prod_{i=1}^n (P(x_i | V_j) P(V_j))$$

Keterangan:

V_j adalah kategori berita $j = 1, 2, 3, 4$ dst.

$P(x_1 | V_j)$ adalah probabilitas x_1 pada kategori V_j

$P(V_j)$ adalah probabilitas dari V_j

Untuk $P(V_j)$ dan $P(x_j | V_j)$ dihitung pada saat pelatihan dimana persamaannya adalah sbb:

$$P(V_j) = \frac{|docs\ j|}{|contoh|}$$

$$P(x_1 | V_j) = \frac{n_k + 1}{n + |kosakata|}$$

$|docs\ j|$ adalah jumlah dokumen pada kategori j

$|contoh|$ jumlah dokumen seluruh kategori

n_k adalah frekuensi kemunculan kata

n jumlah frekuensi kemunculan kata dari setiap kategori

Harapannya adalah dengan penggunaan text mining algoritma NBC tersebut dapat mengklasifikasikan kejadian bencana hidrometeorologi sesuai berita daring. Perkembangan basis data juga dapat tercapai jika seluruh laporan mengenai kejadian bencana tersebut dapat terakses oleh algoritma. Kedepannya hal ini ini dapat membantu perkembangan prakiraan cuaca mengenai kejadian ekstrim dan dapat menjadi self awareness bagi masyarakat untuk paham mengenai cuaca.

Daftar Pustaka

- [1] <https://wahyudisetiawan.wordpress.com/2009/08/03/data-tidak-terstruktur-semiterstruktur-dan-terstruktur/>
- [2] Tjasyono HK, Bayong dan Harijono, Sri Woro B., 2012, *Meteorologi Indonesia II*, Jakarta : BMKG.
- [3] <https://bnpb.cloud/>
- [4] <https://www.jagoanhosting.com/blog/apa-itu-data-mining/>
- [5] Feldman, R & Sanger, J. 2007. *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press : New York.
- [6] Berry, M.W. & Kogan, J. 2010. *Text Mining Application and theory*. WILEY : United Kingdom.
- [7] Efendi, R., & Malik, R. F. (2012). *Klasifikasi dokumen berbahasa indonesia menggunakan naive bayes classifier*
- [8] Weiss, S.M., Indurkha, N., Zhang, T., Damerau, F.J. 2005. *Text Mining : Predictive Methods fo Analyzing Unstructured Information*. Springer : New York

ESSAY

MENGELOLA DATA TIDAK TERSTRUKTUR (UNSTRUCTURED DATA)



DOSEN PENGASUH :

M. Akbar, ST, M.IT

OLEH :

Isti M. Nasichah, ST

192420051

PROGRAM PASCA SARJANA MAGISTER TEKNIK INFORMATIKA

UNIVERSITAS BINA DARMA

MENGELOLA DATA TIDAK TERSTRUKTUR

PENDAHULUAN

Saat ini dunia teknologi berkembang dengan sangat cepat. Hal tersebut menuntut manusia agar dapat mengikuti perkembangan tersebut. Maka dari itu dibutuhkan sumber daya manusia dalam bidang IT dengan kemampuan beradaptasi secara cepat. Dalam dunia teknologi informasi saat ini, kebutuhan terhadap data adalah sebuah kunci yang wajib dimiliki pada setiap lini bisnis. Walaupun bisnisnya bahkan tidak berhubungan dengan dunia teknologi, namun data sangat diperlukan dalam melakukan analisis yang nantinya akan digunakan untuk pengambilan keputusan.

Sebelum membahas lebih lanjut, sebenarnya data itu apa? Dari sejumlah referensi, data dapat didefinisikan sebagai fakta atau hasil dari suatu pengamatan (observasi) langsung terhadap suatu kejadian, yang mewakili objek atau konsep dalam kehidupan sehari-hari. Data dapat berupa tulisan atau gambar yang dilengkapi dengan nilai tertentu. Sebagai contoh, daftar hadir mahasiswa semester 1 Magister Teknik Informatika merupakan sebuah data.

Secara umum, data dapat diklasifikasikan menjadi dua jenis yaitu data terstruktur dan tidak terstruktur. Data terstruktur tersedia sebesar 20% dari seluruh data, serta direpresentasikan dalam bentuk relasi yang mudah dipetakan dan disimpan dalam database relasional. Sedangkan data tidak terstruktur tersedia sebanyak 80% dari seluruh data yang direpresentasikan dalam berbagai bentuk dokumen seperti laporan, artikel berita, *email*, dan konten web.

