

## **BAB IV**

### **ANALISIS DAN PERANCANGAN SISTEM**

Bab ini menyajikan pembahasan serta perancangan sistem prediksi cuaca di Kabupaten Sambas menggunakan algoritma *random forest*. Pembahasan difokuskan pada penguraian kebutuhan-kebutuhan yang diperlukan dalam proses pengembangan sistem.

#### **4.1. Analisis Kebutuhan Non-Fungsional**

Analisis kebutuhan merupakan proses yang dilakukan sebelum mengerjakan tugas akhir ini tentang prediksi cuaca di Kabupaten Sambas menggunakan algoritma *random forest*. yang mencakup kebutuhan untuk perangkat keras dan perangkat lunak.

##### **4.1.1 Kebutuhan Perangkat Keras**

Kebutuhan perangkat keras yang dibutuhkan dalam mengerjakan penelitian tugas akhir ini agar pengerjaan dapat dilakukan karena dalam proses *machine learning* memerlukan perangkat yang mendukung agar dapat berjalan dengan lancar, pada penelitian ini penulis menggunakan spesifikasi perangkat keras sebagai berikut:

- a. Prosesor : Intel(R) Core(TM) i5-5300U CPU @ 2.30GHz (4 CPUs), ~2.3GHz.
- b. RAM : 8 GB.
- c. System type : Windows 10 Pro 64-bit (10.0, Build 19042).

Sedangkan perangkat keras yang dibutuhkan untuk menggunakan aplikasi ini adalah spesifikasi minimum perangkat keras untuk bekerja di kerangka kerja *machine learning* yaitu bisa membuka *web browser* berupa chrome atau yang lainnya.

#### 4.1.2 Kebutuhan Perangkat Lunak

Kebutuhan perangkat lunak yang digunakan dalam membangun aplikasi adalah sebagai berikut:

- a. *Jupyter Notebook*, digunakan untuk menulis baris atau kode program dalam tahapan pemodelan machine learning.
- b. *Visual Studio Code*, digunakan untuk menulis baris atau kode program dalam membangun aplikasi web yang nantinya akan terintegrasi dengan model.
- c. *Anaconda Navigator*, digunakan untuk mengelola *environment python* yang terisolasi.
- d. *Google Chrome*, digunakan untuk pengguna berinteraksi dengan sistem yang dibikin.

#### 4.2. Analisis Kebutuhan Fungsional

Analisis kebutuhan fungsional merupakan penjabaran kebutuhan yang dibutuhkan pada sistem yang akan dibangun. Pada analisis kebutuhan fungsional mencakupi deskripsi global dari aplikasi seperti yang terdapat pada tabel berikut:

Tabel 4. 1 Analisis Kebutuhan Fungsional

No	Kebutuhan	Deskripsi
1.	Aplikasi website	Aplikasi diharapkan dapat berfungsi dengan semestinya, seperti dapat memasukan input data yang diperlukan dan menampilkan hasil pemodelan menggunakan metode <i>Random Forest</i> dari data yang baru di input.
2.	Aplikasi pemodelan	Pemodelan dapat melakukan prediksi <i>regresi</i> untuk membuat prediksi cuaca menggunakan metode <i>Random Forest</i> .

#### 4.3. Analisis Kebutuhan Data

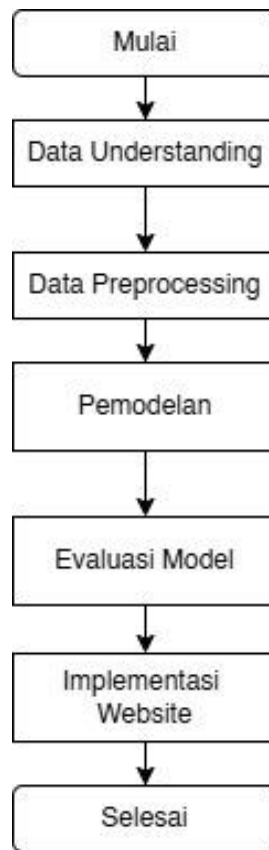
Data yang digunakan merupakan dataset cuaca yang bersumber dari webiste BMKG <http://dataonline.bmkg.go.id/> yang telah dilakukan pengunduhan dengan rentang data tahun 2023-2024 dengan jumlah 366 data dan 11 atribut, seperti yang ditunjukkan pada Tabel 4.2.

Tabel 4. 2 Fitur dan Keterangan

No	Fitur	Keterangan	Jenis Data
1.	Tavg	Berisi data temperatur rata-rata dalam satuan derajat celcius	Kategorikal
2.	RH_avg	Berisi data kelembapan rata-rata dalam satuan derajat celcius	Numerik
3.	RR	Berisi data curah hujan dalam satuan milimeter	Numerik
4.	ss	Berisi data lamanya penyinaran matahari dalam satuan jam	Numerik
5.	ff_x	Berisi data nilai maksimum kecepatan angin yang tercatat dalam satuan meter per detik	Numerik
6.	RR	Berisi data curah hujan dalam satuan milimeter	Numerik
7.	ss	Berisi data lamanya penyinaran matahari dalam satuan jam	Numerik
8.	ff_x	Berisi data nilai maksimum kecepatan angin yang tercatat dalam satuan meter per detik	Numerik
9.	ddd_x	Berisi data arah angin yang tercatat pada saat kecepatan angin mencapai nilai maksimum dalam satuan derajat	Numerik
10.	ff_avg	Berisi data nilai kecepatan angin rata-rata dalam satuan meter per detik	Numerik
22	ddd_car	Berisi data arah angin terbanyak	Kategorikal

#### 4.4.Perancangan Proses Diagram Aliran

Pada perancangan ini dijelaskan alur penelitian dan langkah-langkah mengolah data agar hasilnya dapat berfungsi dengan baik dalam melakukan prediksi.



