



## BAB II

### LANDASAN TEORI

#### © Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

#### A. Tinjauan Pustaka

##### 1. Data

Menurut James O'Brien dan George Marakas (2013: 32), definisi data yaitu:

*“Data is the plural of datum, although data commonly represents both singular and plural forms. Data are raw facts or observations, typically about physical phenomena or business transactions.”*

Data merupakan bentuk jamak dari datum, meskipun dalam penggunaan sehari-hari, data sering digunakan untuk menyatakan bentuk jamak maupun tunggal. Data dapat diartikan sebagai fakta mentah atau observasi, yang biasanya berhubungan dengan fenomena fisik atau transaksi bisnis. Data harus melalui 5 aktivitas utama untuk diproses, hal ini meliputi:

##### a. Input

Input biasanya berbentuk aktivitas penginputan data seperti perekaman dan pengeditan. Pengguna biasanya memasukkan data langsung ke dalam sistem komputer atau merekam data pada jenis media fisik tertentu seperti formulir kertas. Penginputan ini termasuk berbagai aktivitas pengeditan untuk memastikan bahwa mereka telah melakukan masukan data dengan benar. Setelah dimasukkan, data dapat ditransfer ke media yang dapat dibaca mesin, seperti disk magnetik, sampai diperlukan untuk pengolahan.

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



## b. Pemrosesan

Pada tahap ini dilakukan aktivitas pengolahan, seperti perhitungan, perbandingan, pengurutan, pengklasifikasian, dan ringkasan. Aktivitas-aktivitas ini berfungsi untuk mengatur, menganalisis, dan memanipulasi data, sehingga dapat menjadi informasi untuk pengguna akhir. Kualitas dari data yang disimpan dalam sistem informasi harus dijaga untuk dipertahankan melalui proses yang dilakukan terus-menerus melalui pengkoreksian dan pembaruan.

## c. Output

data yang telah menjadi informasi kemudian disajikan kepada pengguna akhir dengan produk informasi yang sesuai. Produk informasi yang umum seperti pesan, laporan, formulir, dan gambar grafis, video, respons audio, produk kertas, dan multimedia.

## d. Penyimpanan

aktivitas sistem informasi di mana data disimpan dengan cara yang terorganisir untuk digunakan di masa depan.

## e. Kontrol

memastikan tahap-tahap sebelumnya telah berjalan dengan sesuai dan melakukan pemantauan dan evaluasi terhadap umpan balik untuk menentukan apakah sistem telah memenuhi standar kinerja yang ditetapkan. Kemudian aktivitas sistem yang sesuai harus disesuaikan sehingga produk informasi yang tepat dapat diproduksi untuk pengguna akhir.

Menurut Zainal A. Hasibuan, PhD (2007 : 135), ada dua cara memperoleh data,

yaitu:

### © Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



a. **Data Primer**

Data primer adalah data yang diambil langsung dari objek penelitian atau merupakan data yang berasal dari sumber asli atau pertama. Data ini tidak tersedia dalam bentuk file. Data primer tersebut harus dicari melalui narasumber atau responden yaitu orang yang kita jadikan objek penelitian atau orang yang kita jadikan sebagai sarana mendapatkan informasi maupun data. Pencarian data primer bisa dilakukan dengan cara wawancara atau *interview* langsung dengan responden, melalui telepon, email dan sebagainya. Dalam mengumpulkan data primer, biasanya peneliti menggunakan instrumen penelitian yang disebut dengan kuesioner.

b. **Data Sekunder**

Data sekunder adalah data yang tidak didapatkan secara langsung dari objek penelitian, melainkan data yang berasal dari sumber yang telah dikumpulkan oleh pihak lain. Data sekunder bisa diperoleh dengan cepat dan mudah karena data ini biasanya sudah tersedia dan kita tinggal mengambil dan mengumpulkan saja. Data sekunder dapat kita kumpulkan dari perpustakaan, perusahaan-perusahaan, organisasi-organisasi perdagangan, biro pusat statistik, dan kantor-kantor pemerintahan seperti data dari Badan Pusat Statistik (BPS), data hasil riset, data dari perusahaan dan lain sebagainya.