Data terstruktur (*structured data*) merupakan data yang direpresentasikan ke dalam bentuk tabel atau relasi yang biasanya ditempatkan dalam suatu basis data. Sedangkan data tidak terstruktur (*unstructured data*) merupakan data yang tidak dapat dengan mudah diinterpretasikan sebelum diolah terlebih dahulu. Contoh data tidak terstruktur adalah dokumen teks, multimedia, dan sebagainya. Data tidak terstruktur jika diolah sedemikian rupa bisa menjadi data terstruktur sesuai dengan kebutuhan.

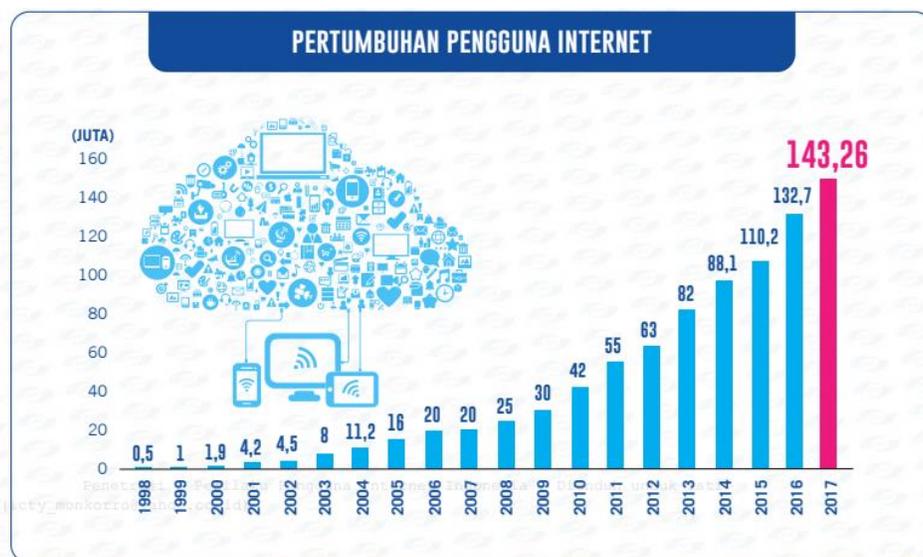
Kemajuan teknologi internet, media digital seperti audio, video, dan teks dapat dikirimkan melalui internet dengan lebih mudah. Saat memiliki data, pertanyaan selanjutnya adalah apa yang dapat dimanfaatkan dengan data ini? Jika hanya menyimpan data tersebut secara *historical*, maka hanya didapatkan storage yang semakin lama akan semakin berkurang. Skala volume data

yang berjumlah besar yang meningkat pesat setiap harinya akan menjadi “sampah” pada penyimpanan apabila tidak diolah menjadi informasi yang berguna. Bagaimana teknik untuk mengatasi “sampah” data yang tidak berguna tersebut agar diolah menjadi informasi yang berguna?

PEMBAHASAN

Perkembangan Data

Seiring dengan perkembangan internet, maka data-data yang tersimpan baik dalam bentuk teks, gambar, suara dan video juga mengalami peningkatan yang sangat cepat dan signifikan. Pada gambar 1, menunjukkan bahwa di Indonesia, pengguna internet pada tahun 1998 hanya 500.000 pengguna. Sedangkan sampai dengan tahun 2017 diproyeksikan pengguna internet sudah mencapai angka 143,26 juta (<https://apjii.or.id>, 2017).



Gambar 1. Data Pengguna Internet di Indonesia

Era industri 4.0 dengan pilar utama, yaitu *Internet of Things (IoT)*, *cloud computing*, *artificial intelligence* dan *big data* telah memproduksi banyak sekali data. Menurut data yang dihimpun oleh IBM (Ohlhorst, 2013) dalam (Suntoro, 2019), di dunia ini skala volume data yang kita buat setiap hari sebesar 2,5 exabyte ($2,5 \times 10^{18}$ byte), dimana 90% data yang berada di dunia ini telah kita buat sejak dua tahun yang lalu. Diprediksi sampai tahun 2020, skala volume data akan meningkat drastis menjadi 40 zettabyte (40×10^{21} byte).

