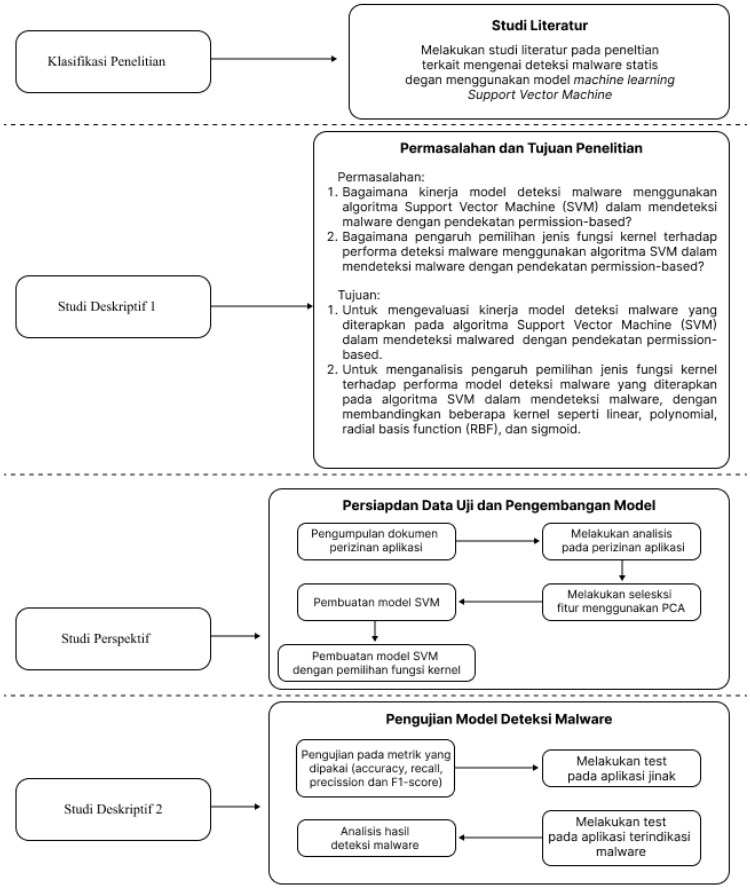
BAB III

METODELOGI PENELITIAN

3.1 Desain Penelitian

Penelitian ini menggunakan metode Design Research Methodology (DRM) yang dikembangkan oleh Blessing dan Chakrabarti. Metode ini melibatkan pendekatan yang terstruktur dalam penelitian ilmu desain, yang terdiri dari empat tahap: Klarifikasi Penelitian, Studi Deskriptif 1, Studi Preskriptif, dan Studi Deskriptif 2 (Blessing & Chakrabarti, 2009). Metodologi ini bertujuan untuk memahami, mengeksplorasi, dan meningkatkan praktik desain secara sistematis, serta menghasilkan pengetahuan deskriptif dan preskriptif (Ebneyamini, 2022). Gambar di atas menggambarkan skema penelitian yang dirancang berdasarkan DRM.



Gambar 3.1 Desain Penelitian

Berdasarkan desain penelitian pada gambar diatas berikut penjelasan dari masing-masing tahapan :

3.1.1 Klasifikasi Penelitian

Pada fase ini, studi literatur dilaksanakan dengan tujuan untuk menemukan penelitian-penelitian terdahulu yang relevan dan berkaitan dengan deteksi malware, penerapan *machine learning* dalam deteksi malware, penggunaan metode SVM dalam deteksi malware, serta optimasi model guna meningkatkan akurasi dan efisiensi model dalam mendeteksi malware. Sumber-sumber yang dijadikan rujukan mencakup jurnal, publikasi laporan ancaman digital dan buku-buku yang berkaitan.

3.1.2 Studi Deskriptif 1

Fase selanjutnya adalah merumuskan permasalahan dan tujuan penelitian berdasarkan pemahaman yang diperoleh dari studi literatur. Permasalahan di sini mengacu pada serangkaian pertanyaan yang perlu dipecahkan. Di sisi lain, tujuan penelitian akan mencerminkan hasil yang diharapkan dari penelitian yang dilakukan, yaitu peningkatan akurasi dalam mendeteksi malware melalui penerapan metode machine learning SVM serta optimasi dalam penggunaan fungsi kernel.