Menurut Zainal A. Hasibuan, PhD (2007 : 135), terdapat dua jenis data berdasarkan sumbernya, yaitu:

a. **Data Internal**

Data internal adalah data yang menggambarkan situasi dan kondisi pada suatu organisasi secara internal.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



## b. Data Eksternal

Data eksternal adalah data yang menggambarkan situasi serta kondisi yang ada di luar organisasi.

Menurut Zainal A. Hasibuan, PhD (2007 : 136), terdapat dua jenis data menurut waktu yaitu:

### a. Data Cross Section

Data cross-section adalah data yang menunjukkan titik waktu tertentu. Contohnya laporan keuangan per 31 Desember 2006, data pelanggan PT. PLN bulan April 2006, dan lain sebagainya.

### b. Data Time Series

Data berkala adalah data yang datanya menggambarkan sesuatu dari waktu ke waktu atau periode secara historis. Contoh *data time series* adalah data perkembangan nilai tukar rupiah terhadap dollar amerika dari tahun 2005 sampai 2006.

## © Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

## 2. Informasi

Menurut Reynolds dan George (2022:4-6), definisi informasi yaitu:

*“Information is a collection of data organized and processed so that it has additional value beyond the value of the individual facts.”*

Informasi merupakan kumpulan data yang terorganisir dan diproses sehingga memiliki nilai tambahan di luar dari nilai yang dimiliki oleh fakta-fakta secara individu. Informasi yang berkualitas dapat dibedakan berdasarkan karakteristiknya, yaitu:



1. Mudah Diakses, informasi harus mudah diakses oleh pengguna yang berwenang sehingga dapat diperoleh dalam format yang tepat dan pada waktu yang tepat untuk memenuhi kebutuhan mereka.
2. Akurat, informasi yang akurat bebas dari kesalahan. Dalam beberapa kasus, informasi yang tidak akurat dihasilkan karena data yang tidak akurat dimasukkan ke dalam proses transformasi. Hal ini biasa disebut sampah masuk, sampah keluar.
3. Lengkap, informasi yang lengkap berisi semua fakta penting.
4. Ekonomis, informasi juga harus relatif ekonomis untuk diproduksi. Para pengambil keputusan harus selalu menyeimbangkan nilai informasi dengan biaya untuk memproduksinya.
5. Fleksibel, informasi yang fleksibel dapat digunakan untuk berbagai tujuan.
6. Relevan, informasi yang relevan adalah informasi yang dapat digunakan untuk pengambil keputusan.
7. Dapat Diandalkan, informasi yang dapat diandalkan adalah informasi yang dapat dipercaya oleh pengguna. Dalam banyak kasus, keandalan informasi bergantung pada metode pengumpulan datanya ataupun sumber informasi.
8. Aman, informasi harus aman dari akses oleh pengguna yang tidak berwenang.
9. Sederhana, informasi harus sederhana, tidak rumit. Informasi yang lengkap dan terperinci mungkin tidak diperlukan. Bahkan, terlalu banyak informasi dapat menyebabkan informasi yang berlebihan sehingga menyulitkan pengambil keputusan karena terlalu banyak informasi dan tidak dapat menentukan informasi mana yang penting.
10. Tepat waktu, informasi yang tepat waktu dapat disajikan pada saat dibutuhkan.

**C Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)**

Hak Cipta Dilindungi Undang-Undang

**Institut Bisnis dan Informatika Kwik Kian Gie**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



11. Dapat diverifikasi, informasi harus dapat diverifikasi. Artinya, seseorang dapat memeriksanya untuk memastikan kebenarannya, mungkin dengan memeriksa banyak sumber untuk informasi yang sama.

