

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Gambaran Umum Objek Penelitian**

Objek penelitian berfokus pada prediksi kelanjutan adopsi WFH pasca COVID-19 berdasarkan preferensi karyawan pada beberapa divisi di GORP Kompas Gramedia Palmerah. Divisi yang akan dijadikan objek penelitian adalah divisi *Corporate Controller Division, Finance & Accounting Division, Merchandising and Distribution Division, dan System and IT Division.*

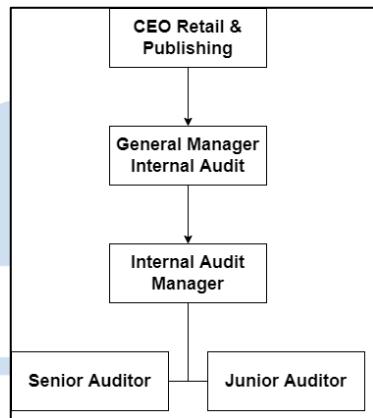
Objek tersebut diambil berdasarkan hasil wawancara narasumber yang mengeluhkan beberapa tantangan dan hambatan selama penerapan kerja WFH. Berikut merupakan penjelasan lebih lanjut mengenai tiap divisi yang akan dijadikan sebagai objek penelitian.

##### **3.1.1 Corporate Controller Division**

Divisi ini bertanggung jawab terhadap perencanaan kegiatan operasional, memastikan semua proses audit dipetakan secara tepat dan akurat, melakukan kontrol terhadap implementasi pengecekan & koreksi atas terjadinya kesalahan, hingga bertanggung jawab atas hasil audit internal dan mempersiapkan implementasi solusi terhadap hasil temuan [39].

*Corporate controller division* terdiri dari kurang lebih 25 karyawan termasuk *General Manager, Manager, Senior Auditor* dan *Junior Auditor* yang terbagi lagi menjadi dua *department* yaitu *Internal Audit 1* dan *Internal Audit 2*. *Internal Audit 1* dan *2* hanya dibedakan berdasarkan wilayah atau regional yang dinaungi. Gambar 3.1. merupakan struktur organisasi untuk *department Internal Audit 1* dan *Internal Audit 2*.

**UNIVERSITAS  
MULTIMEDIA  
NUSANTARA**

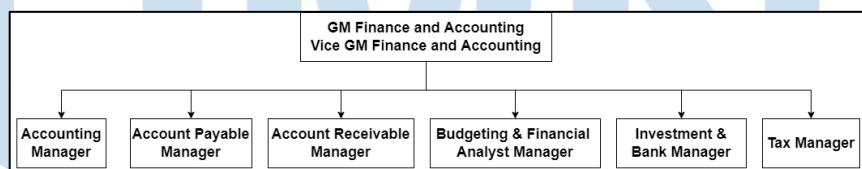


Gambar 3.1 Struktur Organisasi *Internal Audit* [39]

### 3.1.2 Finance and Accounting Division

*Finance and Accounting Division* mempunyai tugas yang berbeda, dimana *finance* bertugas untuk melakukan pengelolaan keuangan dan pengalokasian dana perusahaan, sedangkan *accounting* akan menerima catatan keuangan dari bagian *finance* dan menuliskannya dalam sebuah dokumen atau laporan keuangan sebagai bukti transaksi keuangan [39].

*Finance and Accounting division* terdiri dari kurang lebih 90 karyawan yang terbagi lagi menjadi beberapa *department* yaitu *Accounting Department*, *Account Payable (AP) Department*, *Account Receivable (AR) Department*, *Budgeting & Financial Analyst Department*, *Investment & Bank Department*, dan *Tax Department* seperti struktur organisasi pada Gambar 3.2.



Gambar 3.2 Struktur Organisasi Divisi *Finance & Accounting* [39]

### 3.1.3 Merchandising and Distribution Division

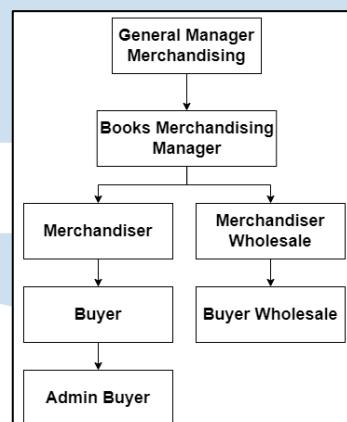
*Merchandising and Distribution Division* mempunyai tugas utama dalam proses pengadaan barang dan negosiasi dengan vendor [39]. *Merchandising and Distribution division* terdiri dari kurang lebih 120 karyawan yang terbagi lagi menjadi dua *department* yaitu *Books*

*Merchandising Department* dan *Non-Books Merchandising Department*.

