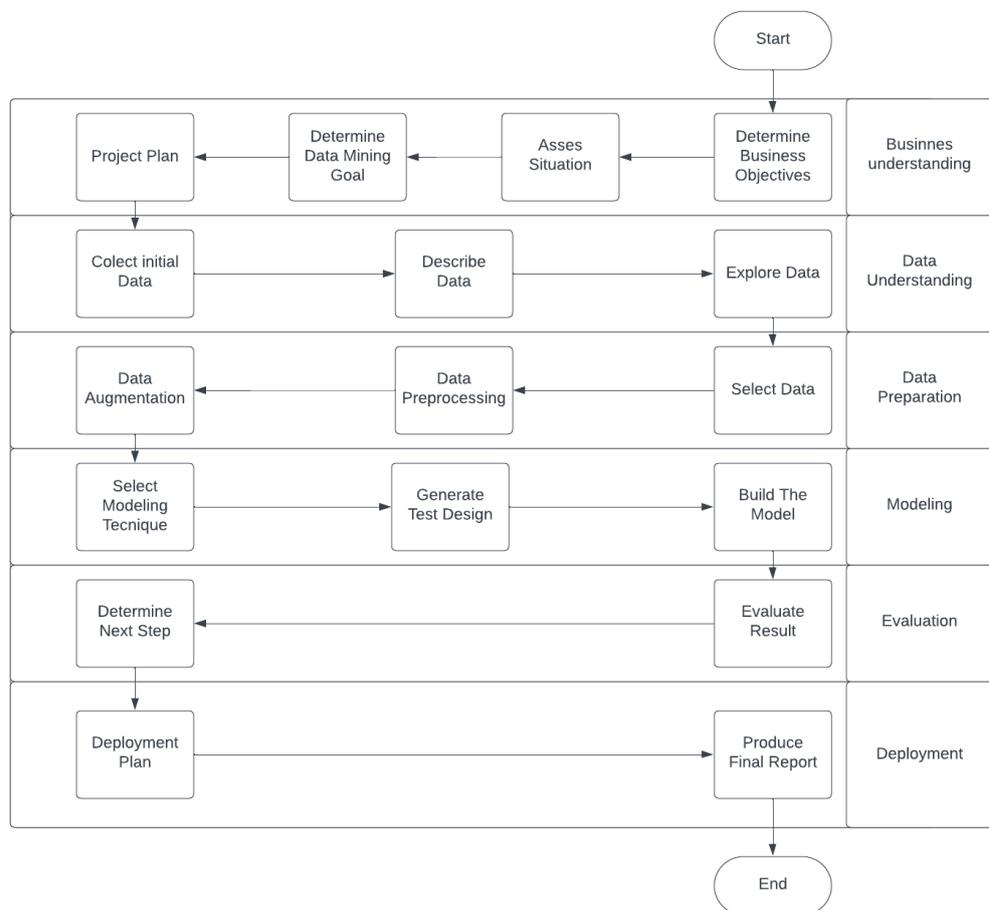


BAB III METODOLOGI PENELITIAN

Pada bab ini membahas mengenai bagaimana peneliti menyelesaikan masalah serta langkah langkah apa saja yang diambil dalam proses penyelesaian masalah dalam penelitian yang akan dilakukan. Adapun pada penelitian kali ini teknik penyelesaian masalah yang digunakan adalah menggunakan metode *CRISP-DM*. Pada metode ini terdapat enam langkah diantaranya *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Adapun agar lebih jelas, tahapan tahapan tersebut akan disajikan pada gambar 3.1.