Data Terstruktur dan Tidak Terstruktur

Jenis data banyak pengelompokannya bergantung dari konteks mana seseorang melihatnya. Berdasarkan pengelolaannya, data dapat dibedakan menjadi 2 jenis, yaitu data terstruktur dan tidak terstruktur. Pada dasarnya semua data berasal dari data tidak terstruktur, namun setelah dilakukan perlakuan tertentu terhadap data, maka data tidak terstruktur dapat berubah menjadi data terstruktur.

Data terstruktur adalah data yang ditempatkan pada item-item tertentu sebagai lokasi penyimpanan dalam suatu basis data. Dengan basis data tersebut proses temu kembali informasi akan lebih cepat dan mudah dibandingkan dengan data yang tidak terstruktur. Data terstruktur diorganisasikan sedemikian sehingga dapat disimpan dalam tabel dan relasi. Data terstruktur biasanya dijalankan untuk mengakses database yang disebut dengan *Structured Query Language* atau yang lebih dikenal dengan SQL. Data terstruktur dapat digunakan untuk menampung lebih dari satu tipe data dalam variabel sehingga mudah saling berhubungan dan dapat disusun dalam satu struktur. Dalam tipe data struktur setiap perubah bisa menyimpan lebih dari sebuah nilai data. Jenis jenis tipe data terstruktur yaitu larik (*array*), rekaman (*record*), himpunan (*set*).

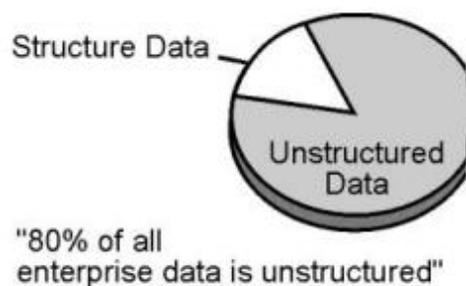
Pada umumnya data terstruktur disimpan dalam berbagai aplikasi basis data, sementara data tidak terstruktur tersimpan dalam berbagai bentuk file seperti dokumen, video, suara, email, attachment, dan lain-lain. Menurut hasil penelitian, sampai saat ini di USA, jumlah data yang dihasilkan sekitar 2 exabyte per tahun, dengan pertumbuhan rata-rata 50%. Demikian pula di Indonesia, perhatian terhadap data tidak terstruktur pada tahun 2007 diperkirakan akan mendapatkan perhatian lebih besar di kalangan perusahaan. Hal itu didorong oleh semakin meningkatnya volume data tidak terstruktur dan kebutuhan perusahaan untuk menekan biaya penyimpanan atau pengarsipan data.

Menurut pimpinan Regional Product Manager Hitachi Data Systems (HDS) Asia Pacific, Lim Beng Lay sampai saat ini pertumbuhan data tidak terstruktur melampaui data-data terstruktur. Beng Lay menyatakan, sejak tahun 2006 sampai 2011, data tidak terstruktur tumbuh sebesar 63,7 persen, sementara data-data terstruktur tumbuh 32,3 persen. Tahun 2010, volume data tidak terstruktur akan jauh melampaui volume data terstruktur.

Untuk keperluan pengarsipan data tidak terstruktur ini, perusahaan-perusahaan biasa mengatasinya dengan menambah storage. Secara tradisional, perusahaan-perusahaan memilih tape storage untuk memenuhi kebutuhan ini. Namun, kemudian timbul masalah baru, ternyata penambahan storage tidaklah sederhana. Dengan jumlah tape yang meningkat, perusahaan juga harus memikirkan pengelolaannya, bagaimana melakukan backup data dan sebagainya.

Sebagaimana di perusahaan, di perpustakaan juga banyak informasi yang berupa data tidak terstruktur. Sebagai contoh informasi hasil-hasil penelitian atau karya tulis. Data seperti itu sulit untuk dikemas sedemikian rupa dan diubah menjadi data terstruktur. Oleh karena itu, maka pendekatan yang dapat dilakukan adalah dengan menganalisis terhadap kandungan subjek dari artikel dan membuatkan metadata dari artikel tersebut. Data seperti itu dikenal dengan nama katalog. Pada masa lalu katalog dibuat pada sehelai kertas berupa kartu yang diketik secara manual. Sekarang metadata tetap digunakan, namun bentuknya sudah berupa data digital dan dilengkapi dengan artikel lengkapnya. Sehingga pengguna yang membutuhkan informasi akan mencari melalui metadata dan untuk lebih meyakinkannya dilihat dari artikel lengkapnya.