Berikut merupakan penjelasan *jobdesc* dan susunan organisasi pada kedua *department* tersebut.

### 1. Books Merchandising Department

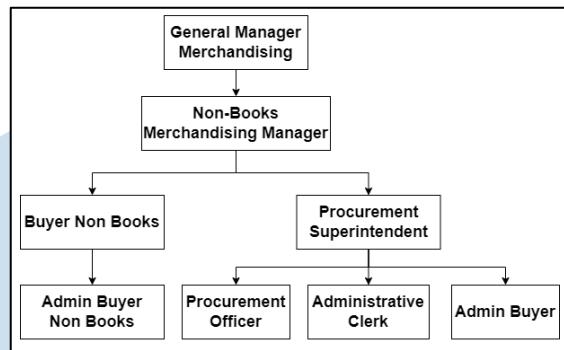
Books Merchandising Department bertanggung jawab dalam proses pengadaan *stock* buku seperti mengelola aktivitas pengadaan dan negosiasi dengan vendor, melakukan kerjasama dengan vendor termasuk untuk promosi bersama, hingga melakukan koordinasi dengan store tentang implementasi *visual merchandising* [39]. Berdasarkan Gambar 3.3 *books merchandising department* terbagi lagi menjadi beberapa *role* atau posisi.



Gambar 3.3 Struktur Organisasi *Books Merchandising Dept* [39]

### 2. Non-Books Merchandising Department

*Non-Books Merchandising Department* bertugas sama seperti *books merchandising department*, tetapi pada *department* ini, proses pengadaan barang hanya difokuskan pada barang non buku (alat tulis, alat musik, dll). Berdasarkan Gambar 3.4. *non-books merchandising department* terbagi lagi menjadi beberapa *role* atau posisi.



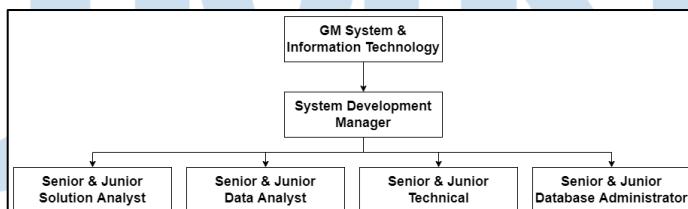
Gambar 3. 4 Struktur Organisasi *Non-Books Merchandising Dept* [39]

### 3.1.4 System and IT Division

*System and IT Division* bertugas dalam melakukan *development* sistem hingga melakukan *maintenance* dan *support (helpdesk)* [39]. *System and IT division* terdiri dari kurang lebih 60 karyawan yang terbagi lagi menjadi dua *department* yaitu *System Development Department* dan *IT Operation Department* [39]. Berikut merupakan penjelasan *jobdesc* dan susunan organisasi pada kedua *department* tersebut.

#### 1. System Developement Department

*System Development Department* tanggung jawab dengan pengembangan dan analisa sistem, seperti pengembangan database perusahaan, dan proses analisa sistem yang dilakukan oleh seorang *data analyst* [39]. Berdasarkan Gambar 3.5 *System Development department* terbagi lagi menjadi beberapa *role* atau posisi.

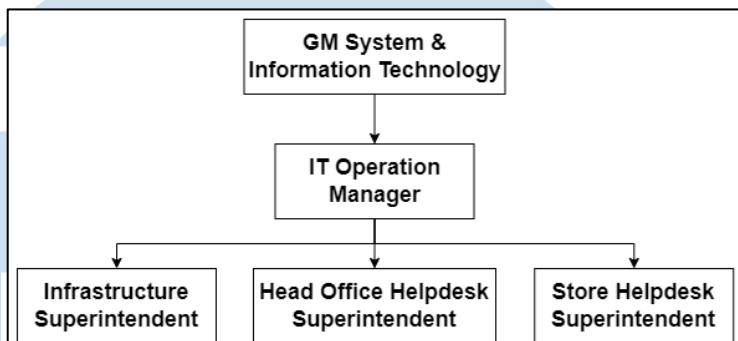


Gambar 3.5 Struktur Organisasi *System Development Dept* [39]

#### 2. IT Operation Department

*IT Operation Department* bertanggung jawab dalam menjaga keamanan infrastruktur seluruh data dari serangan, mengkoordinasi kegiatan pemasangan infrastruktur, mengkoordinasikan kegiatan help desk, melakukan evaluasi terhadap sistem yang digunakan, melakukan

kegiatan manajemen *client* agar kebutuhan dapat terpenuhi [39]. Berdasarkan Gambar 3.6 *IT Operation department* terbagi lagi menjadi beberapa *role* atau posisi.