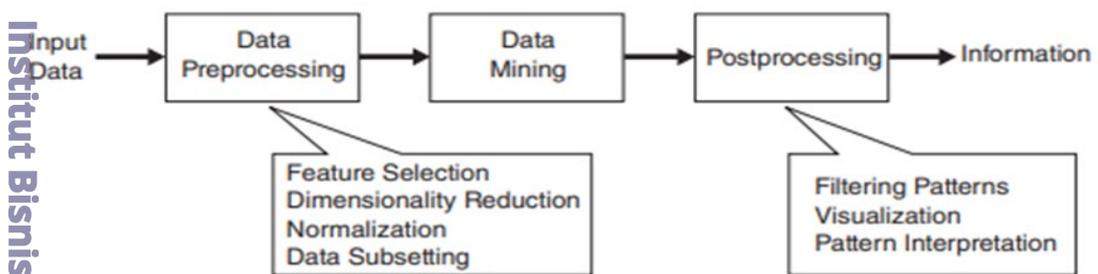
**3. Data Mining**

Hak Cipta Dilindungi Undang-Undang

Menurut Pan-Ning Tan (2019:24) definisi *data mining* yaitu:

“Data mining is the process of automatically discovering useful information in large data repositories”

*Data mining* adalah sebuah proses otomatis mencari informasi yang berguna dalam penyimpanan data yang besar. Data mining merupakan bagian integral dari knowledge discovery in database (KDD), yaitu suatu proses keseluruhan yang mengkonversi data mentah menjadi informasi yang berguna, seperti pada Gambar 2.1 Pada proses ini memiliki beberapa tahap, mulai dari data *preprocessing* hingga *postprocessing*.



**Gambar 2.1**  
**Proses Data Mining**  
Sumber : Pan Ning-Tan (2019:27)

*Input* data dapat berupa berbagai format seperti (file datar, spreadsheet, atau table relational) serta dapat diletakan pada penyimpanan data terpusat atau penyimpanan data yang terdistribusi di beberapa lokasi. Tujuan dari *preprocessing* adalah untuk mengubah input data menjadi format yang sesuai untuk analisis selanjutnya.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Langkah-langkah yang terlibat pada *preprocessing* meliputi penggabungan dari dari berbagai sumber, pembersihan data untuk menghilangkan noise dan duplikasi data, serta memilih *records* dan fitur yang relevan dengan tugas *data mining*.

Tugas dari *data mining* dibagi menjadi dua kategori yaitu:

a. **Predictive Tasks**

Memiliki tujuan untuk memprediksi nilai atribut tertentu berdasarkan nilai atribut lainnya. Atribut yang akan diprediksi biasanya dikenal sebagai variabel target atau dependen, sedangkan atribut yang digunakan untuk membuat prediksi dikenal sebagai variabel penjelas atau independen. (Pan Ning Tan 2019:29).

b. **Descriptive Tasks**

Memiliki tujuan untuk mendapatkan pola (korelasi, tren, cluster, lintasan, dan anomali) yang meringkas hubungan berdasarkan dari data. *Descriptive data mining tasks* sering kali bersifat eksploratif dan sering kali membutuhkan teknik pasca-pemrosesan untuk memvalidasi dan menjelaskan hasilnya. (Pan Ning Tan 2019:29).

