

BAB II

LANDASAN TEORI

2.1 *WordCloud*

WordCloud adalah teknik visualisasi lain yang memiliki tujuan yang sama sekali berbeda karena menyediakan gambaran umum kepada pengguna tentang isi kumpulan teks. Ukuran kata (seperti dalam ukuran font) menunjukkan seberapa sering kata tersebut muncul dalam koleksi teks, biasanya setelah menghapus apa yang disebut kata berhenti. Secara potensial, dengan semakin banyak teks yang dianotasi dengan lokasi yang dikaitkan dengan teks, kata cloud dari teks-teks itu mungkin dapat menambah informasi yang ditemukan di peta, dan bahkan memberikan perspektif yang berbeda tentang lokasi daripada yang ditemukan di peta. (Tessem, 2015)

2.2 *Most Frequent Word*

Most Frequent Word itu adalah Kata-kata frekuensi tinggi didefinisikan sebagai kata-kata yang muncul di semua jenis teks dan mencakup 80% teks itu sendiri, kata-kata yang sering muncul di seluruh teks dan disiplin ilmu dan ada di 2.000 daftar kata pertama yang paling sering, juga disebut sebagai *Most Frequent Words* (Nur, 2015)

2.3 *Beautiful Soup Python*

Library BeautifulSoup dinamai setelah puisi Lewis Carroll dengan nama yang sama di Petualangan Alice di Negeri Ajaib. Dalam ceritanya, puisi ini dinyanyikan oleh seorang tokoh disebut Mock Turtle. (itu sendiri pelesetan pada hidangan Victoria Mock Turtle Soup yang populer dibuat bukan dari kura-kura

tetapi dari sapi Seperti senama *Wonderland*, *BeautifulSoup* mencoba memahami tidak masuk akal; ini membantu memformat dan mengatur web yang berantakan dengan memperbaiki HTML yang buruk menyajikan kepada kita objek Python yang mudah dilintasi yang mewakili XMLstruktur. (Mitchell, 2015)

2.4 *Data Visualization*

Visualisasi data melibatkan penyajian data dalam bentuk grafik atau gambar yang membuat informasi mudah dimengerti. Ini membantu menjelaskan fakta dan menentukan arah tindakan. Ini akan menguntungkan semua bidang studi yang membutuhkan cara-cara inovatif untuk menyajikan informasi yang besar dan kompleks. Munculnya komputer grafis telah membentuk visualisasi modern. (Sadiku, 2016)

2.5 **Python**

Python adalah bahasa yang kuat yang dirancang untuk dengan mudah mengembangkan program sederhana. Dengan Python, ketika Anda ingin menulis sebuah program untuk menyapa seseorang, Anda menulis satu baris kode. Dalam bahasa lain, tugas sederhana ini bisa melibatkan banyak *overhead* (head) tambahan yang bisa membingungkan. *Syntax* Python membantu programmer baru memahami konsep-konsep penting tanpa menggunakan detail yang dapat membuat program sulit dibaca.

Python sering digunakan untuk mengajarkan keterampilan pemrograman dasar. Banyak konsep yang lebih esoteris dalam pemrograman telah dikesampingkan sehingga struktur data dasar dapat dipahami dengan sesedikit mungkin sesengolan. Ini juga bahasa yang bagus untuk berbagai tugas, dan saya yakin itu adalah cara sempurna untuk membuat Anda pemrograman dengan cepat dan percaya diri. (Grant, 2014)

Bahasa pemrograman Python dipilih karena bahasa ini mempunyai banyak keunggulan khususnya untuk pemrograman berbasis *machine learning*. Seperti

umumnya bahasa Java dan C, yang bersifat *Open Source*, interaktif, modular, dinamis, berbasis objek, dan lain-lain. (Karno, 2020)

2.6 Linux

Linux adalah sistem operasi. Bahkan, salah satu *platform* paling populer di planet ini, Android, ditenagai oleh sistem operasi Linux. Sistem operasi adalah perangkat lunak yang mengelola semua sumber daya perangkat keras yang terkait dengan desktop atau laptop Anda. Sederhananya, sistem operasi mengelola komunikasi antara perangkat lunak Anda dan perangkat keras Anda. Tanpa sistem operasi (OS), perangkat lunak tidak akan berfungsi. (Torvalds., 2019)

2.7 PuTTY

PuTTY adalah klien SSH dan telnet, yang dikembangkan oleh Simon Tatham untuk *platform* Windows. Putty adalah perangkat lunak sumber terbuka yang tersedia dengan kode sumber dan dikembangkan dan didukung oleh sekelompok relawan. (Tatham, 2019)

2.8 WinSCP

WinSCP adalah klien SFTP *open source* gratis, klien FTP, klien WebDAV, klien S3 dan klien SCP untuk Windows. Fungsi utamanya adalah transfer file antara komputer lokal dan remote. Selain itu, WinSCP menawarkan fungsi *scripting* dan manajer file dasar. (WinSCP.net, 2019)