Gambar 3.6 Struktur Organisasi *IT Operation dept* [39]

### 3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terbagi menjadi dua yaitu variabel independen dan variabel dependen. Berikut merupakan penjelasan tiap variabel yang digunakan.

#### 3.2.1 Variabel Independen

Variabel independen merupakan variabel bebas yang mempengaruhi atau menjadi sebab variabel dependen. Tabel 3.1 merupakan kelompok variable independen yang akan digunakan dalam penelitian ini.

Tabel 3.1 Variabel Independen

Section A: Employee Profile	
Variable	Reference
Gender	Internal Information Respondent
Ages	
Division	
Department	
Working Model	Emerging from COVID: Online work and its implications for the worker performance in K formance in Kosovo
Section B: Work Improvement	
Variabel	Reference
Motivation & Inovation Improvement	Continuance Adoption of Working from Home after the COVID-19 Outbreak: Empirical Evidence from Saudi Arabia
Maintaining Work Focus	Employee Productivity Modelling on a Work From Home Scenario During the COVID-19 Pandemic: A Case Study Using Classification Trees

Problem Solving Improvement	Continuance Adoption of Working from Home after the COVID-19 Outbreak: Empirical Evidence from Saudi Arabia”
Flexibility Improvement	
Work-Life Balance Improvement	
Time Management Improvement	
<b>Section C: Employee Productivity</b>	
Variabel	Reference
Work Hours during WFO	Additional Questions
Improvement Work Hours during WFH	
Number of Meetings in a Day	
Effort	
Productivity Improvement	Employee Productivity Modelling on a Work From Home Scenario During the COVID-19 Pandemic: A Case Study Using Classification Trees
<b>Section D: Employee Performance</b>	
Variabel	Reference
Quality of Project during WFH	Emerging from COVID: Online work and its implications for the worker performance in K formance in Kosovo
Quantity of Project during WFH	
Project Close on time during WFH	
KPI Improvement	
Performance rate during WFO	
Performance rate during WFH	
Job Satisfaction	
<b>Section E: Work Environment</b>	
Variabel	Reference
Have private work room	Emerging from COVID: Online work and its implications for the worker performance in K formance in Kosovo
Comfortable work environment	Employee Productivity Modelling on a Work From Home Scenario During the COVID-19 Pandemic: A Case Study Using Classification Trees
More distraction during WFH	
Have compatible equipment	
Office Equipment support	Employee Productivity Modelling on a Work From Home Scenario During the COVID-19 Pandemic: A Case Study Using Classification Trees
Internet Connection	
<b>Section D: Communication &amp; Support</b>	
Struggling in Communication	Emerging from COVID: Online work and its implications for the worker performance in K formance in Kosovo
Have a Clear Scope/Job Desc	

Get Supervision Support	Employee Productivity Modelling on a Work From Home Scenario During the COVID-19 Pandemic: A Case Study Using Classification Trees
-------------------------	--

### 3.2.2 Variabel Dependen

Variabel dependen merupakan variabel yang dijelaskan atau dipengaruhi oleh variabel independen. Pada Penelitian ini yang merupakan variabel dependen adalah “Preferensi karyawan mengenai kelanjutan adopsi WFH pasca COVID-19”. Tabel 3.2 akan menunjukkan variable dependen yang digunakan pada penelitian ini.

