



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

## BAB III

### TINJAUAN PUSTAKA

#### 3.1 Python

Python merupakan bahasa pemrograman yang multifungsi yang berarti, tidak seperti JavaScript dan lainnya, Python dapat digunakan untuk banyak tipe pemrograman dan pengembangan perangkat lunak disamping pengembangan web. Python diciptakan dan dikembangkan oleh Guido von Rossum dengan tujuan untuk memfokuskan bahasa pemrograman tersebut pada tingkat yang tinggi, memiliki arti semirip mungkin dengan bahasa manusia [2].

Python mendukung paradigma pemrograman yang beragam seperti pemrograman berorientasi objek, pemrograman imperatif, pemrograman fungsional, dan lainnya. Sebagai bahasa pemrograman dinamis, Python dilengkapi dengan fitur manajemen memori otomatis. Pada umumnya, Python digunakan sebagai bahasa *script* walaupun penggunaan bahasa ini dapat dimanfaatkan dalam hal yang umumnya tidak dilaksanakan dengan menggunakan bahasa *script*.

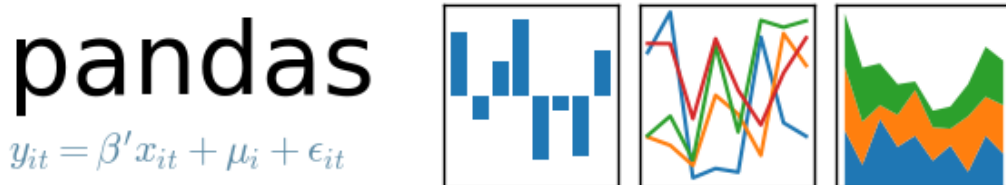
Python dapat dijalankan di atas beragam sistem operasi dan juga dapat dimanfaatkan untuk bermacam-macam keperluan pengembangan perangkat lunak. Python menggunakan sintaksis kode yang jelas dan disertai dengan fungsionalitas dokumentasi yang lengkap serta komprehensif. Karena sifatnya *open-source*, Python didukung oleh komunitas yang besar [3].



Gambar 3.1. Logo Python [4]

### 3.1.1 Pandas

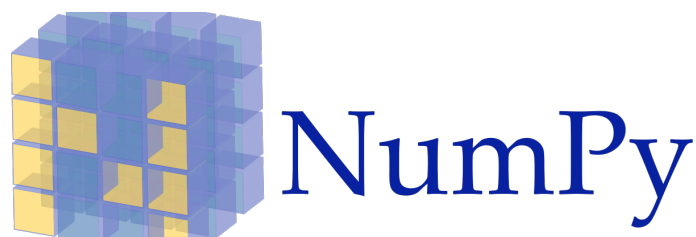
Pandas merupakan salah satu kepastakaan Python yang diciptakan untuk tujuan mempermudah dalam pelaksanaan analisis data. Tingkat kemudahan serta fleksibilitas Pandas memberikan dampak besar terhadap banyak penggunanya terlebihnya terhadap yang berhubungan erat dengan analisis data. Salah satu fitur Pandas yang banyak diminati oleh penggunanya merupakan Pandas mendukung data *wrangling* [5]. *Data Wrangling* merupakan suatu proses transformasi serta *mapping* data dari sebuah data mentah menjadi format yang lebih rapi dimana dalam salah satu kasus penggunaan Pandas yaitu proses pengolahan data baik itu CSV/TSV ataupun SQL menjadi objek Python dengan baris dan kolom yang biasa dikenal dengan istilah *DataFrame*.



Gambar 3.2. Logo Pandas [6]

### 3.1.2 Numpy

NumPy merupakan sebuah package pada Python yang digunakan untuk melakukan komputasi saintifik seperti operasi penjumlahan, pengurangan, perkalian dan sebagainya.



Gambar 3.3. Logo Numpy [7]

### 3.1.3 Scikit-learn

Sklearn merupakan sebuah *library* Python yang menyediakan beragam algoritma untuk klasifikasi, pengelompokan, regresi termasuk *random forrest*, *gradient boosting* dimana *library* ini juga dapat melakukan integrasi antaroperasi dengan *library* saintifik dan numerik lainnya seperti NumPy dan SciPy.



Gambar 3.4. Logo Scikit-learn [8]

### 3.1.4 Psycopg2

Psycopg2 merupakan *library* Python yang berfungsi sebagai *database* adapter PostgreSQL agar dapat digunakan dalam bahasa pemrograman Python. memberikan akses kepada penggunanya agar dapat menggunakan bahasa pemrograman Python pada database PostgreSQL. Psycopg2 di implementasikan dalam C sebagai *libpq wrapper* membuatnya lebih efisien dan juga aman.



