BAB III

TINJAUAN PUSTAKA

3.1. Extract, Transformation, Load (ETL)

Extract, Transformation, Load atau ETL adalah prosedur mengumpulkan atau menyalin data dari satu sumber atau lebih, mengubah data sesuai dengan tujuan, dan memuat data tersebut ke sistem tujuan yang merepresentasikan bentuk berbeda atau dalam konteks yang berbeda dari sumbernya. Proses transformasi pada ETL berlangsung di tempat khusus dan sering melibatkan pengunaan tabel sementara untuk menyimpan data yang sedang diubah dan akhirnya dimuat ke tujuannya.

Transformasi data biasanya melibatkan berbagai operasi, seperti *filtering*, *sorting*, *data cleansing*, dan *data validation*.

3.2. ELK Stack (Elasticsearch, Logstash, Kibana)

ELK Stack (sekarang dikenal sebagai Elastic Stack) adalah *log management platform* dan merupakan akronim dari tiga projek *open source*, yaitu Elasticsearch, Logstash, dan Kibana.

Elasticsearch adalah *search* dan *analytics engine* yang terdistribusi untuk semua jenis data, termasuk teks, numerik, *geospatial*, terstruktur, dan tidak terstruktur. Elasticsearch dibangun di atas Apache Lucene oleh Elasticsearch N.V. yang pertama kali dirilis pada tahun 2010.

Logstash adalah *server-side pipeline* pemrosesan data yang mengambil data dari satu atau banyak sumber secara bersamaan atau real-time, mentransformasikan data, dan mengirimkan data tersebut ke tujuan yang dipilih.

Kibana adalah aplikasi *front-end* yang menyediakan kemampuan pencarian dan visualisasi data untuk data yang telah diindeks di Elasticsearch. Kibana dikembangkan pada tahun 2013 oleh komunitas Elasticsearch.

Gambar 3.1 Logo ELK Stack [2]

3.3. Python

Python adalah bahasa pemrograman yang *high-level*, *general-purpose* dan dapat mendukung banyak paradigma pemrograman seperti *procedural*, *object-oriented*, dan *functional programming*. Python dibuat oleh Guido van Rossum dan pertama kali dirilis pada tahun 1991. Filsafat desain Python menekankan kode yang mudah dibaca manusia dengan pengunaan spasi yang signifikan.

Gambar 3.2 Logo Python [3]

3.4. Ruby

Ruby adalah bahasa pemrograman yang dinamis dan open source dengan fokus pada kesederhanaan dan produktivitas. Ruby memiliki syntax yang elegan,

alami untuk dibaca, dan mudah ditulis. Bahasa pemrograman ini dapat mendukung berbagai paradigma pemrograman seperti *reflective*, *object-oriented*, *functional*, dan *imperative programming*. Ruby dibuat oleh Yukihiro "Matz" Matsumoto dan pertama kali dirilis pada tahun 1995.

Gambar 3.3 Logo Ruby [4]

3.5. Named Entity Recognition

Named Entity Recognition atau NER adalah proses algoritma menerima string atau teks sebagai input, mencari dan mengklasifikasikan kata-kata yang relevan pada string tersebut berdasarkan kategori yang telah ditentukan seperti nama, organisasi, lokasi, jumlah, dll..

3.6. SpaCy

SpaCy adalah *open source natural languange processing library*. *Library* ini dikembangkan menggunakan hasil riset terbaru dan dirancang secara khusus untuk digunakan pada produk yang berada di tahap *production*. SpaCy dapat digunakan untuk membuat sistem *information extraction*, *named entity recognition*, atau untuk *pre-process* teks untuk *deep learning*.

Gambar 3.4 Logo SpaCy [5]

3.7. Features

Features adalah informasi yang merupakan pola tertentu atau fitur khusus yang unik pada gambar yang mudah dilacak dan dibandingkan. *Features* pada gambar dapat berupa struktur khusus seperti tepi, sudut, atau objek yang ada pada gambar tersebut. Algoritma *image processing* menggunakan informasi ini untuk memungkinkan komputer mengindentifikasi atau mengklasifikasi gambar. Untuk melakukan hal tersebut, ada beberapa metode yang digunakan algoritma *image processing* untuk mencari *features* pada suatu gambar.

Metode *feature detection* digunakan untuk mendeteksi dan mendapatkan *features* pada gambar. Metode *feature matching* untuk membandingkan *features* gambar dengan *features* gambar lain.

