

BAB I

PENDAHULUAN

1.1 Latar Belakang

Peramalan cuaca merupakan salah satu aplikasi sains data yang paling krusial dan berdampak tinggi. Akurasi ramalan cuaca memiliki implikasi langsung terhadap berbagai sektor vital, mulai dari pertanian (penentuan jadwal tanam dan panen), transportasi (keselamatan penerbangan dan pelayaran), hingga manajemen bencana. Di Indonesia, urgensi ini semakin nyata. Data dari Badan Nasional Penanggulangan Bencana (BNPB) secara konsisten menunjukkan bahwa bencana yang diakibatkan oleh aktivitas cuaca seperti siklus hidrologi, curah hujan, temperatur, angin dan kelembapan (bencana hidrometeorologi) seperti banjir, tanah longsor, dan puting beliung mendominasi statistik bencana nasional[1]. Sepanjang tahun 2023 saja, tercatat lebih dari 5.000 kejadian bencana, dengan 90% di antaranya merupakan bencana hidrometeorologi. Kegagalan atau ketidakakuratan dalam meramalkan fenomena cuaca ekstrem dapat berujung pada kerugian ekonomi miliaran rupiah dan, yang lebih penting, hilangnya nyawa.

Permasalahan mendasar yang dihadapi dalam peramalan cuaca berbasis *machine learning* terletak pada ketidakmampuan model untuk mengekstrak informasi temporal secara optimal dari data mentah. Data cuaca yang tersedia dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) bersifat multivariat, mencakup variabel seperti suhu minimum, suhu maksimum, suhu rata-rata, kelembapan, dan curah hujan, yang semuanya saling berinteraksi membentuk pola yang berubah dari waktu ke waktu. Kualitas data dan fitur yang dimasukkan ke dalam model sangat menentukan hasil prediksi, karena model yang menerima *input* yang buruk atau tidak representatif akan menghasilkan *output* yang tidak dapat diandalkan [2]. Tanpa proses rekayasa fitur yang tepat, model *machine learning* seperti *Support Vector Regression* (SVR) hanya menerima variabel mentah sebagai input, sehingga informasi penting seperti tren jangka pendek, pola musiman, dan ketergantungan antar waktu tidak dapat direpresentasikan secara eksplisit dalam proses pembelajaran model.

Tantangan teknis dalam permasalahan ini bersifat berlapis dan tidak sederhana. Di antara berbagai algoritma *machine learning*, SVR telah terbukti mampu menangani masalah regresi non-linear melalui penggunaan kernel, salah satunya *Radial Basis Function* (RBF), yang mampu mentransformasi data ke dalam ruang dimensi lebih tinggi sehingga pola yang tidak linier dapat dimodelkan secara efektif [3]. Namun, SVR secara arsitektur tidak memiliki kemampuan bawaan untuk mengenali urutan waktu (*temporal ordering*), yang merupakan karakteristik fundamental dari data cuaca. Hal ini berarti seluruh informasi temporal harus dikodekan secara eksplisit ke dalam bentuk fitur sebelum data dimasukkan ke dalam model [4]. Di sisi lain, penambahan fitur turunan tanpa seleksi yang cermat justru dapat memperburuk performa model. Kondisi ini dikenal sebagai *curse of dimensionality*, yaitu fenomena di mana peningkatan jumlah dimensi fitur pada data dengan ukuran tetap dapat menyebabkan *overfitting* dan menurunkan kemampuan generalisasi model secara signifikan [5]. Oleh karena itu, diperlukan kombinasi yang sistematis antara *feature creation* untuk memperkaya representasi data temporal dan *feature selection* untuk membuang fitur yang tidak relevan, sehingga model SVR dapat bekerja secara optimal.

Sejumlah penelitian terdahulu telah menerapkan SVR pada kasus prediksi cuaca di Indonesia. Imron (2025) membandingkan SVR dengan Random Forest untuk peramalan curah hujan dan menemukan bahwa meskipun kedua model mampu menangani data hidrometeorologi, SVR memiliki keunggulan tersendiri dalam memodelkan hubungan non-linear [6]. Sejalan dengan itu, Safitri (2021) menerapkan SVR pada prediksi curah hujan di Jawa Timur dan membuktikan bahwa SVR dapat diaplikasikan secara efektif pada data cuaca lokal yang memiliki karakteristik geografis tertentu [7]. Setio Pribadi (2025) juga menunjukkan bahwa SVR mampu mengungguli regresi linear dalam memprediksi temperatur udara di Kota Malang, menegaskan bahwa SVR relevan untuk data cuaca yang bersifat fluktuatif [4]. Namun, apabila ditelaah secara menyeluruh, mayoritas penelitian tersebut memperlakukan *feature engineering* hanya sebagai prosedur pra-pemrosesan standar, bukan sebagai variabel eksperimen yang dieksplorasi pengaruhnya secara sistematis terhadap performa model. Belum ada studi yang secara khusus dan terstruktur membandingkan skenario-skenario *feature*