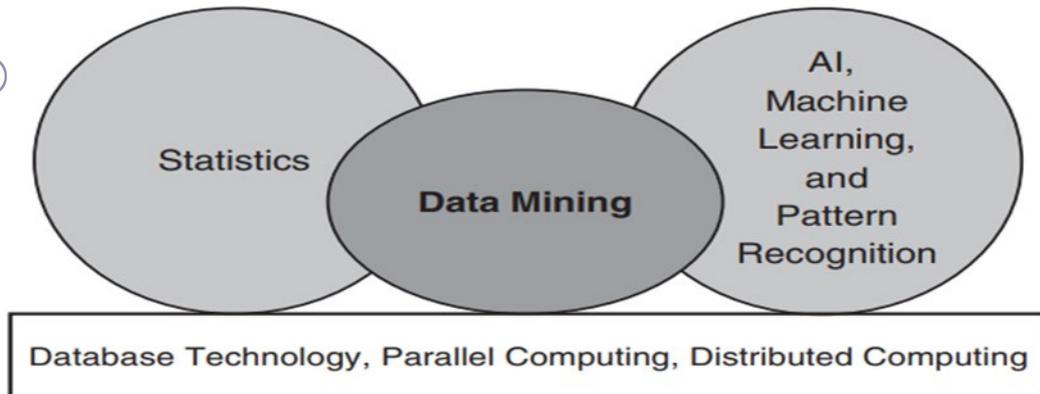
*Data mining* membantu dalam bidang lainnya seperti statistik dengan pengujian hipotesis dan estimasi sampel, untuk *artificial intelligence* seperti algoritma pencarian, teknik *modeling*, *machine learning* dan pengenalan pola. Pada gambar 2.2 menunjukkan relasi antar bidang dari *data mining*:

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



**Gambar 2.2**  
**Relasi Data Mining Antar Bidang**  
Sumber : Pan Ning-Tan (2019:27)

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

Data mining membantu dalam bidang lainnya seperti statistik dengan pengujian hipotesis dan estimasi sampel, untuk *artificial intelligence* seperti algoritma pencarian, teknik *modeling*, *machine learning* dan pengenalan pola. Berikut gambar yang menunjukkan relasi antar bidang dari *data mining*:

a. **Prediksi (*Predictive*)**

Teknik prediksi digunakan apabila suatu nilai memiliki atribut yang berbeda, contoh algoritmanya seperti Algoritma Linear Regression, Neural Network dan lain-lain.

b. **Asosiasi (*Association*)**

Teknik Asosiasi digunakan untuk hubungan antar data, contoh algoritmanya seperti Algoritma Apriori.

c. **Klastering (*Clustering*)**

Teknik clustering digunakan untuk pengelompokan data dalam suatu kelompok tertentu, contoh algoritmanya seperti K-Means, K-Medoids, Self Organization Map(SOM), Fuzzy C-Means.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



#### d. **Klasifikasi (*Classification*)**

Teknik klasifikasi mengelompokkan data yang mempunyai variabel tertentu berbeda dengan clustering yang tidak memiliki variabel yang dependen, contoh algoritma seperti ID3 dan K Nearest Neighbour.

### © Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

#### 4. **Machine Learning**

Menurut Andreas C.Muller (2017:1), definisi *machine learning* yaitu:

*“ is about extracting knowledge from data. It is a research field at the intersection of statistics, artificial intelligence, and computer science and is also known as predictive analytics or statistical learning.”*

*Machine learning* merupakan suatu proses untuk mengekstraksi pengetahuan berdasarkan data. Bidang penelitian ini berkaitan dengan statistik, *artificial intelligence*, dan komputer sains dan juga dikenal sebagai analitik prediktif atau pembelajaran statistik.

#### 5. **Saham**

Menurut Paul Mladjenovic (2020:8), definisi *data mining* yaitu:

*“Stock is a type of security that indicates ownership in a corporation and represents a defined portion (measured in shares) of that corporation’s future success.”*

Saham adalah jenis sekuritas yang menunjukkan kepemilikan dalam sebuah perusahaan dan mewakili bagian tertentu (diukur dalam bentuk saham) dari kesuksesan perusahaan di masa depan. Menurut sumber terdapat 2 jenis saham yang digunakan secara umum, yaitu:

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



a. **Saham Biasa (*Common Stock*)**

Saham yang memberikan hak kepada pemiliknya untuk memberikan suara dalam rapat pemegang saham dan menerima dividen yang dikeluarkan perusahaan.

