

## BAB II KAJIAN LITERATUR

### 2.1 *The State of The Art*

*State of the art* membahas penelitian sejenis yang telah dilakukan peneliti sebelumnya. berikut beberapa penelitian sejenis yang telah dilakukan sebelumnya:

- 1 Pada penelitian yang dilakukan oleh Muhammad Ryan Syah Erwin, Aji Gautama, dan Muhammad Agus Triawan yang berjudul deteksi hama ulat pada tanaman selada berbasis *aquaponic* menggunakan *Convolutional neural network* pada tahun 2021. Penelitian ini menggunakan *CNN* untuk mendeteksi ada atau tidaknya hama ulat pada tanaman selada melalui kamera *Raspberry Pi v1.3*. model *CNN* yang di gunakan pada penelitian ini adalah *VGG16*. Pengambilan dataset menggunakan kamera yang terintegrasi dengan kamera *Raspberry Pi*. hasil yang diperoleh dari penelitian ini adalah akurasi identifikasi sebesar 89% dan *cross entropy loss* sebesar 0,6421 [20].
- 2 Pada penelitian lainnya yang berjudul identifikasi penyakit tanaman tomat menggunakan algoritma *convolutional neural network* dan pendekatan *transfer learning* yang di lakukan oleh Maulana Muhammad Fathul pada tahun 2020. Penelitian ini berfokus pada perbandingan *CNN* dengan pendekatan *transfer learning*. Pada penelitian ini model yang digunakan adalah *VGG*, *ResNet* dan *DenseNet* dengan masing masing 2 variasi model data yang di gunakan berasal dari dataset *platvillage* dengan dimensi *input* 244, 244 pixel, memiliki 9 kelas penyakit dan 1 kelas sehat. Hasil dari penelitian ini adalah model dengan performa terbaik dalam melakukan klasifikasi di capai oleh model *ResNet-50* dengan *accuracy* 96.16% *precision* 97%, *recall* 96% *FI-score* 97% dan *AUC* 0,9792 [21].
- 3 Pada penelitian lain yang di lakukan oleh Costantino Geovany Orlando Lana pada tahun 2020 yang berjudul pengembangan aplikasi *mobile* untuk mengidentifikasi spesies tanaman obat menggunakan metode *convolutional neural network*. Peneliti melakukan identifikasi dengan model *CNN VGG19* dengan dataset yang berasal dari Merapi Farma Herbal Yogyakarta dengan 15 spesies, hasil yang di peroleh menghasilkan akurasi sebesar 99,04% untuk

*training* dan 97.52 untuk *validation*. Untuk implementasi menggunakan *web service* dan bantuan kamera untuk mendeteksi [22].

- 4 Pada penelitian lain yang berjudul Identifikasi Penyakit pada Citra Daun Kentang Menggunakan *Convolutional Neural Network (CNN)* yang dilakukan oleh Alang Mulya Lesmana, Ronna Putri Fadhillah, dan Chaerur Rozikin pada tahun 2022. Peneliti melakukan identifikasi penyakit pada citra daun kentang menggunakan CNN dengan 25 *steps per epoch* dan 4 *validation steps*. Nilai akurasi tertinggi data training mencapai 93% dan akurasi tertinggi data *validation* mencapai 99% [23].
- 5 Pada penelitian lain yang berjudul Implementasi *Convolutional Neural Network* Untuk Identifikasi Jenis Tanaman Melalui Daun yang dilakukan oleh Felix, Jeffry Wijaya, Stephen Putra Sutra, Pyter Wahyu Kosasih, dan Pahala Sirait. Pada penelitian ini peneliti membuat sebuah aplikasi yang berfungsi untuk mengklasifikasikan penyakit berdasarkan citra daun. pengujian dari proses deteksi dan pengenalan jenis daun menghasilkan akurasi sebesar 76% [18].
- 6 Kemudian pada penelitian yang *dipublish* oleh *IEEE*. berjudul *Detection of Lettuce Plant Conditions Based on Images using Backpropagation Method* yang dilakukan oleh Rachman, Hanizar, Purwanto, Yudha dan, Setianingsih. Penelitian ini melakukan deteksi kondisi tanaman selada menggunakan algoritma *CNN*. Data yang digunakan berjumlah 300 data dan memiliki 2 kelas yaitu *good* dan *bad*. kemudian penelitian ini menggunakan kamera mikro computer raspi sebagai input. Performan yang didapat adalah tingkat akurasi terbaik sebesar 99.2% dengan konfigurasi partisi data 90% untuk data latih dan 10% untuk data uji, *learning rate* sebesar 0,0001, jumlah *epochs* 250, dan nilai *batch size* 15 [24].
- 7 Penelitian lain yang *dipublish* oleh *IEEE* pada tahun 2017 yang berjudul *Quality Assessment of Lettuce using Artificial Neural Network*. Yang dilakukan oleh C Valuenza *et al.* Melakukan penelitian mengenai deteksi kualitas dari tanaman selada menggunakan algoritma *ANN*. Data dibagi menjadi 2 yaitu *good* dan *reject*. Hasil yang diperoleh dari dua kali *test* terhitung  $p\text{-value} = 0.487302734$  [25].