Gambar 3.1 *Flowchart* tahapan penelitian

Metode *CRISP-DM* digunakan karena memang sudah sering dipakai pada penelitian penelitian yang melibatkan *data mining* dan *machine learning* [33]. Oleh karena itu pada penelitian kali ini metode *CRISP-DM* akan digunakan dalam

mengatasi permasalahan pada penelitian ini. Adapun penjelasan lebih lanjut mengenai langkah-langkah yang akan dilakukan dijelaskan pada berikut ini:

3.1 *Business Understanding*

Tahapan awal dimana dilakukannya analisis kebutuhan serta tujuan, kemudian mengartikan pengetahuan kedalam bentuk pendefinisian masalah pada data dan kemudian menentukan rencana dan strategi untuk mencapai tujuan menggunakan data yang telah di definisikan. Dalam proses ini juga dilakukan pengumpulan informasi dari berbagai sumber literatur seperti *Google Scholar* baik nasional maupun internasional yang berhubungan dengan penelitian yang dilakukan baik berhubungan dengan *CNN* sebagai algoritma yang digunakan atau penyakit tanaman selada sebagai objek masalah yang akan diselesaikan.

3.1.1 *Determine Business Objective*

Pada tahapan ini akan dijelaskan mengenai tujuan dari penelitian yang akan dilakukan. Hal ini dilakukan dengan mulai mengumpulkan informasi mengenai keadaan dari subjek yang akan diteliti kemudian menentukan tujuan, manfaat, serta peluang yang akan didapatkan dari penelitian yang akan dilakukan. Proses ini akan menghasilkan informasi manfaat dari pembuatan aplikasi identifikasi penyakit pada tanaman selada.

3.1.2 *Asses Situation*

Pada tahapan ini akan melanjutkan salah satu proses sebelumnya yaitu mengumpulkan informasi mengenai keadaan dari subjek yang diteliti secara lebih rinci. Kemudian informasi yang diperoleh akan dinilai dan menghasilkan beberapa pertimbangan baru yang akan membantu dalam menentukan bagaimana penelitian akan dilakukan. Proses ini akan menghasilkan informasi tambahan seperti sumber daya, kebutuhan, dan keterbatasan yang dimiliki penelitian ini.

3.1.3 *Determine Data Mining Goal*

Pada tahapan ini akan memutuskan rencana apa dan bagaimana *data mining* dilakukan serta menentukan tujuan dari *data mining* yang akan dilakukan. Hal ini dilakukan agar dapat mengoptimalkan *resource* ketika mengumpulkan data. Hasil

yang diperoleh dari tahap ini adalah tujuan data yang akan diperoleh beserta informasi tambahan seperti jenis data, sumber data, dan teknik pengambilan data.

3.1.4 Project Plan

Pada tahapan ini merupakan tahapan akhir dari *business understanding*. Pada tahapan *project plan* akan menjelaskan rencana proyek atau penelitian ini dengan menggunakan pertimbangan informasi yang sudah diperoleh pada langkah-langkah sebelumnya. Hasil yang diperoleh pada tahapan ini adalah rencana penelitian yang akan dilakukan.

3.2 Data understanding

Tahapan selanjutnya adalah mengumpulkan dan mendeskripsikan data kemudian memastikan kualitas data agar sesuai dengan penelitian yang dilakukan. Data yang digunakan pada penelitian ini adalah data yang diambil secara langsung dikarenakan data belum tersedia sebelumnya. Data yang digunakan berupa gambar atau citra tanaman selada sehat dan selada yang terkena penyakit berformat jpg. Adapun sampel dari data yang diperoleh ditampilkan pada gambar 3.2.



Gambar 3.2 Sampel data hawar, sehat, dan busuk

3.2.1 Collect Initial Data

Tahapan *collect initial data* merupakan tahapan dimana data yang dibutuhkan pada penelitian mulai dikumpulkan. Data diambil menggunakan kamera *handphone* Xiaomi Redmi 9C dengan pencahayaan normal. Data dibagi menjadi 3 label dengan target total 1.500 citra gambar, yang terdiri dari 500 selada sehat, 500 selada busuk daun, dan 500 selada hawar daun. semua data yang dibutuhkan akan dikumpulkan sesuai dengan kebutuhan penelitian. Hasil dari tahapan ini adalah sekumpulan data mentah yang belum diproses.