Gambar 3.5. Logo Psycopg [9]

### 3.2 PostgreSQL

PostgreSQL merupakan sebuah DBMS object-relational multiguna yang bersifat *open-source*. PostgreSQL memungkinkan penggunanya untuk menambahkan fungsi kustom yang dikembangkan menggunakan bahasa pemrograman seperti C/C++, Java dan lain-lain.



Gambar 3.6. Logo PostgreSQL [10]

### 3.3 Jupyter Lab

Jupyter Lab bertujuan untuk mempermudah penggunanya dalam melakukan pengembangan suatu aplikasi dengan bahasa pemrograman apapun dimana penggunaan Jupyter Lab yang merupakan antarmuka untuk proyek Jupyter ini telah mencakup fitur-fitur seperti Jupyter Notebook, Console, Python *script*, dan lainnya.



Gambar 3.7. Logo Jupyter Lab [11]

### 3.3.1 Jupyter Notebook

Jupyter Notebook merupakan sebuah aplikasi web yang memungkinkan pengguna untuk membuat berkas yang berisi *script* kode, persamaan, visualisasi, dan juga teks naratif. Penggunaannya dapat berupa *data cleaning* dan transformasi, simulasi numerik, pembuatan model statistik, visualisasi data, *machine learning*, dan masih banyak lainnya.



Gambar 3.8. Logo Jupyter Notebook [12]

### 3.4 Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Elastic Compute Cloud (Amazon EC2) merupakan mesin virtual yang disediakan oleh Amazon Web Services, yang memungkinkan pengguna untuk dapat melakukan komputasi pada *cloud* yang memiliki kapasitas yang *scalable*. Dengan menggunakan Amazon EC2 pengguna dapat menggunakan *resource* sebesar yang ia perlukan, dengan komputasi yang lebih cepat, dan dapat mengkonfigurasi keamanan jaringan serta kapasitas penyimpanannya.



Gambar 3.9. Logo Amazon EC2 [13]

### 3.5 Amazon Simple Storage Service (Amazon S3)

Amazon Simple Storage Service (Amazon S3) merupakan servis penyimpanan pada *cloud* berbasis web, yang *scalable*, berkecepatan tinggi dan juga dirancang untuk *backup* daring, dan mengarsipkan data dan aplikasi pada Amazon Web Services. Amazon S3 dibuat dengan kumpulan fitur yang minimal untuk mempermudah komputasi berskala web bagi pengembang.



Gambar 3.10. Logo Amazon S3 [14]

### 3.6 Amazon Redshift

Amazon Redshift merupakan sistem manajemen basis data relasional yang dibangun dengan menggunakan prinsip PostgreSQL. Amazon Redshift memiliki efisiensi kueri yang lebih dari *petabyte* data dan dapat diatur secara keseluruhan.



Gambar 3.11. Logo Amazon Redshift [15]

### 3.7 DBeaver

DBeaver merupakan alat manajemen *database* yang *multi-platform* dan gratis untuk pengembang, *SQL programmer*, *administrator database*, dan juga

analisis. DBeaver diprogram dengan menggunakan bahasa pemrograman Java dan mendukung banyak variasi *database* di antaranya adalah MySQL, PostgreSQL, MariaDB, dan lainnya.



Gambar 3.12. Logo DBeaver [16]

### 3.8 Ubuntu

Ubuntu merupakan salah satu sistem operasi distribusi linux yang bersifat *open-source* dan berbasis pada Debian. Ubuntu telah dirilis dalam tiga edisi yaitu *desktop*, *server*, dan juga *core*. Semua edisi tersebut dapat berjalan pada komputer maupun mesin virtual. Ubuntu terkenal sebagai sistem operasi yang sesuai untuk *cloud computing*, dan mendukung OpenStack.



Gambar 3.13. Logo Ubuntu [17]

### 3.9 Slack

Slack merupakan aplikasi komunikasi lintas *platform* yang digunakan dalam mengkoordinasikan pengerjaan suatu proyek. Dengan menggunakan Slack, beberapa orang dapat bekerja dalam satu kanal dimana mereka dapat saling



membagikan informasi beserta berkas-berkas yang diperlukan satu sama lain.