Tabel 3.2 Variabel Dependen

Variabel	Response
Preferensi karyawan	Ya = 1 Tidak = 0

## 3.3 Metode Penelitian

Metode penelitian yang akan digunakan dibagi lagi menjadi metode penyelesaian dan alur penelitian sebagai berikut

### 3.3.1 Metode Penyelesaian

Penelitian ini akan diselesaikan menggunakan teknik *data mining* dengan prediksi menggunakan metode *classification model*. Mengacu pada penelitian [5], terdapat tiga model *machine learning* yang akan digunakan pada penelitian ini yaitu *decision tree*, *random forest* dan *naïve bayes*. Hasil dari penelitian [4][5] menyebutkan bahwa model yang paling banyak dipakai dan mempunyai performa paling baik adalah *decision tree*, tetapi walaupun begitu Tabel 3.3 akan menunjukkan perbandingan ketiga model klasifikasi tersebut.

Tabel 3.3 Perbandingan *classification model* pada penelitian [22][30][32]

Algoritma	Kegunaan	Kelebihan	Kekurangan
Decision Tree	<ul style="list-style-type: none"> <li>Prediksi klasifikasi dengan model berbentuk <i>hierarchical</i></li> </ul>	<ul style="list-style-type: none"> <li>Mudah untuk menafsirkan <i>decision rules</i></li> <li>Mudah untuk menggabungkan</li> </ul>	Cenderung bekerja lebih baik jika terdapat beberapa attribut

	<ul style="list-style-type: none"> <li>• Menghasilkan <i>tree</i> dengan tiga tipe yaitu <i>internal nodes</i> (pertanyaan yang dicari), <i>nodes/branches</i> (pilihan beberapa alternatif) dan <i>leaf nodes</i> (klasifikasi hasil untuk mengambil keputusan)</li> </ul>	beberapa lapisan data numerik & kategorikal	yang sangat relevan tetapi akan memberikan hasil cukup buruk jika data cukup kompleks, maka dari itu perlu dilakukan <i>feature selection</i> untuk melihat <i>variabel</i> yang paling berpengaruh.
Random Forest	Model prediksi klasifikasi yang dikembangkan dari beberapa <i>decision tree</i> untuk menghasilkan prediksi rata-rata	<ul style="list-style-type: none"> <li>• Mampu memproses data yang besar dengan spasial lebih tinggi</li> <li>• Memberikan prediksi yang akurat tanpa limit <i>variabel</i></li> <li>• Dapat mengukur pentingnya fitur pada <i>training data</i></li> </ul>	Model ini cukup bias dalam mendukung variabel kategorikal yang cukup banyak.
Naïve Bayes	Model prediksi klasifikasi yang handal, cepat, akurat dan berfokus pada probabilitas bersyarat, serta dibuat tanpa estimasi parameter yang rumit	Hanya membutuhkan data <i>training</i> lebih sedikit untuk memperkirakan parameter klasifikasi yang diperlukan	Perlu mengetahui <i>prior probability</i> yang didasarkan pada asumsi dan prosesnya membutuhkan banyak pilihan subjektif antara model dan asumsi.  Solusi: dengan teknik <i>smoothing</i> .

Berdasarkan kegunaan, kelebihan serta kekurangan tiap model yang sudah dijelaskan pada Tabel 3.3, *decision tree* akan dipakai sebagai model utama pada penelitian ini untuk melakukan klasifikasi dan prediksi kelanjutan WFH di PT Kompas Gramedia Palmerah, walaupun begitu

penelitian ini tetap akan dilakukan perbandingan algoritma klasifikasi dengan model *random forest* dan *naïve bayes*.

### 3.3.2 Alur Penelitian

Alur penelitian yang digunakan pada penelitian ini diambil berdasarkan perbandingan tiga *framework data mining* yaitu KDD Process (*Knowledge Discovery in Databases*), SEMMA (*Sample, Explore, Modify, Model and Assess*) dan CRISP-DM (*Cross Industry Standard Process for Data Mining*). Tabel 3.4 akan menjelaskan perbandingan ketiga *framework* tersebut.

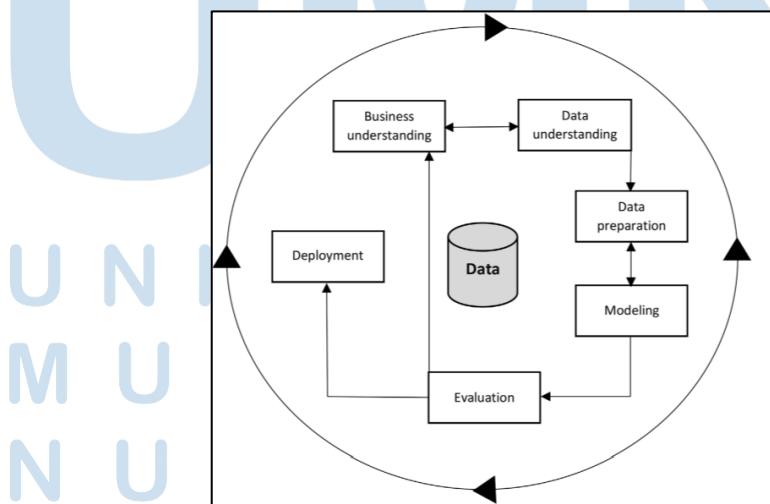
Tabel 3.4 Perbandingan *framework* KDD, SEMMA & CRISP-DM[40][41]