3.1.3 Studi Perspektif

Pada fase selanjutnya, langkah-langkah untuk mengembangkan sistem deteksi malware akan diterapkan dengan memanfaatkan metode machine learning, khususnya menggunakan model SVM. Proses ini melibatkan beberapa tahap penting, mulai dari pengumpulan perizinana aplikasi yang relevan hingga evaluasi terhadap model deteksi yang dihasilkan. Tahap pertama adalah menganalisis perizinan yang bertujuan untuk mempersiapkan data agar siap digunakan. Proses ini meliputi pembersihan data, normalisasi, dan transformasi fitur guna meningkatkan kualitas dataset. Setelah tahap ini, data akan dibagi menjadi dua bagian utama: data latih yang digunakan untuk melatih model SVM dan data uji yang berfungsi untuk mengevaluasi kinerja model.

Proses selanjutnya adalah membuat model SVM dengan data yang sudah melalui proses analisis dan normalisasi, model yang dibuat kemudian dikembangkan mencakup pemilihan jenis kernel yang tepat, seperti *linear*,

polynomial, atau RBF guna memastikan bahwa model mampu memisahkan data malware dari non-malware dengan tingkat akurasi yang tinggi. Model yang sudah dikembangkan kemudian akan dievaluasi melalui pengujian pada aplikasi jinak dan aplikasi berbahaya. Hasil uji yang diperoleh kemudian akan dibandingkan dengan model sebelum pengembangan dengan konfigurasi kernel dan setiap kernel yang digunakan akan dievalusi untuk mengetahui efektifitas setiap konfigurasi kernel yang digunakan.

3.1.4 Studi Deskriptif 2

Pada fase terakhir, setelah proses pelatihan, model di uji cobakan terhadap aplikasi-aplikas jinak dan aplikasi-aplikasi yang terindikasi malware untuk mengetahui hasil pelatihan model. Model kemudian dievaluasi dengan menggunakan berbagai metrik untuk mengetahui kinerja model. Metrik yang digunkan adalah *accurary*, *precision*, *recall*, dan *F1-score*. Tahap optimasi dilaksanakan untuk meminimalkan kesalahan dengan mengimplementasikan fungsi kernel pada model.

3.2 Instrumen Penelitian

Instrumen yang dalam penelitian ini adalah analisis hasil pengujian yang dilakukan melalui pendekatan eksperimental. Tujuan dari penelitian ini adalah untuk mengumpulkan dan menganalisis data yang diperoleh dari pengujian model SVM dalam mendeteksi malware berdasarkan data izin aplikasi. Eksperimen dilaksanakan dengan menerapkan model SVM menggunakan berbagai konfigurasi kernel dan parameter untuk menilai efektivitas metode dalam mendeteksi malware. Hasil analisis dari eksperimen ini digunakan untuk menilai kinerja model, termasuk tingkat *accuracy*, *precission*, *recall*, dan *f1-score* dalam mendeteksi aplikasi yang berpotensi mengandung malware. Eksperimen ini juga disusun untuk menganalisis kinerja berbagai jenis kernel dalam SVM, termasuk kernel linear, polynomial, dan RBF, dengan tujuan untuk menemukan konfigurasi yang paling efisien. Melalui pendekatan eksperimen ini, diharapkan model SVM dapat memperluas jangkauan deteksi malware serta lebih efektif dalam mengidentifikasi potensi ancaman dibandingkan dengan metode deteksi yang konvensional.