Berdasarkan data IDC, jumlah data tidak terstruktur 80% lebih banyak dihasilkan daripada data terstruktur dalam suatu perusahaan atau organisasi (Gambar 2). Dalam data tidak terstruktur, konteks informasi tidak diketahui dengan jelas. Isi informasi dapat diketahui tetapi kadangkala tidak mempunyai arti atau di luar konteks. Pada dasarnya semua data "dilahirkan" tidak terstruktur. Kemudian dengan adanya upaya pengelolaan, data tidak terstruktur dapat diubah menjadi data terstruktur. Namun hasil perubahan tersebut biasanya akan menimbulkan kehilangan beberapa informasi penting yang terkandung di dalamnya.



Gambar 2. Komposisi data terstruktur dan tidak dalam suatu organisasi (Sumber : IDC)

Sebagai contoh pada dokumen transaksi penjualan. Di dalam dokumen tersebut terdapat banyak informasi detail yang disampaikan secara tekstual, seperti langkah-langkah pembeli dan

penjual dalam mencapai harga kesepakatan dan sebagainya. Pembuatan metadata merupakan langkah untuk mengubah data tidak terstruktur menjadi data terstruktur. Namun, apabila dilakukan pengolahan dokumen dengan mengeluarkan metadata dari dokumen tersebut maka tidak semua informasinya dapat diperoleh dengan mudah dan cepat. Informasi yang tidak semua diperoleh itulah yang dimaksudkan sebagai informasi yang hilang.

Beberapa teknik penangkapan data tidak terstruktur supaya tidak hilang informasinya adalah dengan membuatnya menjadi data terstruktur dalam format XML dan HTML. Suatu format yang melibatkan tag-tag data dalam pengelolaan datanya. Data bisnis pelengkap ini sangat penting karena menjadi kunci keberhasilan dalam meningkatkan bisnis menuju keberhasilan pada tingkat yang lebih tinggi yaitu kepuasan pelanggan dan meningkatkan daya saing bisnis.

Kebanyakan informasi yang dihasilkan dan digunakan dari perusahaan dan organisasi berbentuk data tidak terstruktur. Data tidak terstruktur merupakan informasi generasi manusia atau mesin di mana data tidak dapat dengan mudah dibuat menjadi data standar yang terstruktur. Selain itu data juga tidak dapat dengan mudah diakses tanpa bantuan interpretasi manusia atau mesin. Contoh data tidak terstruktur antara lain dokumen, multimedia, isi peta, informasi geografis, gambar satelit, catatan medis dan sebagainya.

Teknik pengelolaan data tidak terstruktur bervariasi bergantung kepada bagaimana data tersebut dibuat dan digunakan. Beberapa contoh data tidak terstruktur dalam organisasi adalah sebagai berikut :

1. Data yang tersimpan di komputer dalam administrasi perkantoran (dokumen, *spreadsheet*, dan materi presentasi).
2. Dokumen arsip dan perpustakaan digital.
3. Data gambar di bank dan perpustakaan untuk keperluan keilmuan dan penelitian.
4. Data dalam suatu aplikasi data warehouse pada sektor publik, keamanan dan telekomunikasi.
5. Data dalam sistem operasi yang terintegrasi dalam dunia bisnis dan catatan kesehatan.
6. Data lokasi dan proyek yang berupa audio, video dan gambar tentang informasi retail, asuransi, perlindungan kesehatan, pemerintahan, dan sistem keamanan publik.
7. Data semantik yang digunakan di akademi, farmasi dan penelitian lain.

Resiko Memiliki Data Tidak Terstruktur

Gartner, perusahaan penelitian dan penasihat terkemuka dunia, memperkirakan bahwa saat ini lebih dari 80% data perusahaan tidak terstruktur dengan baik. Hal ini berarti informasi yang dimiliki oleh organisasi perusahaan tidak dapat dengan mudah ditemukan atau di simpan dengan aman. Ini adalah bagian yang sulit dikelola karena tidak semua data yang dibuat dan dikumpulkan dapat disusun dengan mudah. Misalnya, hampir 270 miliar *email* dibuat setiap hari. Banyak dari *email* tersebut memiliki informasi pribadi atau *sensitive*. Selain itu, data yang tidak terstruktur semakin banyak setiap tahun nya. Setiap PDF, video, dokumen *Word* dan bahkan unggahan pada media social pun ikut menambah masalah ini. Tipikal praktek kerja modern dengan lingkungan bertekanan tinggi juga dapat dipersalahkan dalam menyumbang banyaknya data yang tidak terstruktur tersebut. Pola pikir yang menciptakan budaya berdasarkan data menjadi hal yang problematik. Begitu juga dengan pendekatan terhadap keamanan data.