engineering yang dirancang spesifik untuk model SVR pada konteks peramalan cuaca multivariat di Indonesia. Kesenjangan inilah yang menjadi motivasi utama penelitian ini.

Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan memberikan dua kontribusi utama. Pertama, secara metodologis, penelitian ini mengusulkan dan menguji tiga skenario *feature engineering* secara terstruktur menggunakan kerangka kerja CRISP-DM [17], yaitu skenario *baseline* tanpa rekayasa fitur, skenario *feature creation* dengan fitur lag, *rolling statistics*, dan *seasonal encoding* berbasis sinus-cosinus, serta skenario kombinasi *feature creation* dan *feature selection* berbasis korelasi Pearson. Perbandingan ketiga skenario ini menghasilkan bukti empiris yang terukur mengenai sejauh mana *feature engineering* mampu meningkatkan akurasi peramalan SVR pada data cuaca multivariat. Kedua, secara praktis, temuan penelitian ini diharapkan dapat menjadi rekomendasi teknis bagi pengembangan sistem peramalan cuaca berbasis *machine learning* yang lebih andal, khususnya untuk mendukung upaya mitigasi bencana hidrometeorologi di Indonesia[1].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, yang dapat dirumuskan masalah dalam penelitian ini adalah:

1. Bagaimana pengaruh penerapan teknik *feature engineering* terhadap akurasi model *Support Vector Regression* dalam meramalkan data cuaca multivariat?
2. Bagaimana kinerja model terhadap skenario *Feature Engineering* pada model *Support Vector Regression* dalam peramalan cuaca?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini untuk :

1. Mengimplementasikan teknik *feature engineering* pada data cuaca multivariat, berupa *feature creation* dan *feature selection* untuk mendukung proses peramalan menggunakan *Support Vector Regression* (SVR).
2. Mengetahui dan menganalisis pengaruh masing-masing teknik *feature engineering* terhadap peningkatan akurasi model SVR.

3. Menentukan kombinasi fitur serta parameter SVR yang paling optimal dalam meramalkan data cuaca multivariat.

1.4 Batasan Penelitian

Agar penelitian ini tetap berada dalam ruang lingkup yang telah ditentukan, maka ditetapkan beberapa batasan masalah sebagai berikut:

- a. Penelitian ini hanya berfokus pada peramalan data cuaca multivariat, yang mencakup variabel seperti curah hujan, suhu, kelembapan, dan tekanan udara.
- b. Dataset yang digunakan merupakan data historis cuaca bulanan yang diperoleh dari laman [datonline.bmkg](https://datonline.bmkg.go.id) (sumber resmi BMKG), tanpa memasukkan variabel eksternal lain seperti fenomena ENSO, topografi, ataupun data satelit.
- c. Model yang digunakan terbatas pada algoritma *Support Vector Regression* (SVR) dengan pendekatan kernel RBF sesuai cakupan penelitian.
- d. Teknik *feature engineering* yang diterapkan dibatasi pada dua skenario, yaitu *feature creation* (*lag*, *rolling statistics*, *encoding* musiman) dan *feature selection* (*korelasi*, *mutual information*, atau metode seleksi fitur lain yang relevan).
- e. Evaluasi model dilakukan menggunakan metrik standar regresi, yaitu *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi (R^2).
- f. Visualisasi hasil prediksi hanya ditampilkan dalam bentuk grafik perbandingan antara nilai aktual dan hasil peramalan.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat seperti yang tercantum di bawah ini:

1) Bagi Pengguna

Penelitian ini bertujuan menyediakan model peramalan cuaca yang lebih akurat dan efisien, sehingga pengguna terutama instansi terkait seperti pertanian, transportasi, dan pengelolaan bencana dapat memperoleh informasi cuaca yang lebih reliabel untuk mendukung pengambilan keputusan.