3.2.1 Analisis Pengumpulan Data Penelitian

Penelitian yang dimulai dengan pengumpulan data berupa dokumen perizinan aplikasi, yang berfungsi sebagai informasi dasar untuk analisis lebih lanjut. Data tersebut mencakup berbagai jenis perizinan yang diajukan oleh aplikasi selama proses instalasi atau penggunaannya. Dataset yang digunakan adalah kumpulan data perizinan aplikasi yang dapat diakses secara publik di internet, termasuk aplikasi yang berasal dari Google Play Store. Dataset ini diperoleh dari situs web Kaggle.com, yang diunggah oleh Saurabh Shahane pada tahun 2021. Terdapat total 29.999 entri data yang berkaitan dengan perizinan aplikasi yang dapat dilihat pada lampiran 8, sehingga dijagikannya sebagai sumber data yang signifikan dan beragam untuk mendukung analisis yang mendalam. Dataset ini mencakup data mengenai 184 jenis atribut sesuai yang tercantum pada lampiran 9**,** dengan 174 fitur yang berhubungan dengan izin aplikasi dan 10 atribut sisanya adalah informasi umum aplikasi. Setiap entri dalam dataset ini menggambarkan izin yang diajukan oleh aplikasi tertentu, mencakup informasi seperti nama aplikasi, jenis izin, kategori aplikasi, serta data relevan lainnya. Dengan informasi yang tersedia, dataset ini mendukung berbagai metode analisis, termasuk identifikasi pola perizinan, analisis risiko privasi, dan evaluasi keamanan aplikasi. Selanjutnya, dilakukan analisis terhadap data perizinan untuk mengidentifikasi pola atau karakteristik tertentu yang dapat menunjukkan adanya malware. Tujuan dari tahap ini adalah untuk memahami hubungan antara jenis perizinan yang diminta oleh aplikasi dengan kemungkinan aplikasi tersebut mengandung malware. Hasil dari analisis ini akan menjadi landasan untuk membangun model deteksi yang berbasis pada machine learning.

Kelayakan data yang digunakan dalam penelitian ini dapat dijelaskan dengan beberapa aspek. Aspek pertama, dateset yang diambil dari Kaggle.com merupakan hasil pengumpulan data yang berasal dari Google Play Store maupun website pengadaan aplikasi *open-source.* Setiap aplikasi yang diunggah di Google Play Store wajib mencantumkan daftar perizinan yang dibutuhkan wajib mencantumkan daftar perizinan yang dibutuhkan sebagai dari kebijakan transaparansi dan perlindungan pengguna. Data yang diambil oleh Saurabh Shahane

mengacu dari publikasi yang diunggah oleh Arvind Mahindru pada tahun 2018 (Mahindru, A., & Sangal, A. L. 2021).

3.2.2 Analisis Pengumpulan Data Hasil Penelitian

Untuk memastikan adanya jaminan mengenai optimasi SVM dengan implementasi fungsi kernel dalam peningkatan efektifitas deteksi malware, dapat diukur dengan *metric-metric* pengujian. *Metric-metric* digunakan untuk menghitung akurasi hasil dari model deteksi malware yang telah melewati proses pelatihan dan proses optimasi dengan implementasi fungsi kernel. *Metric* yang umun digunakan untuk mengukur keberhasilan sebuah ML seperti yang dibahas oleh (Sitarz, 2023). Dalam penelitian ini, *metric* yang digunakan pada penelitian ini yaitu *accucary, precision, recall* dan *F1-score.* Pemilihan metri-metrik ini didasarkan pada kebutuhan untuk mengevaluasi model secara menyeluruh, tertutama dalam melihat ketidak seimbangan data yang mungkin terjadi antara jumlah aplikasi malware dan aplikasi jinak.

1. *Accuracy*

Accuracy pada implementasi klasifikasi biner penggunaan metrik ini untukmengukur proporsi prediksi TP dan TN. Metrik ini digunakan untuk memberikan gambaran umum dari performa dari model yang digunakan (Balayla, 2021).

2. *Precision*

Metrik ini digunakan untuk mengukur seberapa tepat model dalam mengidentifikasi data tanpa menghasilkan banyak kesalahan dalam mengklasifikan aplikasi yang sebenarnya aman sebagai malware atau FP. *Precision* yang tinggi dapat digunakan untuk memastikan bahwa model memberikan peringatan yang akurat dalam mendeteksi model (Tripathi dkk., 2021).