3.2.2 Describe Data

Tahapan *Describe Data* merupakan tahapan dimana semua data yang sudah terkumpul akan dijelaskan dalam bentuk informasi agar lebih mudah memahami apa yang ada pada data yang sudah dikumpulkan. kemudian data dibagi kedalam beberapa kelas sesuai dengan informasi yang sudah didapatkan. Hasil dari tahapan ini adalah data *image* penyakit selada yang sudah dibagi kedalam beberapa kelas.

3.2.3 Explore Data

Tahapan *explore Data* merupakan proses dimana keseluruhan data akan diperiksa dan *diexplore* untuk memperoleh informasi data secara keseluruhan. Seperti jumlah data perkelas, ukuran, format, dan informasi lain mengenai data.

3.3 Data Preparation.

Pada tahapan ini data di persiapkan. Kumpulan data yang telah diperoleh dan akan digunakan akan persiapkan agar sesuai dengan model. Beberapa hal yang akan dilakukan mencakup *data selection*, *data preproceasing*, dan *data augmentation*. Masing masing tahapan *data preparation* akan dijelaskan sebagai berikut:

3.3.1 Data Selection

Pada tahapan ini data akan dicek dan diseleksi. Data yang digunakan merupakan data yang harus berhubungan dengan penelitian serta memiliki kualitas yang jelas dan baik. Serta memastikan tidak ada gambar diluar subjek penelitian yang masuk kedalam data yang akan digunakan kedalam model.

3.3.2 Data Preprocessing

Pada tahapan *data preprocessing* data akan disiapkan agar sesuai dengan model. Ada beberapa proses yang dilakukan pada data langkah *preprocessing* ini. Yang pertama melakukan merubah nama. Kemudian *import data*, lalu *split data*, dan terakhir konfigurasi tambahan pada data. Pembagian data dilakukan untuk membagi data kedalam dua jenis data yaitu *data train* dan *data validation*. Data dibagi kedalam dua kategori secara acak. *Data train* adalah data yang akan dilatih pada pemodelan sedangkan *data validation* merupakan data akan menjadi bahan evaluasi hasil dari pemodelan. Lalu dilakukan juga konfigurasi tambahan seperti mengubah *class mode* menjadi kateorogikal mengatur *batch size* dan mengatur *seed*. Hasil dari tahapan ini adalah data yang sudah terbagi menjadi data *train* dan *test* juga sudah melalui pengaturan yang sudah ditentukan sebelumnya.

3.3.3 Data Augmentation

Data augmentation merupakan teknik yang dibuat untuk meningkatkan jumlah data dengan cara memanipulasi data yang sudah ada sebelumnya. Teknik augmentasi dilakukan ketika data dinilai kurang saat akan melakukan *modeling*. pada penelitian kali ini ada 3 metode augmentasi yang akan dilakukan yaitu:

1. *Random Flip*

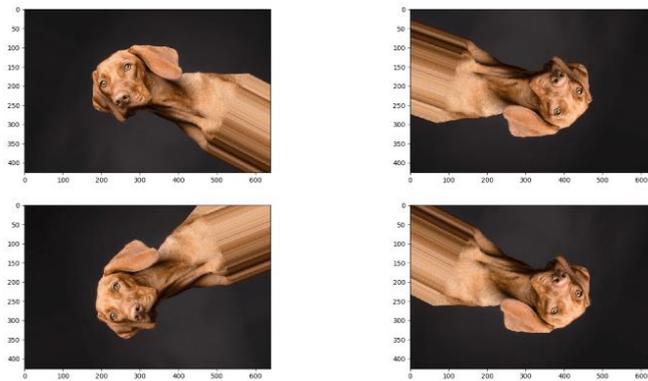
Random flip adalah sebuah metode augmentasi yang secara acak memutar gambar selama *training*. Contoh dari *random flip* diilustrasikan pada gambar 3.3



Gambar 3.3 Contoh *random flip* [49]