2) Bagi Penulis

Penelitian ini memberikan kesempatan bagi penulis untuk mengembangkan kemampuan analitis dan teknis dalam pengolahan data cuaca, penerapan *feature*

engineering, serta implementasi model *Support Vector Regression*. Selain itu, penelitian ini menjadi portofolio yang berharga dalam pengembangan keahlian pada bidang *data science* dan *machine learning*.

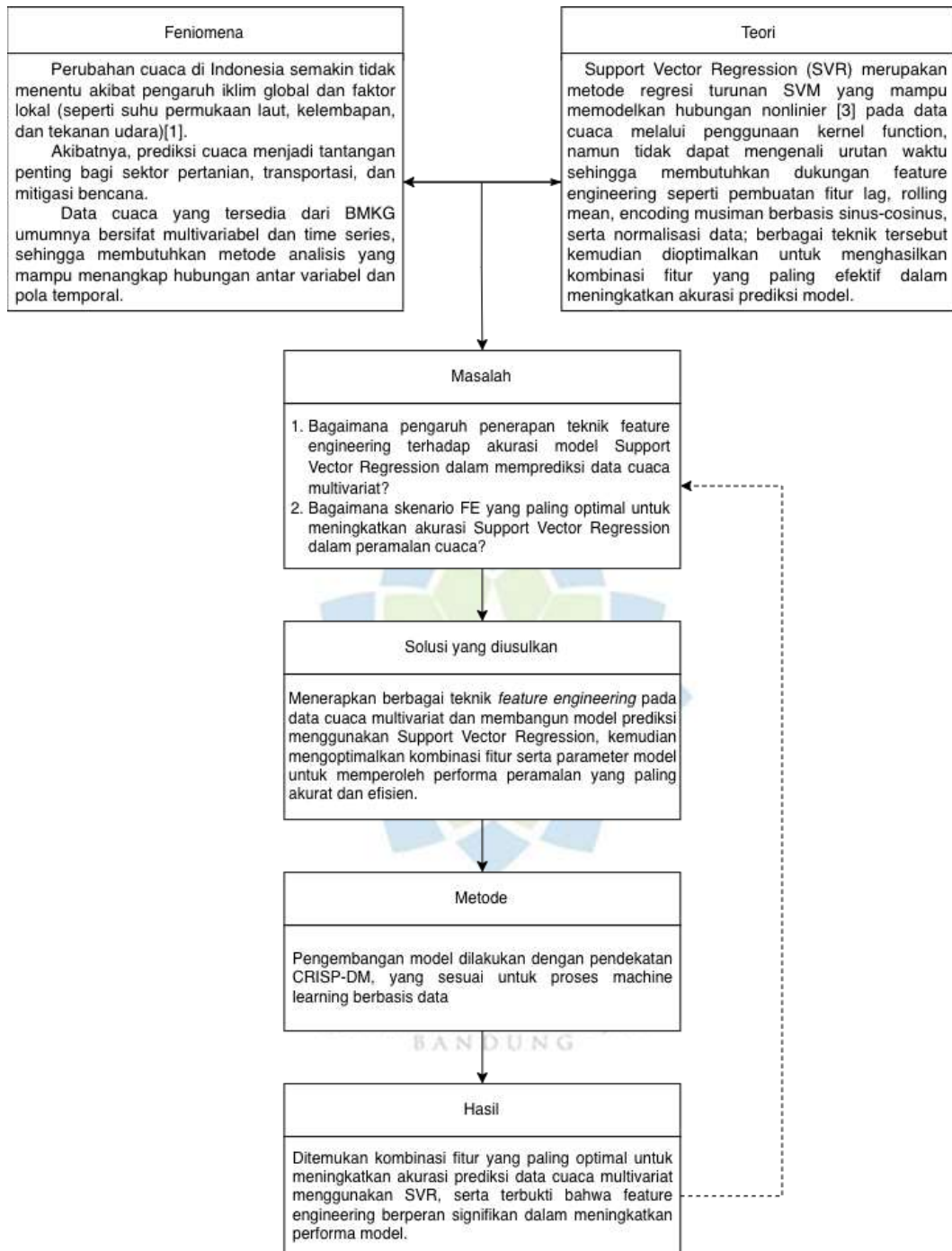
3) Bagi Akademik

Hasil penelitian ini dapat dijadikan referensi atau acuan bagi penelitian berikutnya yang membahas peramalan cuaca, optimasi fitur, maupun penerapan SVR dalam data *time series*. Penelitian ini juga dapat menjadi bahan pembelajaran untuk memperluas wawasan terkait teknik *forecasting* pada data multivariat.