3. *Recall*

Metrik ini digunakan untuk mengukur kemampuan model dalam mendeteksi semua aplikasi malware, yang berkaitan dengan FN, dimana aplikasi malware dapat terlewatkan dan dianggap sebagai aplikasi yang aman. *Recall* yang tinggi dapat membantu dalam memastikan bahwa model memberikan hasil yang lebih akurat, dengan mengurangi risiko dalam mendeteksi FN (Gonzalez-Abril dkk., 2017).

4. *F1-Score*

Metrik ini digunakan untuk mengevaluasi hasil dari kombinasi *precision dan recall* untuk memberikan penilaian seimbang terhadap performa model. *F1-score* berperan dalam menilai model secara objecktif, terutama ketika terdapat trade-off, sehingga memastikan bahwa model dapat secara efektif medeteksi malware tanpa menghasilkan jumal FN yang banyak (Balayla, 2021).

Dalam kontek ini, TP merupakan aplikasi yang terindikasi bahaya, TN merupakan aplikasi yang tidak berbahaya. FN merupakan aplilkasi bahaya yang terindikasi sebagai aplikasi yang tidak berbahaya dan FP merupakan aplikasi yang tidak berbahaya yang terindikasi sebagai aplikasi yang berbahaya.

3.3 Alat dan Bahan Penelitian

Pada penelitian ini, menggunakan alat dan bahan sebagai berikut:

3.3.1 Alat Penelitian

Alat penelitian yang digunakan sebagai lingkungan pengembangan dan peluncuran aplikasi adalah sebagai berikut

1. Perangkat Keras (*Hardware*)

Penelitian ini menggunakan beberapa alat perangkat keras (*hardware)* untuk menunjang pelaksanaan penelitian, diantaranya seperti yang dijelaskan pada tabel dibawah berikut:

Tabel 3.1

Kepustakaan Perangkat keras (*Hardware*) yang Digunakan

Kepustakaan / Alat	Funsgi
Processor Intel Core i5 12400F gen 12	Sebagai otak dari komputer yang mengelola semua perintah dan operasi yang dilakukan oleh perangkat.
Installed RAM 16 GB	Sebagai penyimpanan sementara yang digunakan oleh sistem untuk menyimpan data yang sedang diolah
Android	Sebagai alat input dan output utama untuk mendapatkan data perizinan yang dibutuhkan.

2. Perangkat Lunak (*Software*)

Selain perangkat keras (*hardware*) penelitian ini juga menggunakan beberapa perangkat lunak (*software)* untuk menunjang pelaksanaan penelitian, diantaranya seperti yang dijelaskan pada tabel dibawah berikut:

Tabel 3.2

Kepustakaan Perangkat Lunak (*Software*) yang Digunakan

Kepustakaan / Alat	Funsgi
VS *Code*	Alat pengembangan untuk menulis dan mengedit kode yang digunakan untuk menyusun dan memodifikasi skrip pengujian
Python	Bahasa pemrograman yang digunakan untuk membuat skrip.
Jupyter NoteBook	Alat pengembangan berbahasa *Python* dalam membuat skrip agar lebih terstruktu
Apk Analyzer	Alat untuk menganalisis perizinan dan semua data yang ada dalam sebuah aplikasi pada Android
Apk Tools	Alat untuk menganalisis perizinan dan semuda data yang ada dalam sebuah aplikasi pada desktop

3.3.2 Bahan Penelitian

Bahan penelitian yang digunakan sebagai lingkungan pengembangan dan peluncuran aplikasi adalah sebagai berikut

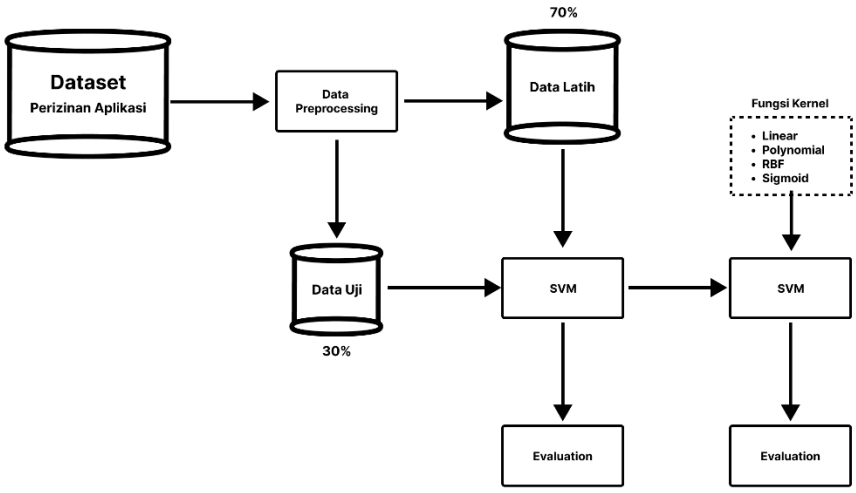
1. Aplikasi Jinak dan Malware yang di Uji Coba