Kategori	KDD	SEMMA	CRISP-DM
First Development	Dasar metodologi data mining terstruktur pertama kali yang mendukung ekstraksi <i>knowledge</i> dengan data	Dikembangkan oleh SAS Institute tahun 2005 berbasis model KDD	Dikembangkan pada tahun 2000 sebagai tanggapan alternatif lebih baik dari model KDD.
Number of Steps	9	5	6
Name of Steps	<p><i>Developing &amp; Undestanding the Application</i></p> <p><i>Creating a Target Dataset</i></p> <p><i>Data Cleaning &amp; Pre Processing</i></p> <p><i>Data Transformation</i></p> <p><i>Choosing the Suitable Data Mining Task</i></p> <p><i>Choosing the Suitable Data Mining Algorithm</i></p> <p><i>Employing Data Mining Algorithm</i></p> <p><i>Interpreting Mined Patterns</i></p> <p><i>Using Discovered Knowledge</i></p>	<p>-----</p> <p><i>Sample</i></p> <p><i>Explore</i></p> <p><i>Modify</i></p> <p><i>Model</i></p> <p>-----</p> <p><i>Assessment</i></p> <p>-----</p> <p><i>Iterative, fokus pada data</i></p>	<p><i>Business Understanding</i></p> <p><i>Data Understanding</i></p> <p><i>Data Preparation</i></p> <p><i>Modeling</i></p> <p><i>Evaluation</i></p> <p><i>Deployment</i></p>
Advantages	<i>Iterative &amp; User-driven</i>	<i>Iterative, fokus pada data</i>	<i>Iterative dan stages dapat</i>

		<i>management, models &amp; supports</i>	diulang kembali, terbagi menjadi beberapa fase, <i>tasks</i> dan aktivitas
<i>Disadvantages</i>	Tidak menjelaskan tasks & aktivitas, meninggalkan GAP antara data <i>interpretation</i> dan <i>visualization</i> , tidak mempunyai repository informasi	Tidak melakukan <i>project analysis, design</i> dan <i>implementation</i> , tidak menjelaskan aktivitas yang dilakukan dan menghilangkan fase <i>evaluation</i> .	Tidak termasuk <i>monitoring, communication</i> dan <i>team management</i> , serta menghilangkan fase <i>maintenance</i> atau <i>updating models</i> .
<i>Industry</i>	<i>Industry &amp; Academic</i>	<i>Industry</i>	<i>Industry &amp; Academic</i>

Berdasarkan Tabel 3.4, terlihat bahwa *framework* paling *simple* adalah SEMMA, tetapi SEMMA melewatkkan dua process, sehingga pada penelitian ini akan digunakan alur penelitian dengan *framework data mining* CRISP-DM. Pemilihan CRISP-DM didukung berdasarkan penelitian [40] yang menyatakan bahwa hingga saat ini CRISP-DM banyak digunakan oleh penelitian dari sektor industri dan akademik sehingga sudah dianggap sebagai standar “de-facto” *framework data mining*. CRISP-DM juga masih menempati peringkat nomor satu berdasarkan *survey* yang dilakukan oleh organisasi *Data Science Project Management USA* di tahun 2020 [42].

Berikut merupakan *framework* CRISP-DM yang menggambarkan alur penelitian ini.



Gambar 3.7 Alur Penelitian dengan *Framework* CRISP-DM [24]

Berdasarkan Gambar 3.7, berikut penjelasan yang akan dilakukan pada penelitian ini:

### **1. Business Understanding**

Fase pertama pada *framework* CRISP-DM diawali dengan *business understanding* dimana difase ini difokuskan pada pemahaman tujuan bisnis dan *requirement* yang ada [24]. Penentuan tujuan merupakan salah satu aspek terpenting seperti jenis *data mining* yang akan digunakan serta kriteria keberhasilan apa yang akan dicapai [43]. Penelitian ini mempunyai tujuan utama untuk melakukan prediksi kelanjutan WFH pada PT Kompas Gramedia Palmerah dengan teknik data mining menggunakan perbandingan tiga algoritma klasifikasi yaitu *decision tree* sebagai model utama, *random forest* dan *naïve bayes*.