Berdasarkan penelitian sejenis yang sudah disebutkan sebelumnya kemudian akan disusun dalam bentuk tabel sebagai berikut.

Tabel 2. 1 Tabel *State of The Art*

No.	Peneliti	Tahun	Metode	Tujuan
1	Muhammad Ryan Syah Erwin, Aji Gautama, dan Muhammad Agus Triawan.	2021	<i>Convolutional Neural Network (CNN) arsitektur VGG.</i>	Mengidentifikasi hama ulat pada tanaman selada melalui kamera <i>Raspberry Pi</i> .
2	Maulana Muhammad Fathul.	2020	<i>Convolutional Neural Network Arsitektur VGG, ResNet, dan DenseNet</i>	Membandingkan hasil akurasi model transfer learning pada penyakit tanaman tomat.
3	Costantino Geovany Orlando Lana	2020	<i>Convolutional Neural Network (CNN) arsitektur VGG.</i>	Mengembangkan aplikasi mobile untuk identifikasi spesies tanaman obat
4	Alang Mulya Lesmana, Ronna Putri Fadhillah, dan Chaerur Rozikin	2022	<i>Convolutional Neural Network (CNN)</i>	Mengidentifikasi penyakit pada citra daun kentang.
5	Felix, Jeffry Wijaya, Stephen Putra Sutra, Pyter Wahyu Kosasih, dan Pahala Sirait	2020	<i>Convolutional Neural Network (CNN)</i>	mengidentifikasi jenis tanaman melalui bentuk daun.
6	Rachman, Hanizar, Purwanto, Yudha dan, Setianingsih,	2020	<i>Artificial Neural Network (ANN) Backpropagation</i>	Menidentifikasi kondisi tanaman selada berdasarkan daun.

7	C Valuenza	2020	<i>Artificial Neural Network (ANN)</i>	Mendeteksi kualitas tanaman selada.
8	Harry Akbar Fauzan	2023	<i>CNN</i>	Penyakit Selada

Berdasarkan tabel 2.1 topik penelitian *image recognition* dengan objek selada lain yaitu mengenai identifikasi kualitas selada [24], [25], identifikasi jenis selada menggunakan *CNN* [26], dan identifikasi abnormalitas selada menggunakan *SVM* [27]. Untuk model sendiri peneliti akan menggunakan model *CNN* arsitektur *DenseNet201* karena jarang digunakan dan kalah populer dibanding *VGG* atau *Resnet* [15]. Walaupun dalam beberapa penelitian *DenseNet* memperoleh nilai akurasi dibawah arsitektur lain namun peneliti tetap menggunakan *DenseNet201* dengan pertimbangan nilai akurasi yang tidak terlalu tertinggal jauh seperti pada penelitian yang dilakukan peng zhaoo dkk. Akurasi *DenseNet201* berada dibawah *Resnet* hanya terpaut 1.91% saja [28]. Kemudian pada penelitian lain yang dilakukan oleh M putra selisih akurasi yang diperoleh *DenseNet* adalah 3% dengan *ResNet* dan *MobileNet* [29].

Sehingga pada penelitian ini peneliti berfokus pada identifikasi penyakit pada tanaman selada dengan data gambar sebagai input yang mana belum ada penelitian yang berfokus pada subjek tersebut. Karena belum adanya penelitian yang berfokus pada identifikasi penyakit tanaman selada mengakibatkan tidak adanya data gambar penyakit selada yang dapat digunakan sehingga peneliti perlu mencari dan mengumpulkan data gambar secara mandiri berdasarkan ciri informasi yang ada pada penelitian lain. Proses identifikasi penyakit tanaman selada dilakukan dengan menggunakan algoritma *Convolutional Neural Network* arsitektur *DenseNet201*. Model terbaik yang sudah dibuat kemudian akan di-*deploy* kedalam *FastAPI*.