b. **Saham Preferen (*Preferred Stock*)**

Saham yang biasanya tidak memberikan hak suara, tetapi memiliki beberapa hak yang melebihi hak-hak saham biasa. Pemegang saham preferen, misalnya, memiliki perlakuan istimewa dalam kondisi tertentu, seperti menerima dividen sebelum pemegang saham biasa jika terjadi likuidasi atau kebangkrutan perusahaan. Selain itu, saham preferen juga memiliki fungsi yang mirip dengan obligasi bagi investor yang mencari pendapatan stabil.

**6. Algoritma Long Term-Short Memory (LSTM)**

Menurut Ian Goodfellow, Yoshua Bengio, et al (2017:410) definisi long term-short memory yaitu:

*“The clever idea of introducing self-loops to produce paths where the gradient can flow for long durations is a core contribution of the initial long short-term memory (LSTM).”*

Ide cerdas untuk memperkenalkan *self-loop* untuk menghasilkan jalur di mana gradien dapat mengalir dalam jangka waktu yang lama merupakan kontribusi inti dari jangka pendek awal (LSTM). Dengan membuat bobot dari *self-loop* ini terjaga (dikontrol oleh unit tersembunyi lainnya), skala waktu integrasi dapat diubah secara dinamis. Meskipun LSTM memiliki parameter yang tetap, skala waktu integrasi dapat berubah berdasarkan urutan input, karena konstanta waktu adalah output dari model itu sendiri.

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



## 7. Algoritma Regresi Linear

Menurut Ibnu Daqiqil (2022:28) definisi Regresi Linear yaitu:

*“Regresi adalah salah satu teknik dalam , dimana teknik ini berasal dari ilmu statistik untuk mencari parameterparameter dalam persamaan linear yang dapat memetakan input dan output.”*

Analisis regresi setidaknya memiliki 3 kegunaan, yaitu mendeskripsi fenomena data atau kasus yang sedang diteliti; untuk tujuan kontrol; serta untuk tujuan prediksi.

Regresi mampu mendeskripsikan fenomena data melalui terbentuknya suatu model hubungan yang bersifat numerik.

## 8. Time Series Forecasting

Menurut Charu C. Aggarwal (2015 : 464 ) definisi Time Series Forecasting yaitu:

*“Forecasting is one of the most common applications of time series analysis. The prediction of future trends has applications in retail sales, economic indicators, weather forecasting, stock markets, and many other application scenarios. In this case, we have one or more series of data values, and it is desirable to predict the future values of the series using the history of previous values.”*

Forecasting merupakan salah satu aplikasi yang paling umum dari *timeseries analysis*. Aplikasi prediksi tren masa depan ditemukan pada berbagai bidang, seperti penjualan ritel, indikator ekonomi, prakiraan cuaca, pasar saham, dan banyak skenario aplikasi lainnya. Dalam hal ini, apabila terdapat satu atau lebih rangkaian nilai data, maka prediksi nilai masa depan dari rangkaian tersebut dapat menggunakan riwayat nilai sebelumnya.

## 9. Python

Menurut Pajankar dan Joshi (2022:28) definisi Python yaitu:

*“Python is a modern programming language. It has features of object-oriented and procedural programming. It runs on a variety of platforms. And the major*

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



*advantage is that the code written on one platform runs on the other platform without any major changes to the code (except for the platform-specific code).”*

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

Python adalah bahasa pemrograman modern. Bahasa ini memiliki fitur pemrograman berorientasi objek dan prosedural. Bahasa pemrograman ini dapat berjalan di berbagai platform. Keuntungan utama dari Python adalah kode yang ditulis di satu platform dapat berjalan di platform lain tanpa memerlukan perubahan besar pada kode.