Hasil pemodelan yang akan didapatkan selanjutnya akan diprediksi dengan faktor-faktor pendukung keberhasilan WFH yang sudah ditentukan yaitu *work improvement*, *employee performance*, *employee productivity*, *work environment* dan *communication & support*. Keberhasilan penelitian ini ditandai dengan model algoritma klasifikasi mampu memberikan performance dengan hasil akurasi yang baik.

### **2. Data Understanding**

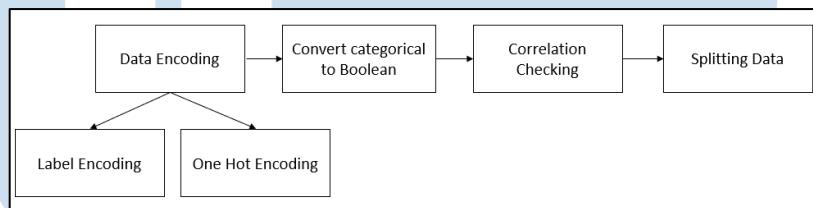
*Data Understanding* merupakan fase kedua pada *framework* CRISP-DM dimana fase ini akan dilakukan pengumpulan dataset yang akan digunakan, eksplorasi dan deskripsi data hingga memeriksa kualitas data sebelum diolah [43]. Pemahaman data sangat diperlukan agar dapat digunakan untuk mengungkap *knowledge* tersembunyi yang akan dicari [24]. Informasi awal untuk mendukung penelitian ini didapatkan dari hasil wawancara terhadap beberapa karyawan, selanjutnya proses pengambilan dataset dilakukan melalui survey kuesioner dengan target 200 responden pada divisi *Corporate Controller, Finance & Accounting, Merchandising & Distribution, System & IT* di GORP PT Kompas Gramedia Palmerah.

Dataset yang sudah diperoleh nantinya akan dikelompokkan menjadi lima faktor yang cukup berpengaruh pada kelanjutan WFH.

Pemahaman dataset akan diperjelas dengan proses *Exploratory Data Analysis* (EDA), dimana deskripsi dataset akan ditampilkan dengan visualisasi agar lebih mudah untuk dipahami.

### 3. Data Preparation

Fase ketiga adalah *data preparation* yang melibatkan proses untuk mempersiapkan data agar siap untuk diolah seperti *data reduction*, *cleansing* hingga *transformation*, karena kualitas data yang buruk akan sangat mempengaruhi hasil yang ingin dicapai [24]. Terdapat lima tahapan yang akan dilakukan pada fase *data preparation* berdasarkan Gambar 3.6.



Gambar 3.8 Alur Data Preparation

#### a. Data Encoding

Proses *encoding* data bertujuan untuk mengkonversi atau merubah *variabel* dengan tipe *categorical* menjadi bentuk *integer*. Adapun proses *encoding* data terbagi menjadi dua yaitu *label encoding* dan *one hot encoding*. Berikut merupakan variabel yang akan dilakukan encoding:

- *Label Encoding*: *age*, *number\_of\_meetings*, *workhours\_wfo*
- *OneHotEncoding*: *division*, *department*, *working\_model\_preference*

#### b. Convert Categorical to Boolean

Tahap selanjutnya adalah melakukan konversi untuk tipe data *categorical* yang hanya mempunyai dua option ('Ya', 'Tidak') menjadi bentuk boolean yaitu 1='Ya', dan 0='Tidak'. Variabel yang akan dikonversi ke dalam bentuk *boolean* adalah *gender*, *motivation\_inovation*, *work\_focus*, *time\_management*, *problem\_solving*, *flexibility*, *worklife\_balance*, *improvement\_workhours*, *effort*,