## 2.2 Studi Literatur

### 2.2.1 Selada

Selada (*Lactuca Sativa*) adalah tanaman yang termasuk dalam famili *Asteraceae*. Selada berasal dari Asia Barat yang kemudian menyebar di Asia dan

negara negara beriklim sedang. Negara yang mengembangkan selada diantaranya Jepang, Thailand, Taiwan, Amerika Serikat serta Indonesia [30]. Daerah yang membudidayakan selada di Indonesia saat ini masih terbatas diantaranya Cipanas, Lembang, dan Pengalengan Jawa Barat. Selada adalah tanaman sayuran yang biasanya dapat dimakan secara mentah, hal ini dikarenakan selada memiliki kandungan mineral yang cukup tinggi. Tanaman selada mengandung mineral, vitamin, antioksidan, potassium, zat besi, folat, karoten, vitamin C dan vitamin [31]. Selada sangat bermanfaat bagi tubuh karena dapat membantu pembentukan sel darah putih dan sel darah merah dalam susunan sum-sum tulang. kemun mengurangi resiko terjadinya kanker, tumor dan penyakit katarak, membantu kerja pencernaan dan organ-organ di sekitar hati serta menghilangkan gangguan anemia [21].

### **2.2.2 Penyakit Selada**

Beberapa pengertian penyakit menurut para ahli: menurut Thomas Timmereck penyakit adalah keadaan dimana terdapat gangguan terhadap bentuk dan fungsi tubuh sehingga berada dalam keadaan tidak normal [33]. menurut Rahmat Rukmana dan Sugandi Saputra penyakit tanaman adalah sesuatu yang menyimpang dari keadaan normal, cukup jelas menimbulkan gejala yang dapat dilihat, menurunkan kualitas atau nilai ekonomis, dan merupakan akibat interaksi yang cukup lama [34]. Kemudian menurut Agrios penyakit adalah kondisi dimana sel dan jaringan tanaman tidak berfungsi secara normal yang ditimbulkan karena gangguan secara terus menerus oleh agen pathogen atau faktor lingkungan dan akan menghasilkan perkembangan gejala [35]. Sehingga dapat disimpulkan bahwa penyakit adalah sebuah gangguan pada organisme hidup yang membuat sesuatu menjadi menyimpang dari keadaan normal yang disebabkan baik oleh patogen ataupun faktor lingkungan.

Penyebab penyakit dapat dibagi menjadi dua yang pertama biotik yaitu penyakit yang disebabkan oleh parasit dan yang kedua adalah non biotik atau non parasit [36]. Parasit atau patogen yang dapat menyerang tanaman di antaranya meliputi jamur, bakteri, nematode, virus, viroid, riketsia, fitoplasma, algae, protozoa, dan tumbuhan tingkat tinggi parasit [37].

Beberapa penyakit yang sering menyerang tanaman selada diantaranya: Penyakit hawar daun, hawar daun alternaria atau penyakit bercak kering atau hawar daun alternia yang disebabkan oleh jamur *Alternaria SP.* Gejala yang Nampak pada serangan daun ini adalah terlihat ada bercak-bercak coklat tua sampai hampir hitam. Bercak berbentuk hitam berukuran kecil namun jika dibiarkan akan membesar dan bergabung menjadi satu serangan dimulai dari daun bawah kemudian naik keatas. Terkadang daunnya akan berlubang karena bercak bercak itu mengering lalu jatuh, dan daun menggulung kriting. Gambar 2. Menampilkan contoh dari penyakit hawar daun yang menyerang tanaman selada.



Gambar 2. 1 Penyakit hawar daun pada tanaman selada

Penyakit yang kedua adalah penyakit busuk daun (*downy mildew*) disebabkan oleh *bremia lactuae*. Gejala yang ditimbulkan tulang-tulang daun terjadi bercak bersudut, bewarna hijau muda pucat sampai kuning. Pada permukaan bawah daun dapat berbentuk kapang (jamur) putih. Bagian daun yang terinfeksi saling berhubungan dan berubah menjadi bercak coklat yang besar [10]. Gambar 2.2 Menampilkan contoh dari pernyakit busuk daun yang menyerang tanaman selada.