1.6 Kerangka Pemikiran Penelitian

Kerangka pemikiran yang dijadikan sebagai alur dari penelitian ini dijelaskan melalui Gambar 1.1 berikut.





Gambar 1.1 Kerangka Pemikiran

Gambar 1.1 tersebut menggambarkan alur pemikiran penelitian yang berfokus pada optimalisasi teknik *feature engineering* untuk meningkatkan akurasi prediksi data cuaca multivariat menggunakan model *Support Vector Regression* (SVR). Penelitian ini berangkat dari fenomena perubahan cuaca di Indonesia yang semakin tidak menentu akibat pengaruh faktor iklim global dan lokal, sehingga

kemampuan memprediksi kondisi cuaca menjadi semakin penting bagi sektor pertanian, transportasi, hingga mitigasi bencana. Data cuaca yang tersedia umumnya bersifat multivariabel dan berbentuk deret waktu, sehingga diperlukan metode analisis yang mampu menangkap hubungan antar variabel serta pola temporal.

Berdasarkan teori, SVR merupakan metode regresi yang mampu memodelkan hubungan nonlinier melalui penggunaan *kernel function*, namun tidak dapat mengenali urutan waktu secara alami. Oleh karena itu, diperlukan dukungan *feature engineering* seperti pembuatan fitur *lag*, *rolling statistics*, *encoding* musiman sinus-cosinus, dan normalisasi data untuk memperkuat kemampuan model dalam memahami pola cuaca. Dari fenomena dan teori tersebut, muncul dua masalah utama: bagaimana pengaruh teknik *feature engineering* terhadap akurasi model SVR, dan teknik mana yang paling optimal untuk meningkatkan kemampuan prediksi cuaca.

Sebagai solusi, penelitian ini mengusulkan penerapan berbagai teknik *feature engineering* pada data cuaca multivariat, kemudian membangun model prediksi berbasis SVR dan mengoptimalkan kombinasi fitur serta parameter model guna memperoleh performa peramalan yang paling akurat dan efisien. Seluruh proses pengembangan model dilakukan menggunakan pendekatan CRISP-DM, yang mencakup pemahaman data, persiapan data, pemodelan, evaluasi, hingga penyusunan rekomendasi implementasi.

Hasil yang diharapkan dari penelitian ini adalah ditemukannya kombinasi fitur yang paling efektif dalam meningkatkan akurasi prediksi cuaca, serta bukti bahwa *feature engineering* memiliki peran signifikan dalam meningkatkan performa model SVR. Temuan ini diharapkan dapat menjadi rekomendasi untuk pengembangan sistem peramalan cuaca yang lebih akurat dan efisien di Indonesia.

1.7 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum tentang proses penulisan yang dilakukan untuk menyelesaikan penelitian ini. Sistematika penulisan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menguraikan latar belakang penelitian, perumusan masalah, tujuan penelitian, serta batasan yang ditetapkan agar penelitian tetap fokus. Selain itu, bab ini juga menjelaskan kerangka pemikiran yang menjadi dasar penelitian, metodologi yang digunakan, serta gambaran sistematika penulisan sebagai panduan dalam membaca penelitian ini.

BAB II STUDI PUSTAKA

Bab ini membahas teori-teori dan konsep dasar yang relevan dengan topik penelitian. Kajian pustaka ini bertujuan untuk memberikan pemahaman mendalam serta mendukung analisis dalam penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah metodologi yang digunakan dalam penelitian, mulai dari tahap awal hingga akhir. Pembahasan mencakup analisis produk serta perencanaan eksekusi aplikasi agar penelitian berjalan secara sistematis dan terarah.

BAB IV HASIL DAN PEMBAHASAN

Bab ini terdiri dari dua hal utama, pertama pemaparan tentang temuan atau hasil penelitian berdasarkan tahapan penelitian yang dilakukan. Pemaparan hasil penelitian disesuaikan dengan urutan masalah penelitian. Kedua pembahasan hasil atau temuan penelitian untuk menjawab rumusan penelitian.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan, serta kritik dan saran yang dapat menjadi masukan bagi penelitian selanjutnya. Kesimpulan yang disajikan merupakan intisari dari keseluruhan penelitian dan dapat dijadikan dasar pengembangan lebih lanjut.