*extra\_time, productivity\_improvement, private\_workplace, comfortable\_workplace, distraction, suitable\_equipment, get\_asset, internet\_connection, dependence\_on\_internet, struggle\_communication, clear\_scope, supervision\_support, employee\_preference\_wfh*

c. *Correlation Checking*

Tahap ini akan dilakukan pengecekan korelasi antara features variabel, dan target variabel (*employee\_preference\_wfh*). Pengecekan korelasi akan dilakukan dengan membandingkan hasil untuk 10 *features* terbaik dan *features* dengan korelasi minimal sebesar 0,4. *Features* tersebut akan ditampung dalam *data frame* baru dan akan digunakan untuk proses *modeling classification*.

d. *Splitting Data*

Proses *splitting* data dilakukan dengan membagi data menjadi data *training* dan data *testing*. Ratio persentasi *splitting* data dilakukan sesuai dengan dataset, kebutuhan model dan tidak ada persentasi *splitting* yang optimal [44], tetapi terdapat dua masalah utama saat memutuskan *splitting* data yaitu jika *training* data lebih sedikit, model akan menunjukkan nilai *training* yang lebih besar, dan jika data pengujian/*testing* lebih sedikit, evaluasi model kinerja akan memiliki nilai yang lebih besar [44].

Berdasarkan teori tersebut, *splitting data* pada penelitian ini akan dilakukan pada dua *data frame* baru yang menampung 10 *attributes* dan 7 *attributes* terbaik dengan persentase perbandingan sebesar 70% *data training* dan 30% *data testing*. Proporsi tersebut diambil mengingat jumlah dataset yang tidak terlalu banyak.

#### 4. *Modeling*

Fase keempat merupakan tahapan modelling yang terdiri dari pemilihan teknik pemodelan berdasarkan masalah bisnis dan data yang akan digunakan [43]. Tahap pemodelan biasanya juga melibatkan model analitik yang biasa dilakukan dengan algoritma *machine learning* [24]. Penelitian ini akan dilakukan dengan mengembangkan dan membandingkan algoritma

machine learning yaitu *decision tree*, *random forest*, dan *naïve bayes*. Pemilihan algoritma ini didasarkan pada penelitian [4][5].

### 5. Evaluation

Fase kelima dalam *framework* CRISP-DM adalah *evaluation* dimana pada tahap ini akan dilakukan evaluasi dan peninjauan kembali dari hasil analisis dalam konteks kriteria keberhasilan tertentu [24]. Pada penelitian ini, tahap evaluasi model akan dilakukan dengan mengukur serta membandingkan *performance* algoritma klasifikasi. Evaluasi model akan dilakukan dengan melihat hasil nilai akurasi dengan 30% *testing data* menggunakan *confusion matrix* dan membandingkannya untuk melihat model mana yang paling cocok digunakan lalu dilanjutkan untuk memprediksi kelanjutan WFH dan melihat faktor atau variabel mana yang cukup mempengaruhi kelanjutan WFH.

### 6. Deployment

Fase terakhir pada *framework* CRISP-DM adalah *deployment* yang dapat berupa laporan akhir atau hasil dari *knowledge* baru yang diperoleh dari analisa data dan memudahkan *end-user* dalam memahami hasil sehingga dapat dilanjutkan dengan rekomendasi bisnis yang dapat ditindaklanjuti [24]. Tahap *deployment* pada penelitian ini akan dilakukan dengan implementasi model terbaik untuk menghasilkan *knowledge* berupa prediksi kelanjutan WFH pasca COVID-19 berdasarkan preferensi karyawan GORP Kompas Gramedia Palmerah yang akan dilengkapi dengan visualisasi *confusion matrix*, untuk melihat kesalahan prediksi yang terjadi.

#### 3.4 Teknik Pengumpulan Data

Data yang dipakai merupakan data primer yang diambil langsung dari hasil penyebaran kuesioner dan wawancara untuk mendukung latar belakang masalah penelitian. Berikut merupakan penjelasan untuk masing-masing teknik pengumpulan data yang dilakukan.

## 1. Wawancara

Wawancara bertujuan untuk mendukung latar belakang masalah penelitian dan dilakukan kepada beberapa *sample* karyawan sebagai perwakilan di *Group of Retail and Publishing* PT Kompas Gramedia Palmerah. Hasil yang didapatkan dari wawancara tersebut yaitu terdapat empat divisi yang cukup terhambat dalam menjalakan proses kerja selama penerapan WFH yaitu *corporate controller, finance & accounting, merchandising and distribution* dan *system & IT*.