Gambar 2.2 Penyakit busuk daun pada tanaman selada

### 2.2.3 *Image processing*

Identifikasi adalah proses pengenalan, menempatkan objek atau individu dalam suatu kelas sesuai dengan karakteristik tertentu. Identifikasi berasal dari kata Inggris *identify* yang artinya mengenali, mengategorikan, atau menentukan. Ada juga yang mengartikan bahwa identifikasi adalah penempatan atau penentu identitas seorang atau benda [38]. Sehingga identifikasi dapat juga diartikan sebagai kegiatan yang bertujuan untuk memeriksa dan menganalisa lebih mendalam akan suatu hal, suatu proses, ataupun benda.

Pengolahan citra digital atau (*Digital Image Processing*) merupakan bidang ilmu yang mempelajari tentang bagaimana suatu citra itu dibentuk, diolah, dan dianalisis sehingga menghasilkan informasi yang dapat dipahami oleh manusia. Citra dibagi menjadi 2 jenis yaitu citra analog dan citra digital. Citra analog adalah citra yang dibentuk dari sinyal analog yang bersifat kontinu, sedangkan citra digital adalah citra yang dibentuk dari sinyal digital yang bersifat diskrit. Citra analog dihasilkan dari alat akuisisi citra analog, contohnya adalah mata manusia dan kamera analog. Gambaran yang tertangkap oleh mata manusia dan foto atau film yang tertangkap oleh kamera analog merupakan contoh dari citra analog [39].

### 2.2.4 *Machine learning*

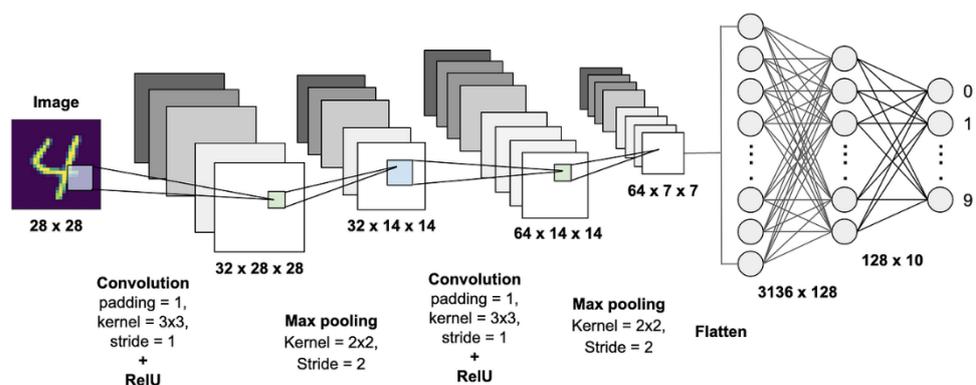
*Machine learning* adalah sebuah studi algoritma yang mempelajari bagaimana cara melakukan tugas tertentu secara otomatis oleh *system* itu sendiri. *Machine learning* berkerja dengan menggunakan berbagai kegiatan yang dilakukan sebelumnya sebagai sumber dan dimanfaatkan dalam perintah yang diberikan berikutnya. Tujuan *machine learning* adalah dapat mengeksekusi *system* pembelajaran mandiri dengan masukan sebelumnya [40].

*Machine learning* terbagi menjadi tiga kategori yaitu *Supervised Learning*, *Unsupervised Learning*, *Reinforcement Learning*. *Supervised learning* adalah sebuah metode *machine learning* yang mana kumpulan data sepenuhnya diberikan label. Contoh beberapa algoritma *supervised learning* adalah *Naïve Bayesian*, Metode *Rocchio*, *Decision Tree*, *k-Nearest Neighbor*, *Neural Network* [41]. Lalu *unsupervised learning* atau sering disebut *clustering* adalah metode *machine learning* yang menggunakan kumpulan data yang tidak diberikan label secara

mandiri. Metode *unsupervised learning* akan melakukan pembagian kelas dengan otomatis. Metode *clustering* terbagi menjadi 2 metode yaitu *hierarchical* dan *non-hierarchical*. *Hierarchical Agglomerative Clustering* adalah metode pengelompokan yang menggunakan kesamaan objek terdekat kemudian dilanjutkan pada objek lain yang terdekat yang kemudian membentuk hierarki dari paling mirip hingga tidak mirip [42]. kemudian *non-hierarchical* adalah metode pengelompokan yang melibatkan pembentukan *cluster* baru dengan menggabungkan atau memisahkan *cluster*. *Reinforcement learning* adalah metode yang bekerja dengan *system trial and error*. *Reinforcement learning* berinteraksi dan mengamati konsekuensi atas tindakan sebagai tanggapan atas hasil yang diterima [43].