3.8. OpenCV

OpenCV adalah *computer vision* dan *machine learning library* yang *open source*. *Library* ini dibuat untuk menyediakan infrastruktur yang umum untuk aplikasi *computer vision* dan untuk mempercepat pengembangan dan pengunaan *computer vision* pada produk komersial. OpenCV memiliki lebih dari 2500 algoritma yang telah dioptimalkan dan mencakup serangkaian komprehensif algoritma *computer vision* yang klasik dan yang mutakhir. Algoritma-algoritma ini dapat digunakan untuk mendeteksi dan mengenali wajah, mengidentifikasi objek, dll.

Gambar 3.5 Logo OpenCV [6]

3.9. Oriented FAST and rotated BRIEF

Oriented FAST and rotated BRIEF atau ORB adalah algoritma *feature detection* yang telah disediakan oleh OpenCV. ORB pada dasarnya adalah perpaduan dari FAST *keypoint detector* dan BRIEF *descriptor* dengan modifikasi untuk meningkatkan performa. ORB dirancang sebagai alternatif untuk SIFT dan SURF yang lebih efisien. Algoritma ini dikembangkan oleh Ethan Rublee dan rekan pada tahun 2011.

3.10. Fast Library for Approximate Nearest Neighbors

Fast Library for Approximate Nearest Neighbors atau FLANN adalah *library* untuk melakukan pencarian *fast approximate nearest neighbor* di *high dimensional spaces*. *Library* ini berisi kumpulan algoritma yang berfungsi dengan baik untuk pencarian *nearest neighbor* dan FLANN memiliki sistem yang secara otomatis memilih algoritma terbaik dan parameter yang optimal tergantung pada *dataset* pengguna. FLANN dibuat dengan bahasa pemrograman C++ dan dapat digunakan dengan bahasa pemrograman lain seperti C, Python, MATLAB, dan Ruby.

3.11. Flask

Flask adalah WSGI *web application framework* yang ringan dan dirancang dengan kemampuan untuk memulai dengan cepat dan mudah. Flask berawal dari *wrapper* yang sederhana untuk Werkzeug dan Jinja dan telah menjadi salah satu *web application framework* Python yang paling populer. Flask dikembangkan oleh Armin Ronacher dan pertama kali dirilis pada tahun 2010.

Gambar 3.5 Logo Flask [7]

3.12. Pickle

Pickle adalah *standard library* atau *package* yang sudah termasuk dalam Python untuk melakukan *object serialization* dan *unserialization*. Semua objek dalam Python dapat di *pickled* sehingga dapat di simpan dalam *file*. Ketika objek di *pickled*, objek tersebut di *serialises* lalu di *write* ke suatu *file*.

3.13. Pip

Pip adalah *package manager* untuk meng-*install* dan mengelola *package*. Pip sudah termasuk dalam Python versi 3.4 atau yang lebih baru. Pertama kali diperkenalkan sebagai pyinstall dan dirilis pada tahun 2008 oleh Ian Bicking sebagai alternatif dari easy_install.

3.14. Visual Studio Code

Visual Studio Code atau VS Code adalah program pengolah *source-code* yang dibuat oleh Microsoft untuk membuat kode aplikasi. Program ini dapat mengolah berbagai bahasa pemrograman yang salah satunya adalah Python. VS Code memilliki *tools* dan *extensions* yang mempermudah developer untuk membuat suatu aplikasi dan menyusun kode aplikasi tersebut secara otomatis sehingga mudah dibaca, *package*, dan *deploy*.

Gambar 3.5 Logo Visual Studio Code [8]

3.15. RabbitMQ

RabbitMQ adalah aplikasi *open-source message broker* yang pertama kali mengimplementasikan Advanced Message Queuing Protocol. RabbitMQ dirancang untuk memungkinkan aplikasi lain terhubung satu sama lain dan pengiriman *message* yang dilakukan RabbitMQ bersifat asinkron. Message broker ini dirilis pertama kali pada tahun 2007 oleh Rabbit Technologies Ltd. dan dikembangkan dengan bahasa pemrograman Erlang dan *framework* Open Telecom Platform untuk *clustering* dan *failover*. *Library* klien untuk *interface* dengan RabbitMQ tersedia untuk semua bahasa pemrograman yang populer.

Gambar 3.5 Logo RabbitMQ [9]

3.16. Click

Click atau Command Line Interface Creation Kit adalah *package* Python untuk membuat *command line interface* dengan menggunakan kode sesedikit mungkin. *Package* ini memungkinkan pengembang script Python untuk *parse argument* melalui *command line interface* ke script tersebut. Click pertama kali dirilis pada tahun 2014 oleh The Pallets Projects.

Gambar 3.5 Logo Click [10]