2. *Random Rotation*

Random rotation adalah sebuah metode augmentasi yang secara acak memutar gambar selama *training*. Contoh dari *random rotation* Diilustrasikan pada gambar 3.4



Gambar 3.4 Contoh *random rotation* [49]

3. *Random Zoom*

Random Zoom merupakan sebuah metode augmentasi yang mana secara acak melakukan *zoom out* atau *zoom in* pada gambar selama *training*. Ukuran yang diperoleh pada gambar yang diaugmentasi akan tetap sama seperti sebelum augmentasi. Contoh dari *random zoom* diilustrasikan pada gambar 3.5.



Gambar 3.5 Contoh *random zoom* [49]

4. *Image Normalization*

Data normalization adalah sebuah proses mengganti *range of pixel* pada gambar. Tujuannya adalah untuk memperjelas gambar. Pada penelitian kali ini *image normalization* digunakan dengan menggunakan fungsi *rescale*. Dengan menggunakan fungsi *rescale* $1./255$ *pixel* pada data gambar akan dikonversi dari rentang $[0,255]$ menjadi rentang $[0,1]$. Dengan mengskalakan setiap gambar menjadi $[0,1]$ akan membuat gambar berkontribusi lebih merata terhadap total loss. Tanpa fungsi *rescale* akan menyebabkan perbedaan bobot tiap gambar menjadi

lebih rumit karena *pixel* pada gambar yang terang akan lebih tinggi daripada gambar gelap.

3.4 Modeling

Setelah semua tahapan sebelumnya selesai maka akan masuk pada tahapan selanjutnya. Pada tahapan ini secara langsung melibatkan *machine learning* untuk menentukan model yang di buat. Kemudian membuat *design model* dengan berbagai variasi yang sudah ditentukan, dan melakukan evaluasi hasil secara singkat sebelum masuk ke tahap *evaluation* yang mana akan melakukan evaluasi yang lebih mendalam. Pada tahap ini peneliti dapat kembali ke proses sebelumnya jika dataset yang digunakan tidak sesuai atau kurang.

3.4.1 Select Modeling Technique

Pada tahapan ini model yang akan dibuat ditentukan. Konsep dan struktur dasar model dibuat. Pada penelitian ini algoritma model yang dipilih adalah model dengan algoritma *CNN* atau *Convolutional Neural Network*. Pada penelitian ini peneliti tidak membuat model *Convolutional Neural Network* sendiri. Melainkan menggunakan salah satu model transfer learning yaitu arsitektur *DenseNet201*. Pemilihan dari *DenseNet201* sendiri dikarenakan performanya yang baik namun kurang populer dibanding *VGG19* dan *MobileNet*. Seperti pada penelitian perbandingan algoritma yang mana didalamnya hanya ada *AlexNet*, *VGG*, *ResNet*, dan *SqueezeNet* tanpa ada *DenseNet* [50], [51]. Oleh karena itu penelitian kali ini memilih untuk menggunakan *DenseNet201* yang kurang populer namun memiliki tingkat akurasi yang tidak kalah baik dibanding arsitektur lain terutama dalam klasifikasi gambar [52].

3.4.2 Generate Test Design

Pada tahapan ini model yang tadi sudah dipilih kemudian dilakukan pembuatan *design test* model yang akan dilakukan. Sebelumnya kita sudah menentukan arsitektur *DenseNet201* sebagai main model atau model utama yang akan digunakan. Sebuah model tidak dapat dipakai kecuali dilakukan penyesuaian lanjutan agar dapat digunakan pada penelitian yang berbeda. Pada penelitian ini akan menambahkan *densenet layer* dengan *softmax* sebagai *activation* yang

berfungsi untuk menentukan jumlah kelas yang akan diidentifikasi. Untuk *skenario test design* pada penelitian kali ini akan melakukan percobaan yang berbeda pada *optimizer, learning rate, dan epoch*. Percobaan pada *optimizer, learning rate, dan epoch* ditentukan berdasarkan penelitian yang sudah pernah dilakukan sebelumnya [16], [53]–[55]. Sehingga akan dilakukan pembuatan model sebanyak 18 dengan variasi konfigurasi seperti yang disajikan pada tabel 3.1.