### 2.2.5 Convolutional Neural Network (CNN)

*Convolutional Neural Network* (CNN) adalah jenis *neural network* yang biasa digunakan dalam pemrosesan *machine learning* dengan menggunakan data gambar. CNN terdiri dari neuron yang memiliki *weight*, *bias* dan *activation function*. *Convolutional layer* juga terdiri dari *neuron* yang tersusun sedemikian rupa sehingga membentuk sebuah filter dengan panjang dan tinggi. Cara kerja *CNN* adalah: pertama gambar akan dipecah menjadi bagian-bagian kecil dan tumpang tindih. Kemudian memasukan setiap gambar yang lebih kecil ke dalam *small neural network*. Kemudian menyimpan masing-masing gambar kecil ke dalam *array* baru. Proses selanjutnya yaitu *downsampling* atau mengecilkan ukuran *array* biasanya menggunakan *maxpooling*. Kemudian pada *layer* berikutnya melakukan prediksi. Ilustrasi jaringan CNN digambarkan pada gambar 2.1.



Gambar 2.3 Ilustrasi jaringan CNN [14]

Secara garis besar CNN terdiri dari dua proses yaitu *feature extraction layer* dan *fully connected layer*. Pada proses *feature extraction layer* gambar akan di ekstrak sehingga diperoleh fitur berupa angka angka yang merepresentasikan gambar tersebut. Pada layer ini akan dilakukan *convolutional layer*, *Rectified Linear Unit (ReLU) layer*, dan *pooling layer*. *Feature* yang telah dihasilkan dari proses sebelumnya kemudian akan diproses oleh *fully connected layer*. Proses ini berfungsi untuk proses klasifikasi gambar. Pada *fully connected layer* memiliki cara kerja yang sama dengan MLP dan memiliki bagian yang sama yaitu beberapa *input layer*, *hidden layer*, *activation function*, dan *output layer*. Namun sebelum ke proses kedua ini hasil pada proses pertama perlu dilakukan *flatten* dan *reshape* karena sebelumnya *feature map* hasil dari *feature extraction layer* masih berbentuk *multidimensional array* [17].

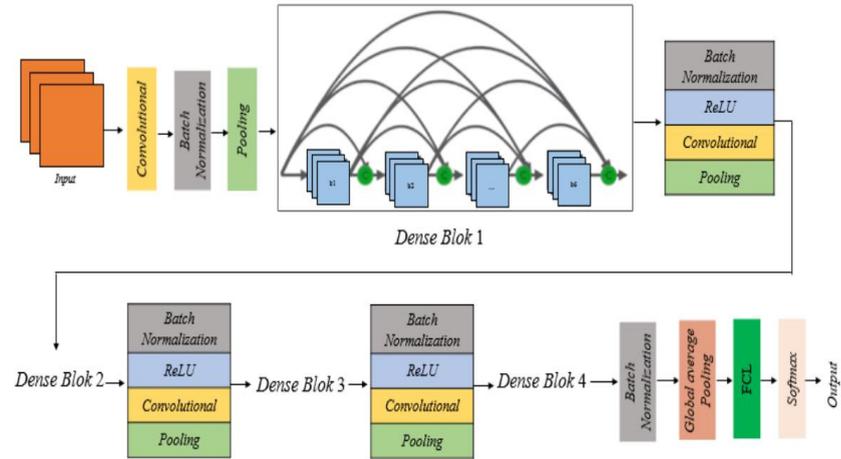
### 2.2.6 Transfer Learning

*Transfer learning* adalah penggunaan model yang sudah dilatih sebelumnya pada masalah yang berbeda. *Transfer learning* bukan termasuk algoritma *machine learning* melainkan sebuah struktur atau desain pada suatu algoritma tertentu yang mana pada penelitian ini algoritma CNN. *Transfer learning* merupakan sebuah teknik machine learning yang mana memanfaatkan model yang sudah dilatih sebelumnya (*pretrained model*) dan digunakan untuk mengklasifikasikan dataset yang baru. Walaupun model sudah ada dan sudah dapat dipakai namun tetap harus ada penyesuaian agar model *transfer learning* dapat digunakan pada penelitian yang akan dilakukan [44].