2.9 R

R adalah bahasa dan lingkungan yang kuat untuk komputasi statistik dan grafik. Ini adalah *domain* publik (disebut "GNU") proyek yang mirip dengan bahasa S komersial dan lingkungan yang dikembangkan di Bell Laboratories (sebelumnya AT&T, sekarang Lucent Technologies) oleh John Chambers dan rekannya. R dapat dianggap sebagai implementasi, dan banyak digunakan sebagai bahasa pendidikan dan alat penelitian. (Paul Torfs, 2017)

2.10 Barplot

Barplot digunakan untuk menampilkan hubungan antara variabel numerik dan kategori. Bagian ini juga mencakup barplot bertumpuk dan barplot yang dikelompokkan di mana dua tingkat pengelompokan ditampilkan. (Holtz, 2018) dan untuk *Top 10* akan divisualisasikan *most frequent word* dalam bentuk barplot (Kassambra, 2020)

2.11 Power BI

Power BI Desktop menyertakan dua mesin (Power Pivot dan Power Query), serta lapisan visualisasi baru yang disebut Dasbor, Dasbor terlihat agak seperti Power View, tetapi sedangkan Power View agak membuat frustrasi setengah langkah, Dasbor Power BI sangat kuat / lengkap. Mereka menawarkan banyak jenis visualisasi yang tidak tersedia di Excel asli, daftar visualisasi bertambah dengan setiap rilis, dan mereka telah membuka *platform* sehingga pemrogram pihak ketiga dapat menambahkan jenis visualisasi kustom mereka sendiri. Power BI Desktop dan layanan *cloud* pendampingnya karenanya layak untuk bagian mereka sendiri (Collie, 2016)

2.12 Matrix

Aturan pemformatan bersyarat *default* dan kustom dapat diterapkan ke tabel dan matriks visual agar lebih mudah mengidentifikasi pengecualian dan nilai yang sama. Power BI saat ini mendukung skala warna latar belakang, skala warna font, dan pemformatan bersyarat bilah data untuk tabel dan visual matriks. (Powell, 2018)

2.13 Penelitian Terdahulu

Tabel 2 1 Penelitian Terdahulu

Nama Peneliti	Nama Jurnal	Judul Penelitian	Hasil Penelitian
Bjørnar Tessem	Journal of location based services – volume 9, 2015 – issue 4	Word cloud visualisation of locative information	Penelitian ini menjelaskan studi eksploratif tentang prototipe untuk visualisasi informasi lokatif dari Wikipedia dalam bentuk wordcloud.

Shadab Irfan	International Journal of Engineering Science and Computing, April 2018 Volume 8 Issue No.4	Web Mining for Information Retrieval	Penelitian ini menunjukkan bahwa banyak data yang berasal dari website dan tidak terstruktur, maka penelitian ini memberi step bagaimana teknik-teknik web mining dengan tepat
-----------------	---	--	--

Geetika Chawla	International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 21, October 2018	Big Data Analytics for Data Visualization: Review of Techniques	Penelitian ini menjelaskan bagaimana cara agar data yang banyak dan berantakan bisa divisualisasikan dengan baik
-------------------	--	--	--

Shipra Saini	International Journal of Computer Applications (0975 – 8887) Volume 118 – No. 18, May 2015	Review on Web Content Mining Techniques	Penelitian ini menjelaskan bagaimana cara mengelola data mulai dari gambar, teks maupun grafik dan menyajikan secara rinci dan komprehensif dengan pendekatan data terstruktur, tidak terstruktur, semi-terstruktur
--------------	--	---	---

Lonela MANIU	Sea – Practical Application of Science Volume VIII, Issue 24 (3/2020)	Insight From IT Jobs Market With Text Mining Approach	Jurnal ini menjelaskan tentang bagaimana cara mencari <i>Most Frequent Words</i> dari website dalam bentuk barplot dan juga <i>WordCloud</i>
-----------------	---	---	--

Berdasarkan jurnal diatas ada beberapa *tools* yang dapat digunakan untuk membuat visualisasi berikut ini adalah *tools* yang digunakan :

1. R : Grafik online, analitik, dan alat statis untuk individu serta *library* grafik ilmiah untuk R.
2. *BeautifulSoup* : *Library* yang digunakan untuk *data cleansing* website dalam bentuk *HTML*
3. Microsoft Power BI : Untuk visualisasi, laporan, dan dasbor interaktif, menggunakan kueri bahasa di dasbor.

Jadi berdasarkan dari *tools* penelitian sebelumnya yang paling tepat untuk mengolah untuk melakukan *data cleansing* maupun visualisasi data dalam bentuk barplot dan *wordcloud* adalah R dan *library beautifulsoup*, sedangkan untuk membuat tabel perbandingan matrix adalah Microsoft Power BI.