Gambar 3.14. Logo Slack [18]

### 3.10 Git

Git adalah *version-control system* yang dikembangkan oleh Linus Torvalds. Git memungkinkan penggunaannya untuk mengelola berbagai versi *source code* pada pengembangan aplikasi perangkat lunak. Git digunakan untuk kolaborasi dan koordinasi antara banyak kontributor dalam sebuah proyek. Git didistribusikan secara gratis dan terdapat banyak *repository* berbasis Git seperti GitHub, GitLab, dan BitBucket.



Gambar 3.15. Logo Git [19]

#### 3.10.1 Github

GitHub adalah sebuah titik temu atau kumpulan beragam jenis repository yang bersifat terbuka. Github digunakan untuk manajemen proyek dan sekaligus platform jaringan sosial yang dibuat untuk pengembang. Github memungkinkan penggunaannya untuk bekerja secara kolaboratif dengan orang lain di seluruh dunia, merencanakan proyek dan melacak pekerjaan.



Gambar 3.16. Logo Github [20]

### 3.11 JIRA

JIRA adalah sebuah perangkat lunak berbasis web yang dikembangkan oleh Atlassian. JIRA digunakan untuk pelacakan *bug*, pelacakan masalah, dan manajemen proyek. Penggunaan dasar aplikasi ini adalah untuk melacak masalah dan bug yang terkait dengan perangkat lunak dan aplikasi pengguna. Dasbor JIRA terdiri dari banyak fungsi dan fitur yang berguna untuk memudahkan penanganan masalah.

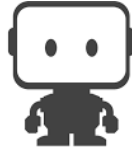


Gambar 3.17. Logo JIRA [21]

### 3.12 DataRobot

DataRobot adalah perangkat lunak analisis prediktif tingkat perusahaan untuk *business analysts*, *data scientists*, *executives*, dan *IT professionals*. DataRobot memanfaatkan kekuatan *machine learning* untuk memberdayakan pengguna agar secara konsisten membuat keputusan bisnis yang cepat, sederhana, dan akurat. DataRobot menganalisis berbagai algoritma pembelajaran mesin yang inovatif untuk membangun, menerapkan, dan membangun model prediksi untuk setiap situasi. Beberapa fitur yang menonjol termasuk *machine learning*, berbagai

jenis *database*, ETL dan alat visualisasi, persiapan data, dan kumpulan dari algoritma.



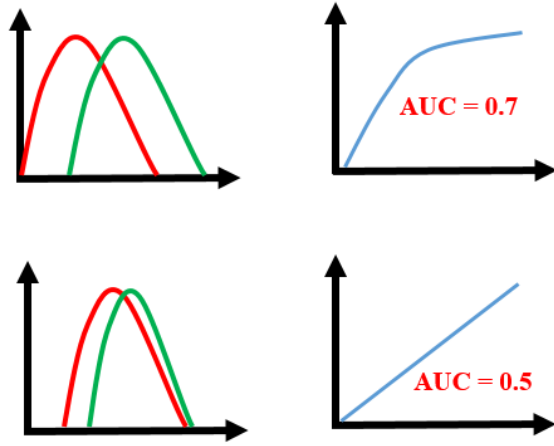
**DataRobot**

Gambar 3.18. Logo DataRobot [22]

### 3.13 Metrik Evaluasi Kurva AUC-ROC

Pada *machine learning* pengukuran performa merupakan hal yang esensial. Kurva AUC-ROC merupakan tolak ukur yang dapat diandalkan dalam model prediksi *classification* dan merupakan metrik evaluasi yang paling penting dalam memeriksa performa dari model *classification*. Kurva AUC - ROC adalah pengukuran performa untuk masalah *classification* di berbagai pengaturan *thresholds*. ROC adalah kurva probabilitas dan AUC mewakili derajat atau ukuran dari sebuah keterpisahan. Ini menunjukkan seberapa mampu model membedakan antara kelasnya [23]. Semakin tinggi AUC berarti semakin baik model dalam membedakan antarkelas dan memprediksi pun menjadi lebih akurat. Dengan analogi, semakin tinggi nilai AUC, semakin baik modelnya membedakan antara peminjam yang baik dan peminjam yang berisiko.

Pada saat AUC bernilai 1 atau dapat dinyatakan dalam persen yaitu 100%, adalah kondisi ideal dimana model memiliki ukuran keterpisahan yang ideal, secara sempurna dapat membedakan antarkelasnya. Pada saat AUC bernilai di antara 1 dan 0.5, Sedangkan pada saat AUC bernilai 0.5 atau dapat dinyatakan dalam persen yaitu 50%, adalah kondisi terburuk model karena model tidak memiliki kemampuan diskriminasi sama sekali untuk membedakan antarkelasnya.



Gambar 3.19. AUC bernilai 0.7 dan 0.5 [24]