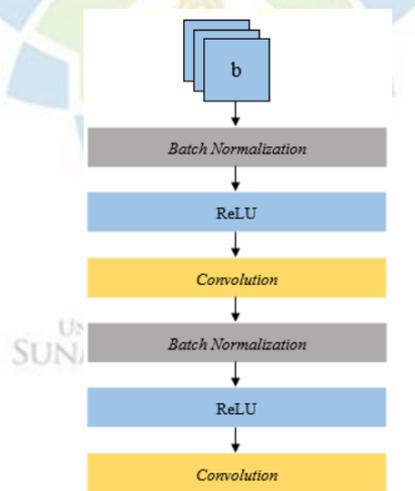
### 2.2.7 DenseNet

*Densely Connected Convolutional Networks (DenseNet)* merupakan salah satu model arsitektur *Deep Learning* yang menghubungkan setiap *layer* beserta *feature-maps* ke seluruh *layer* berikutnya, sehingga *layer* berikutnya menerima *input feature-maps* dari seluruh *layer* sebelumnya. *Densenet* memiliki arsitektur yang mana setiap *layer* ke *layer* lain terhubung secara *feed-forward*. Sehingga fitur

yang didapat lebih beragam dan kompleks. Ilustrasi jaringan *DenseNet* ditunjukkan oleh gambar 2.2.



Gambar 2.4 Ilustrasi jaringan *DenseNet* [45]



Gambar 2.5 Ilustrasi jaringan tiap blok [45]

Pada sebuah *dense blok* terdapat beberapa *blok* yang didalamnya terdiri dari beberapa *BN*, *ReLU*, dan *Convolutional* seperti yang diilustrasikan pada gambar 2.4. Didalam *blok* terdapat *concatenation layer* sebagai penghubung antara satu *blok* dengan *blok* lainnya. Kemudian penghubung antara *DenseBlok* terdapat *transition layer* yang terdiri dari *Convolutional layer* dan *pooling layer*. Kemudian diakhiri oleh *average pooling* dan *softmax* [45].

### 2.2.8 *Confusion matrix*

*Confusion matrix* adalah sebuah pengukuran yang biasa digunakan dalam proses *machine learning*. *Confusion matrix* adalah tabel yang berisi 4 kombinasi dari nilai prediksi dan nilai actual. 4 kombinasi yang ada pada *confusion matrix* diantaranya *True Positive* dimana prediksi bernilai positif dan itu benar, *True Negative* dimana prediksi bernilai *negative* dan itu benar, *False Positive* dimana bernilai positif dan itu salah, dan *False Negatif* dimana bernilai negatif dan itu salah. Dengan Informasi yang diperoleh dari 4 kombinasi *confusion matrix*, salah satu informasi yang dapat diperoleh adalah akurasi. Akurasi merupakan nilai yang menggambarkan seberapa akurat model dalam mengklasifikasikan nilai yang benar. Nilai akurasi diperoleh dengan persamaan:

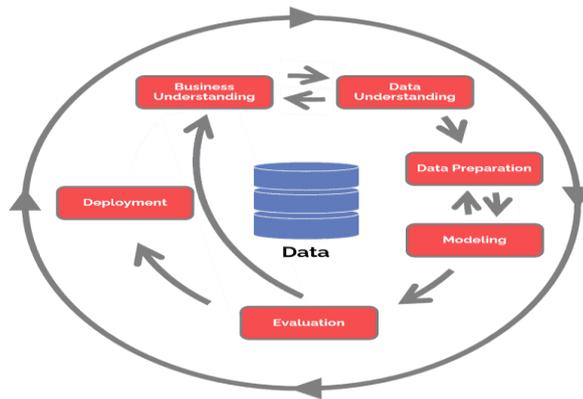
$$Accuracy = \left( \frac{(TP+TN)}{(TP+FP+FN+TN)} \right) \quad (2.1)$$

Nilai akurasi diperoleh dengan menjumlahkan semua data benar atau *true positive* dengan *true negative* kemudian dibagi dengan jumlah seluruh percobaan atau *true positive, false positive, false negative, dan true negative*.