Karyawan yang membutuhkan data ketika pergi keluar kantor (mungkin ketika mereka bekerja dari rumah atau saat meeting di luar kantor), memiliki kecenderungan mengambil data dari sumber yang sudah terstruktur dan menyimpannya ke dalam format yang tidak terstruktur. Data tersebut kemudian tidak lagi terlindung atau aman dan tidak dapat di akses dan dikelola dengan mudah

Memiliki banyak data yang tidak terstruktur, dan tidak adanya sistem yang mampu melacak data tersebut, membuat organisasi rentan terhadap beberapa risiko. Pertama, data yang tidak terstruktur merupakan tantangan karena organisasi tidak mengetahui dimana data yang tidak terstruktur tersebut disimpan dan apa informasi yang terkandung didalamnya. Tidak mengetahui kapan data dibuat, atau kapan data disalin dari sumber data yang terstruktur ke dalam format yang tidak terstruktur, akan mempersulit organisasi untuk mengetahui seberapa besar masalah yang ada dan langkah-langkah baru apa yang perlu diperkenalkan untuk memecahkan masalah. Kedua, proteksi data pribadi merupakan topik hangat saat ini, khususnya dengan diperkenalkannya GDPR. Jika sebuah organisasi tidak dapat menemukan data dalam sumber yang tidak terstruktur, maka organisasi tersebut berisiko tidak dapat memberikan informasi yang dibutuhkan dalam jangka waktu tertentu. Selain itu, organisasi yang berbagi data dengan mitranya mungkin saja meneruskan data yang tidak terstruktur tersebut dan kemudian tidak lagi memiliki akses kepada datanya sendiri. Masalah hukum dengan pelanggan atau mitra dapat kemudian muncul karena persyaratan yang disetujui pada awal perjanjian dibuat melalui

email. Dengan dibuatnya data di sumber yang tidak terstruktur tersebut, akan mempersulit pencarian dan penarikan kembali informasi yang dibutuhkan sebagai bukti. Ketiga, dari segi keamanan, penjahat dunia maya merasa semakin sulit untuk mengakses jaringan dan data perusahaan karena solusi keamanan yang semakin efektif. Dengan alasan tersebut mereka mencari jalan yang lebih lemah dan data yang tidak terstruktur merupakan target utama. Keempat, dari segi finansial, organisasi yang mengumpulkan data tanpa mengelolanya dalam sumber data yang terstruktur memiliki resiko peningkatan biaya penyimpanan data digitalnya dari tahun ke tahun. Hal ini berarti perlu membayar lebih untuk ruang, alat dan teknologi untuk menyimpan data, alih-alih memecahkan inti masalah. Jika sebuah organisasi mengekstraksi data menjadi sumber yang terstruktur, maka organisasi itu hanya menyimpan data yang dibutuhkan dan dapat menggunakan kapasitas penyimpanannya dengan lebih efisien.

Mengelola Data Tidak Terstruktur

Dengan adanya sistem pengelolaan basis data, data tidak terstruktur sudah dapat dikelola dalam jumlah besar. Sebagai contoh basis data yang digunakan untuk membuat katalog, dokumen referensi, gambar, dan isi media yang disimpan dalam file.

Untuk menyimpan data tidak terstruktur ke dalam basis data, penggunaan tabel basis data Binary Large Objects atau BLOBs telah digunakan dalam beberapa dekade belakangan ini. BLOBs, untuk beberapa tahun Oracle telah mengintegrasikan jenis dan struktur data secara optimal dengan melakukan analisis dan manipulasi dokumen XML, multimedia content, teks dan informasi geospasial. Contohnya dengan aplikasi Oracle database 11g, data tidak terstruktur telah dapat dikelola dengan memperbaiki performa dan keamanannya.

Dalam mengaplikasikan suatu sistem pengelolaan data tidak terstruktur ada 10 hal yang harus diperhatikan, yaitu :

1. Visibilitas

Beberapa alternatif pemecahan dalam pengelolaan dan pengendalian data harus ada dengan representasi visual yang jelas terhadap akses data sebagaimana didefinisikan dalam jaringan.

2. Pengendalian (*control*)

Alternatif pemecahan masalah harus sudah termasuk semua mekanisme, uji, menetapkan dan mengembalikan hak akses file dan folder. Secara spesifik sistem perlu menyediakan penekanan atau kepastian mengubah hak akses secara langsung ke file di server. Mekanisme

seharusnya sudah termasuk pilihan untuk menekan perubahan secara jelas dengan administrator sistem, atau secara otomatis menggunakan pengatur.