## 2. Kuesioner

Dataset yang akan diolah diambil berdasarkan hasil dari survey kuesioner melalui *Google Form* dan akan disebarluaskan ke empat divisi pada GORP Kompas Gramedia Palmerah dengan target responden sebesar 200 karyawan. Pertanyaan yang diajukan pada survey kuesioner diambil berdasarkan penelitian terdahulu.

## 3.5 Tools

### 3.5.1 Programming Language

Penelitian ini akan dilakukan dengan bahasa pemrograman Python untuk melakukan *preparation data, modeling, evaluation, dan prediction*. Untuk mendukung proses *preparation, modeling, evaluation* dan *prediction*, Tabel 3.5 akan menampilkan beberapa *library* utama yang dipakai pada penelitian ini yaitu:

Tabel 3.5 *Library* utama pada penelitian

Library	Tujuan
Pandas	<i>Library</i> yang mendukung proses <i>import</i> , membaca file dengan berbagai format (.txt, .csv, .tsv) dan pengolahan data lainnya.
Numpy	<i>Library</i> yang mendukung proses operasi matematika tertentu.
Matplotlib	<i>Library</i> yang mendukung proses untuk visualisasi data.
Seaborn	<i>Library</i> peningkatan dari matplotlib yang digunakan juga untuk visualisasi data dan menentukan hubungan antar dua variabel.
Sklearn	<i>Library</i> yang mendukung berbagai tugas <i>machine learning</i> dalam proses pembuatan model.

### 3.5.2 Software

*Software* utama yang akan digunakan untuk pengolahan dan analisa data dipilih berdasarkan perbandingan *tools* yang sering digunakan untuk melakukan implementasi algoritma *machine learning* yaitu *Platform Python*, salah satunya *Jupyter Notebook* dan *RapidMiner*. Tabel 3.6 akan menunjukkan perbandingan kedua *software* tersebut.

Tabel 3.6 Perbandingan *Python Platforms* dan *RapidMiner* [36]

Comparison	Python Platform	RapidMiner
<i>Interface</i>	Pengguna membutuhkan <i>basic programming skills</i> dalam menggunakan Python karena diharuskan untuk melakukan <i>code editor</i> secara manual. Pengguna yang dapat memahami penggunaan dan mempunyai <i>basic programming skills</i> Python akan mempunyai lebih banyak kesempatan menggunakan <i>machine learning tools</i> .	Mempunyai <i>Graphical User Interface (GUI)</i> yang lebih baik, serta lebih mudah bagi pemula tanpa <i>basic programming skills</i> untuk menjalankan model dan mendapatkan hasilnya dengan mengklik dan memilih opsi yang sudah disediakan oleh RapidMiner.
<i>Algorithms</i>	Python memberikan lebih banyak fleksibilitas pada user untuk mengimplementasikan berbagai algoritma <i>machine learning</i> , serta dapat menemukan lebih dari seribu variasi algoritma.	Auto modelling yang diberikan RapidMiner membuat user hanya dapat memilih enam algoritma <i>machine learning</i> untuk masalah prediksi.
<i>Performance Metrics</i>	Untuk mendapatkan hasil <i>performance metrics</i> ,	Dapat dengan mudah menampilkan hasil

	diperlukan <i>programming statement</i> lebih banyak, sehingga <i>metrics</i> yang umum digunakan di Python adalah akurasi dan korelasi untuk kasus masalah prediksi.	<i>performance metrics</i> , karena sudah menyediakan <i>metrics</i> yang dapat langsung dipilih untuk evaluasi algoritma <i>machine learning</i> , salah satunya <i>performance metrics</i> berupa R Squared, RMSE, square error, relative error dan absolute error.
--	---	---

Berdasarkan Tabel 3.5, *software* utama yang akan digunakan pada penelitian ini adalah Python karena proses *code editor* pada Python terbilang lebih fleksibel dan cenderung mempunyai beragam variasi algoritma untuk melakukan prediksi. Berikut merupakan list software yang digunakan untuk menunjang penelitian ini yaitu:

1. Jupyter Notebook 6.0.3: digunakan sebagai *tools* utama untuk *code editor* dan pengolahan data menggunakan bahasa pemrograman *Python*.
2. Anaconda Navigator 1.9.12: digunakan untuk mengelola semua *library* yang akan dipakai pada saat pengolahan data.
3. Power BI Desktop 2.105.923.0: digunakan sebagai *tools* untuk melakukan eksplorasi data dengan visualisasi.