### 10. Orange

Orange Data Mining adalah perangkat lunak *open source* yang digunakan untuk analisis data dan pemodelan prediktif. Orange dirancang untuk membantu pengguna, terutama yang tidak memiliki pengetahuan tentang pemrograman, untuk memahami dan menganalisis data secara visual.

Berikut contoh yang bisa dilakukan pada Orange Data Mining:

#### a. *Interactive Data Visualization*

Orange hadir dengan visualisasi data yang membantu menemukan pola data tersembunyi, memberikan intuisi dalam prosedur analisis data, atau mendukung komunikasi antara ilmuwan data dan ahli domain. *Widget* visualisasi yang ditawarkan mencakup diagram pencar, diagram kotak dan histogram, serta visualisasi-model khusus seperti dendrogram, scatter plot, dan visualisasi tree, di samping visualisasi lainnya yang tersedia dalam *add-on* seperti visualisasi jaringan, Word Cloud, peta geografis, dan masih banyak lagi.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



b. **Visual Programming**

Orange alat penambangan data yang sangat baik untuk pemula maupun ahli ilmu data. Berkat antarmuka pengguna yang dimilikinya, pengguna dapat fokus pada analisis data daripada melakukan coding yang melelahkan, sehingga membuat konstruksi jalur analitik data yang kompleks menjadi sederhana.

c. **Add-ons Extend Functionality**

Pengguna dapat memanfaatkan *add-on* yang tersedia dalam Orange untuk mengekstraksi data dari sumber data eksternal, melakukan pengolahan bahasa alami dan *text mining*, melakukan analisis jaringan, mengekstraksi itemset yang sering muncul, serta melakukan penambangan aturan asosiasi.

**© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)**

Hak Cipta Dilindungi Undang-Undang

## 11. Penelitian Kuantitatif

Menurut Umesh Kumar dan D.P. Kothari (2022:6) definisi penelitian kuantitatif yaitu:

*“Quantitative research is variables based whereas qualitative research is attributes based. Quantitative research is based on measurement or quantification of the phenomenon under study. That is , it is data based and hence more objective and popular.”*

Penelitian kuantitatif berbasis variabel sedangkan penelitian kualitatif berbasis atribut. Penelitian kuantitatif didasarkan pada pengukuran atau kuantifikasi dari fenomena yang diteliti. Artinya, ini berbasis data dan karena lebih objektif dan populer.



## B. Penelitian Terdahulu

### 1. Perbandingan Algoritma Linear Regression, LSTM, dan GRU dalam Memprediksi Harga Saham Dengan Model Time Series

Penelitian yang telah dilakukan oleh Khalis Sofi, Aswan Supriyadi Sunge, Sasmitoh Rahmad Riady, dan Antika Zahrotul Kamalia adalah pelatihan menggunakan algoritma Linear Regression, LSTM, dan GRU terhadap data harga saham KEJU yang berbentuk *time series* dari tanggal 15 November 2019 sampai dengan 08 Juni 2021 dengan pengukuran nilai RMSE, MSE, dan MAE. Setelah dilakukan proses *training* dan *testing*, hasil yang diperoleh adalah bahwasanya algoritma GRU memiliki *performance* terbaik dibandingkan dua algoritma lainnya dalam memprediksi harga saham, dibuktikan dengan nilai RMSE, MSE, dan MAE dari uji coba GRU paling rendah, yaitu RMSE 0.034, MSE 0.001, dan MAE 0.024.

### 2. Komparasi Metode Multilayer Perceptron (MLP) dan Long Short Term Memory (LSTM) dalam Peramalan Harga Beras

Penelitian yang dilakukan oleh Steven Sen, Dedy Sugiarto, dan Abdul Rochman adalah prediksi terhadap data *time series* harga beras pada PT. FoodStation, perbandingan dilakukan dengan model Multilayer Perceptron (MLP) dan arsitektur Long Short Term Memory (LSTM). Pengujian dilakukan dengan menggunakan data harga beras dari tanggal 1 Januari 2019 hingga 31 Juli 2019 sebagai data percobaan.