3. Auditing

Audit secara rinci harus ada dengan memperhatikan semua aspek penggunaan data. Informasi harus dapat dilengkapi dan ditemukan kembali dengan mudah.

4. Keamanan

Keamanan harus tersedia secara otomatis, terutama dalam pengembalian hak akses data.

5. Penampilan.

6. Skala

Karena kegiatan organisasi dan pertumbuhan data tidak terstruktur tinggi, sistem harus menyediakan ruang untuk pertumbuhan untuk kegiatan setahun (12 bulan) maka *storage* harus disediakan 2 kali lipatnya.

7. Kemudahan dalam instalasinya.

8. Kemudahan dalam penggunaannya.

9. Kemudahan dalam pengintegrasian.

Aplikasi harus mendukung server dan peralatan tempat penyimpanan (*storage*). Tingkat kompatibilitas yang tinggi sangat penting untuk meningkatkan integritas antar aplikasi atau data.

10. Biaya keseluruhan rendah (*low total cost of ownership*).

Aplikasi harus dapat mendatangkan benefit yang dapat dihitung dari segi waktu dan efisiensi penggunaan sumber daya.

KESIMPULAN

Era industri 4.0 dengan pilar utama, yaitu *Internet of Things* (IoT), *cloud computing*, *artificial intelligence* dan *big data* telah memproduksi banyak sekali data. Berdasarkan pengelolaannya, data dibedakan menjadi 2 jenis, yaitu data terstruktur dan tidak terstruktur. Data terstruktur (*structured data*) merupakan data yang direpresentasikan ke dalam bentuk tabel atau relasi yang biasanya ditempatkan dalam suatu basis data. Sedangkan data tidak terstruktur (*unstructured data*) merupakan data yang tidak dapat dengan mudah diinterpretasikan sebelum diolah terlebih dahulu.

Memiliki banyak data yang tidak terstruktur, dan tidak adanya sistem yang mampu melacak data tersebut, membuat organisasi rentan terhadap beberapa risiko. Dengan adanya sistem pengelolaan basis data, data tidak terstruktur dapat dikelola dalam jumlah besar. Untuk menyimpan data tidak terstruktur ke dalam basis data, penggunaan tabel basis data Binary Large Objects atau BLOBs telah digunakan dalam beberapa decade. Untuk beberapa tahun Oracle telah mengintegrasikan jenis dan struktur data secara optimal dengan melakukan analisis dan manipulasi dokumen XML, multimedia content, teks dan informasi geospasial.

DAFTAR PUSTAKA

1. A.C. Eberendu. Unstructured Data : An Overview Of The Data Of Big Data. *International Journal of Computer Trends and Technology (IJCTT)* Vol. 38 Number 1, August 2016.
2. Afifanto, Choerul. (2015). Integrasi Data Terstruktur dan Tidak Terstruktur Dalam Sistem Intelegensi Bisnis. Available :
3. A. Tolic and A. Brodnik. Deduplication in Unstructured Data Storage Systems. *Elektrotehni~Ski Vestnik* 82(5): 233–242, 2015.
4. <https://apjii.or.id/>. (2017). Penetrasi dan Perilaku Pengguna Internet Indonesia. Retrieved from <https://apjii.or.id/>
5. I.R. Mansuri and S. Sarawagi. Integrated Unstructure Data Into Relational Databases. In *Proceeding of the 22nd International Conference on Data Engineering (ICDE'06)*, 2006.
6. Wahyunita, Laili. Rekayasa Web Klasifikasi Pada Data Tidak Terstruktur. *Jurnal Komunikasi, Media dan Informatika* Vol. 8 No. 2, November 2019.

ESSAY

MATA KULIAH

DATA BASE

Penggunaan Jaringan Bayesian dalam Informasi

Ekstraksi dari Sumber Data Tidak Terstruktur



D
I
S
U
S
U
N

OLEH

MUHAMMAD IQBAL RIVANA

192420057

MAGISTER TEKNIK INFORMATIKA
UNIVERSITAS BINA DARMA

Penggunaan Jaringan Bayesian dalam Informasi

Ekstraksi dari Sumber Data Tidak Terstruktur

A. Pendahuluan

1. Unstructure Data dan Structure Data

Data terstruktur (structured) adalah data yang berada dalam satu tempat baik berbentuk sebuah file termasuk data yang berada dalam database maupun spreadsheet. Data terstruktur adalah yang membuat model data. Contohnya adalah data CRM, Industry Research Data dan lain sebagainya. Data terstruktur biasanya dijalankan untuk mengakses database yang disebut dengan Structured Query Language atau yang lebih dikenal dengan SQL.