Tabel 3. 1 Skenario Pengujian

skenario	<i>Optimizer</i>	<i>Learning Rate</i>	<i>epoch</i>
1	<i>Adam</i>	0,001	5
2	<i>RMSprop</i>	0,001	5
3	<i>SGD</i>	0,001	5
4	<i>Adam</i>	0,001	10
5	<i>RMSprop</i>	0,001	10
6	<i>SGD</i>	0,001	10
7	<i>Adam</i>	0,01	5
8	<i>RMSprop</i>	0,01	5
9	<i>SGD</i>	0,01	5
10	<i>Adam</i>	0,01	10
11	<i>RMSprop</i>	0,01	10
12	<i>SGD</i>	0,01	10
13	<i>Adam</i>	0,0001	5
14	<i>RMSprop</i>	0,0001	5
15	<i>SGD</i>	0,0001	5
16	<i>Adam</i>	0,0001	10
17	<i>RMSprop</i>	0,0001	10
18	<i>SGD</i>	0,0001	10

Untuk pengaturan lain seperti *loss* dan *metrics* model akan menggunakan pengaturan yang sama. Untuk *loss* akan menggunakan *categorical_crossentropy*, dan untuk *metrics* akan menggunakan *accuracy*.

3.4.3 Build the Model

Pada tahapan ini semua skenario yang sudah ditetapkan sebelumnya dijalankan dan dilatih. Tahapan ini akan menampilkan *source code* model dan proses *training* dari model.

3.5 Evaluation

Tahap ini dilakukan dengan melihat tingkat performa dari pola yang dihasilkan oleh model yang sudah dilatih. Parameter yang akan digunakan dalam proses evaluasi komparasi adalah *confusion matrix* dengan konfigurasi nilai akurasi, dan *evaluation* yang akan diterima yaitu *loss*, *accuracy*, *val_loss*, dan *val_accuracy*. Tahapan evaluasi terbagi menjadi dua yaitu.

3.5.1 Evaluation Result

Pada tahapan *evaluation result* semua model yang sudah di latih akan ditampilkan dan dibandingkan kemudian pada tahapan ini juga akan ditampilkan grafik yang menunjukkan nilai akurasi dan *epoch* dengan menggunakan *matplotlib.pyplot*. proses evaluasi akan dilakukan dengan menampilkan hasil dari 3 skenario dengan *epoch* dan *learning rate* yang sama agar lebih mudah membandingkan hasil akurasi.

3.5.2 Determine Next Step

Pada tahapan model yang sudah dievaluasi kemudian dipilih yang terbaik yang kemudian dinilai apakah dapat dilanjutkan ke tahapan selanjutnya atau tidak. Jika model dianggap dapat digunakan maka akan disiapkan untuk menjadi model yang akan di-*deploy* pada tahapan selanjutnya. kemudian model akan dikonversi kedalam bentuk *Data Hierarkis (HDF)*.

3.6 Deployment

Tahapan deployment merupakan tahapan terakhir pada penelitian ini. Pada tahapan deployment model terbaik yang sudah ditentukan sebelumnya akan dibuat menjadi sebuah *API*. Tahapan deployment terbagi menjadi dua tahap yaitu:

3.6.1 Deployment Plan

Pada tahapan ini proses deploy akan dijelaskan. Dimulai dari *flowchat API* hingga proses *deployment* model ke dalam sebuah *API (Application Programming Interface)*.

3.6.2 Produce Final Report

Pada tahapan ini menjelaskan tahapan *deployment* dan melakukan *review* hasil *deployment* seperti keberhasilan aplikasi dan akurasi dalam melakukan klasifikasi pada hasil *deployment* yang sudah dibuat.