Pada penelitian ini, aplikasi yang dianalisis untuk pengujian model terbagi menjadi dua kategori utama, yaitu aplikasi jinak dan aplikasi berbahaya. Pemisahan ini bertujuan untuk menilai kemampuan model dalam membedakan antara kedua jenis aplikasi berdasarkan data izin yang diberikan. Aplikasi yang digunakan dalam pengujian ini diambil dari Google Play Store, yang merupakan platform resmi untuk distribusi aplikasi Android. Pemilihan Google Play Store didasarkan pada

reputasinya sebagai sumber aplikasi yang umumnya aman dan dapat dipercaya. Sementara itu, aplikasi malware yang akan dianalisis dalam model deteksi malware berasal dari kumpulan dataset ashishb/android-malware: Collection of android malware samples (github.com), yang merupakan koleksi sampel malware Android yang dapat diakses secara publik. Dataset ini mencakup berbagai sampel malware yang mencerminkan beragam teknik dan strategi yang diterapkan oleh pengembang perangkat lunak berbahaya untuk mengeksploitasi perangkat pengguna. Aplikasi modifikasi juga digunakan sebagai sebagai digunakan untuk dianalisi dalam model deteksi malware yang diperoleh dari website apkdone.com.

3.4 Prosedur Penelitian

Prosedur penelitian akan mengikuti langkah-langkah yang telah ditetapkan dalam desain penelitian yang telah diuraikan sebelumnya. Rincian lebih lanjut mengenai prosedur penelitian ini dapat dilihat pada gambar dibawah berikut ini:



Gambar 3.2 Prosedur Penelitian

Pada gambar 3.2 menjelaskan prosedur penelitian yang dimulai dengan pengumpulan data yang berupa dokumen perizinan aplikasi, yang berfungsi sebagai informasi dasar untuk analisis lebih lanjut. Data tersebut mencakup berbagai jenis perizinan yang diajukan oleh aplikasi selama proses instalasi atau penggunaannya. Selanjutnya, dilakukan analisis terhadap data perizinan untuk mengidentifikasi pola atau karakteristik tertentu yang dapat menunjukkan adanya malware. Tujuan dari tahap ini adalah untuk memahami hubungan antara jenis perizinan yang diminta

oleh aplikasi dengan kemungkinan aplikasi tersebut mengandung malware. Hasil dari analisis ini akan menjadi landasan untuk membangun model deteksi yang berbasis pada machine learning.

Tahap berikutnya melibatkan pembuatan model SVM, yang dirancang untuk mendeteksi aplikasi yang berpotensi mengandung malware. Dalam proses ini, pemilihan fungsi kernel yang paling tepat dilakukan untuk meningkatkan akurasi model. Setelah model selesai dikembangkan, pengujian dilakukan pada aplikasi tertentu yang sudah dianalisis sebelumnya, seperti untuk mengevaluasi performa awal model dalam mendeteksi potensi ancaman. Model juga diuji dengan menggunakan data aplikasi yang telah terindikasi mengandung malware, untuk mengukur efektivitasnya dalam situasi nyata. Pada tahap akhir, analisis hasil deteksi malware dilakukan untuk mengevaluasi kinerja model secara keseluruhan, termasuk tingkat akurasi, presisi, serta kemampuannya dalam meminimalkan false positives dan false negatives. Analisis ini memberikan wawasan yang penting untuk pengembangan lebih lanjut dalam deteksi malware yang berbasis SVM.