Sedangkan data tidak terstruktur (unstructured) juga data yang tidak mudah diklasifikasi dan dimasukkan kedalam sebuah kotak dengan rapi. Contohnya adalah foto, gambar grafis, streaming instrument data, webpages, PDF, Power Point presentations, konten blog dan lain sebagainya.

Karakteristik Data Tidak Terstruktur:

- a. Data tidak sesuai dengan model data atau memiliki struktur apa pun.
- b. Data tidak dapat disimpan dalam bentuk baris dan kolom seperti pada Database
- c. Data tidak mengikuti semantik atau aturan apa pun
- d. Data tidak memiliki format atau urutan tertentu
- e. Data tidak memiliki struktur yang mudah diidentifikasi
- f. Karena kurangnya struktur yang dapat diidentifikasi, itu tidak dapat digunakan oleh program komputer dengan mudah

Sumber-sumber Data Tidak Terstruktur antara lain adalah : Halaman web, Gambar (JPEG, GIF, PNG, dll.), Video, memo, laporan, dokumen Word dan persentasi PowerPoint, survei. Keuntungan dari Data Tidak Terstruktur adalah data ini mendukung data yg tidak memiliki format atau urutan yang tepat, juga data tidak mudah di batasi oleh skema, juga data sangat Fleksibel karena tidak adanya skema, data bersifat portabel, data sangat scalable, data ini

juga dapat dengan mudah berurusan dengan heterogenitas sumber, jenis data ini juga memiliki beragam aplikasi intelijen bisnis dan analitik.

Kerugian Dari Data Tidak Terstruktur:

Sulit untuk menyimpan dan mengelola data yang tidak terstruktur karena kurangnya skema dan struktur, pengindeksan data sulit dan rawan kesalahan karena struktur tidak jelas dan tidak memiliki atribut yang ditentukan sebelumnya. Karena yang hasil pencariannya tidak terlalu akurat, memastikan keamanan data adalah tugas yang sulit.

Masalah yang dihadapi dalam menyimpan data yang tidak terstruktur:, Ini membutuhkan banyak ruang penyimpanan untuk menyimpan data yang tidak terstruktur. Sulit untuk menyimpan video, gambar, audio, dll. Karena struktur yang tidak jelas, operasi seperti memperbarui, menghapus, dan mencari sangat sulit., biaya penyimpanan tinggi dibandingkan dengan data terstruktur, pengindeksan data tidak terstruktur sulit, solusi yang mungkin untuk menyimpan data tidak terstruktur:

B. PEMBAHASAN

1. DATA TIDAK TERSTRUKTURE PADA JARINGAN BAYESIAN

Bayesian Networks mendukung ekstraksi informasi dari yang tidak terstruktur, ungrammatis, dan sumber data tidak koheren untuk anotasi semantik. Alat telah dikembangkan yang menggabungkan ontologi, pembelajaran mesin, dan ekstraksi informasi dan teknik penalaran probabilistik untuk mendukung proses ekstraksi. Akuisisi data dilakukan dengan bantuan pengetahuan yang ditentukan dalam bentuk ontologi. Karena ukuran variabel informasi yang tersedia di berbagai sumber data, yaitu sering kali data yang diekstraksi berisi nilai yang hilang untuk variabel minat tertentu. Diinginkan dalam situasi seperti itu untuk prediksi nilai yang hilang.