### 2.2.9 *CRISP-DM*

Model *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) merupakan salah satu model yang sering digunakan dalam proses *data mining* yang dibuat pada tahun 1996 oleh 5 perusahaan yaitu Teradata, *Integral Solutions Ltd* (ISL), *Daimler AG*, *OHRA*, dan *NCR Corporation* yang kemudian berkembang hingga menjadi *methodology standard non-proprietary* dalam *data mining*.

*CRISP-DM* saat ini masih banyak digunakan sebagai metodologi data mining baik untuk tujuan akademi maupun industri. Karena memiliki keunggulan dalam menyelesaikan banyak masalah pada kasus kasus *data mining* [46].



Gambar 2.6 CRISP-DM Diagram [47]

Model *CRISP-DM* terdiri dari 6 tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Masing masing tahapan dijelaskan sebagai berikut:

### 1. *Business Understanding*

Tahapan awal dimana dilakukannya analisis kebutuhan serta tujuan berdasarkan sudut pandang bisnis, kemudian mengartikan pengetahuan kedalam bentuk pendefinisian masalah pada data dan kemudian menentukan rencana dan strategi untuk mencapai tujuan menggunakan data yang telah didefinisikan menganalisis. Pada tahapan *business understanding* terdapat 4 hal yang perlu dilakukan. Yang pertama adalah menentukan objektif bisnis. Pada tahap menentukan objektif bisnis hal hal yang harus diperhatikan diantaranya background, tujuan utama, dan kriteria keberhasilan. Yang kedua adalah menentukan tujuan data. Pada tahapan ini hal yang wajib ditentukan yaitu menentukan definisi data yang akan diambil. Yang ketiga adalah *asses situation*. Pada tahapan ini kita perlu menuliskan apa saja keperluan keperluan yang diperlukan untuk penelitian yang akan dilaksanakan. Langkah terakhir adalah *plan activities* yaitu membuat *planning* penelitian dan menentukan *tools* yang akan digunakan.

### 2. *Data Understanding*

Tahapan selanjutnya adalah mengumpulkan dan mendeskripsikan data kemudian memastikan kualitas data agar sesuai dengan penelitian yang dilakukan

agar tidak ada potensi masalah sehingga dapat mencegah penyimpangan tak terduga. Ketika melakukan *modeling* nanti. Beberapa masalah yang ada adalah *outliner* dan *null value*, pada tahapan ini terdapat 3 tahapan yang dilakukan. Pertama mengumpulkan data awal, yang kedua *describe data*, dan terakhir *explore data*.

### 3. *Data Preparation*

Tahapan ini secara garis besar adalah memperbaiki dan menyiapkan data. Pada tahapan Kumpulan data yang telah diperoleh dan akan digunakan akan dipersiapkan agar sesuai dengan model. Beberapa hal yang akan dilakukan mencakup *filtering* data, pemilihan data, serta melakukan transformasi data sesuai kebutuhan penelitian. Pada *data preparation* terdapat 3 tahapan yang sering dilakukan yaitu *data selection*, *data preprocessing* dan *data transformation* [48].

### 4. *Modeling*

Tahapan ini merupakan tahapan dimana algoritma *machine learning* atau metode statistika dijalankan. Pada tahap ini peneliti dapat kembali ke proses sebelumnya jika dataset yang digunakan tidak sesuai atau kurang. Pada tahapan *modeling* terdapat 3 tahapan yaitu menentukan teknik *modeling*, membuat *test design*, dan membuat model.

### 5. *Evaluation*

Tahap ini dilakukan dengan melihat tingkat performa dari pola yang dihasilkan oleh algoritma. Parameter yang akan digunakan dalam proses evaluasi komparasi adalah *confusion matrix* dengan aturan nilai akurasi. Pada tahapan *evaluation* terdapat 2 hal yang dilakukan yaitu *evaluate result* dan menentukan langkah selanjutnya.

### 6. *Deployment*

Tahapan *deployment* merupakan tahapan terakhir Dari metode *CRISP-DM*. Tahapan ini dilakukan dengan tujuan men-*deploy* model *machine learning* ke dalam bentuk aplikasi kemudian dilakukan penyebaran contohnya dengan membuat laporan. Tahapan *deployment* terbagi menjadi 2 yaitu *planning* dan *review* hasil [48].