Dalam tahap forecasting, data 6 bulan terakhir harga beras digunakan sebagai *data testing* untuk menghasilkan output prediksi harga dari kedua model. Dari hasil perbandingan output, didapatkan kesimpulan bahwa model LSTM lebih akurat dibandingkan dengan MLP. Hal ini terlihat dari rentang dengan harga aktual dan nilai RMSE yang lebih kecil pada model LSTM.

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IBI RKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Nilai RMSE pada model LSTM yang didapatkan adalah sebesar 0.49 untuk training data score dan 0.27 untuk *test data score*. Selain itu, hasil grafik plot menunjukkan bahwa tidak ada perbedaan signifikan antara hasil train dan *testing* dengan data aktual. Oleh karena itu, dapat disimpulkan bahwa model LSTM lebih cocok digunakan dalam memprediksi harga beras pada PT. FoodStation.

#### 4. Perbandingan Algoritma Long Short-Term Memory dengan SVR pada Prediksi Harga Saham di Indonesia

Hasil penelitian yang dilakukan oleh Adhib Arfan, dan Lussiana ETP menunjukkan bahwa semakin besar rentang data yang digunakan oleh SVR, maka semakin besar nilai MSE yang dihasilkan. Sedangkan nilai MSE yang dihasilkan LSTM memiliki nilai yang sama walaupun menggunakan rentang data yang berbeda. Berdasarkan hasil pengujian, LSTM mampu memprediksi harga saham pada tahun 2017-2019 dengan performa yang baik, hal ini ditunjukkan dengan tingkat kesalahan yang kecil. Dengan demikian dapat disimpulkan bahwa LSTM mampu menanggulangi ketergantungan jangka panjang dan mampu memprediksi harga saham dengan hasil yang akurat.

#### 4. Perbandingan Perbandingan Algoritma Regresi Linier dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19

Hasil penelitian yang dilakukan oleh Syakirah Fachid, dan Agung Triayudi adalah model Regresi Linier menghasilkan tingkat akurasi sebesar 94% dengan RMSE sebesar 3031.127 dan MAPE sebesar 47.66. Sedangkan pada model Random Forest dengan menggunakan hyperparameter  $max\_depth = 10$  dan  $n\_estimators = 100$  dapat menghasilkan tingkat akurasi sebesar 97.7% dengan RMSE sebesar 1886.555 dan nilai MAPE sebesar 14.85. Dapat disimpulkan bahwa pada penelitian



model random forest lebih baik daripada regresi linier, karena tingkat akurasiya 97.7%. Terdapat perbedaan diantara kedua algoritma tersebut salah satunya pada nilai MAPE dari masing masing algoritma. Jika merujuk pada interpretasi MAPE bahwa model random forest dikategorikan 'baik' untuk digunakan dimasa yang akan datang sedangkan model regresi linier hanya di kategorikan 'layak' digunakan dimasa yang akan mendatang. Ini menunjukkan bahwa model Regresi Random Forest lebih baik daripada regresi linier.

## 5. Perbandingan Multi Time Steps Prediction dengan Recurrent Neural Network Long Short Term Memory

Hasil penelitian yang dilakukan oleh Ahmad Ashril Rizal, Siti Soraya adalah sistem prediksi yang dibangun untuk kunjungan wisatawan dengan Recurrent Neural Network Long Short Term Memory (RNN LSTM), bahwa training RNN LSTM dengan menggunakan tiga model memberikan hasil yang beragam. Dari ketiga model LST yang dilakukan yaitu LSTM regression, LSTM dengan sliding window dan LSTM dengan time steps tidak ada model yang memberikan hasil optimal dari sisi training dan *testing* sekaligus. Hasil terbaik pada proses training untuk prediksi kunjungan wisatawan diperoleh dengan menggunakan model regresi dengan RMSE 6529,42.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.