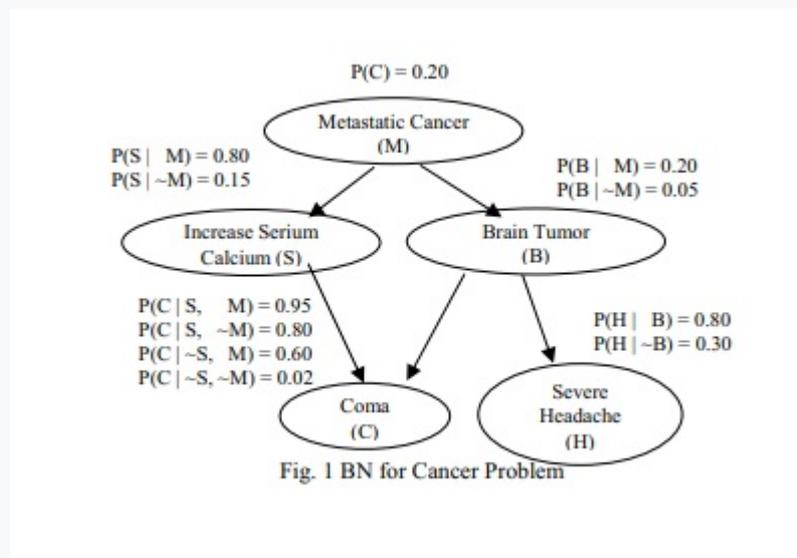
Metodologi yang disajikan dalam makalah ini, pertama-tama mempelajari jaringan Bayesian dari data pelatihan dan kemudian menggunakannya untuk memprediksi data yang hilang dan untuk menyelesaikan konflik. Eksperimen punya telah dilakukan untuk menganalisis kinerja yang disajikan metodologi. Hasilnya terlihat menjanjikan sebagai metodologi mencapai tingkat presisi tinggi dan penarikan informasi ekstraksi dan akurasi yang cukup baik untuk memprediksi hilang nilai-nilai.

Selama dua dekade terakhir, Bayesian Networks telah muncul sebagai alat paling populer untuk pemodelan dan alasan dalam domain yang tidak pasti. Mereka telah berhasil diterapkan di beberapa domain termasuk medis diagnosis, terorisme, peramalan, penggabungan informasi, sistem pemecahan masalah, dll. BN adalah representasi grafis dari probabilitas distribusi. Ini terdiri dari dua komponen. Yang pertama adalah a diarahkan grafik asiklik di mana setiap node mewakili a variabel acak, sedangkan himpunan busur yang menghubungkan pasangan node mewakili sifat independensi kondisional tertentu.

Komponen ini menangkap struktur probabilitas distribusi. Komponen kedua adalah koleksi parameter yang menggambarkan probabilitas bersyarat masing-masing variabel yang diberikan induknya dalam grafik. Bersama-sama, keduanya komponen mewakili distribusi probabilitas unik . Gambar 1 menunjukkan sampel BN untuk masalah diagnosis medis. Ada lima variabel acak (node). Tautan yang terhubung variabel yang berbeda menunjukkan ketergantungan langsung di antara ini variabel. Nilai yang dilampirkan ke node mewakili probabilitas sebelum dan kondisional yang sesuai. Secara umum, distribusi probabilitas gabungan variabel (X_1, X_2, \dots, X_n) dalam BN dihitung menggunakan distribusi lokal ini sebagai berikut memudahkan proses elisitasi pengetahuan. Secara matematis,

$$P(X_1, X_2, \dots, X_n) = \prod P(X_i | pa(X_i))$$

di mana $pa(X_i)$ mewakili jumlah himpunan dari variabel X_i



Gambar. 1

Secara tradisional praktiknya adalah membangun BN dengan mengakuisisi pengetahuan dari para ahli domain. Namun, jika data tersedia, BN dapat dipelajari dari data itu. Dalam beberapa tahun terakhir, banyak pekerjaan telah dilakukan untuk mempelajari struktur dan parameter BN dari kumpulan data. Dalam semua upaya tersebut, ketidakpastian tentang struktur a BN direpresentasikan sebagai distribusi probabilitas di atas semuanya struktur yang mungkin. Distribusi sebelumnya ditugaskan untuk masing-masing struktur dan distribusi lokal mengingat hal itu struktur. Kemudian aturan Bayes diterapkan untuk mendapatkan posterior distribusi untuk struktur dan parameter yang diberikan data D. Secara matematis,

$$P(S_i|D) = \frac{P(D|S_i)P(S_i)}{P(D)}$$

Di mana $P(D)$ adalah konstanta normalisasi yang tidak bergantung pada struktur apa pun.

$P(D|S_i)$ adalah kemungkinan marginal fungsi, $P(S_i|D)$ adalah distribusi posterior struktur diberikan data, dan $P(S_i)$ adalah probabilitas sebelumnya dari struktur.

Daftar Pustaka

1. www.geeksforgeeks.org/difference-between-structured-semi-structured-and-unstructured-data
2. www.geeksforgeeks.org/what-is-structured-data
3. Quratulain N. Rajput, Sajjad Haider, Jurnal Use of Bayesian Network in Information Extraction from Unstructured Data Sources (2009)