Gambar 4. 1 Perancangan Proses Diagram Aliran

Gambar 4.1 menjelaskan alur penelitian dan langkah-langkah mengolah data. Adapun penjelasan dari langkah-langkah pada gambar 4.1 sebagai berikut:

1. *Data Understanding*

Tahapan *Data Understanding* merupakan proses awal dalam pemodelan prediktif yang bertujuan untuk memahami struktur, kualitas, serta karakteristik dari dataset yang digunakan. Pemahaman yang baik terhadap data akan membantu dalam menentukan proses *preprocessing*, pemilihan fitur, serta jenis model yang sesuai.

2. *Data Preprocessing*

*Preprocessing* merupakan proses penyiapan data agar data dapat diolah oleh model dengan benar. Adapun tahapan dalam *preprocessing* yaitu *data cleaning*, *data transformation*, *feature engineering*.

### 3. Pemodelan

Setelah data melalui tahap data preprocessing, langkah berikutnya adalah membangun model prediksi menggunakan algoritma *Random Forest*. Algoritma ini dipilih karena kemampuannya dalam menangani data dengan banyak fitur, mengurangi risiko *overfitting*. Untuk mendapatkan performa model yang optimal, dilakukan proses *hyperparameter tuning* menggunakan *GridSearchCV*.

### 4. Evaluasi Model

Setelah dilakukan proses pemodelan menggunakan metode *Random Forest*, langkah selanjutnya adalah melakukan evaluasi terhadap performa model. Evaluasi dilakukan menggunakan metrik evaluasi regresi yang umum digunakan, yaitu *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan  $R^2$  *Score*. Metrik-metrik ini digunakan untuk mengukur sejauh mana hasil prediksi model mendekati nilai sebenarnya, serta untuk menilai akurasi dan ketepatan model dalam memprediksi data cuaca di Kabupaten Sambas.

### 5. Implementasi Website

Hasil dari proses pemodelan dan evaluasi akan disimpan ke format *pickle* dan selanjutnya digunakan untuk diintegrasikan ke dalam sebuah website interaktif menggunakan bantuan kerangka kerja *Streamlit*. Tujuan dari integrasi ini adalah untuk mempermudah pengguna dalam melakukan prediksi cuaca berdasarkan input data secara langsung melalui antarmuka web yang sederhana.

#### 4.4.1 Perancangan *Data Understanding*

Pada perancangan ini dijelaskan tahapan *data understanding* yang digunakan untuk penelitian ini.



Gambar 4. 2 Perancangan Proses *Data Understanding*

Gambar 4.2 menjelaskan tahapan-tahapan dalam *data understanding*.

Adapun penjelasan dari tahapan-tahapan pada gambar 4.2 sebagai berikut:

1. Sumber Data

Dataset yang digunakan dalam penelitian ini merupakan data cuaca harian yang diperoleh dari Data Harian Cuaca BMKG Kab. Sambas yang mencakup periode dari bulan juli 2023 hingga juni 2024.

2. Eksplorasi Nilai Yang Tidak Valid

Setelah dilakukan inspeksi awal terhadap data, ditemukan beberapa baris dengan nilai hilang (*missing values*) dan nilai tidak valid (8888) pada sejumlah kolom. Penanganan nilai hilang ini sangat penting karena dapat mempengaruhi hasil model prediksi.

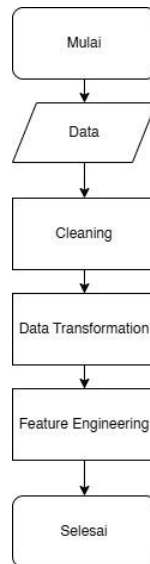
3. Eksplorasi Distribusi Data

Untuk memahami karakteristik distribusi dari masing-masing fitur numerik, dilakukan analisis deskriptif statistik (min, max, mean,

median, std), serta visualisasi seperti histogram untuk mengidentifikasi adanya outlier dan distribusi skewness.

#### 4.4.2 Perancangan *Data Preprocessing*

Pada perancangan ini dijelaskan tahapan *data preprocessing* yang digunakan untuk penelitian ini.



Gambar 4. 3 Perancangan Proses *Data Preprocessing*

Gambar 4.3 menjelaskan tahapan-tahapan dalam *data preprocessing*.

Adapun penjelasan dari tahapan-tahapan pada gambar 4.3 sebagai berikut:

##### 1. *Cleaning*

Membersihkan data dari nilai yang tidak valid atau hilang agar model tidak "tertipu" oleh *noise* dan kesalahan input. Langkah yang dilakukan ialah, menghapus baris yang berisi nilai hilang atau tidak valid.

##### 2. *Data Transformation*

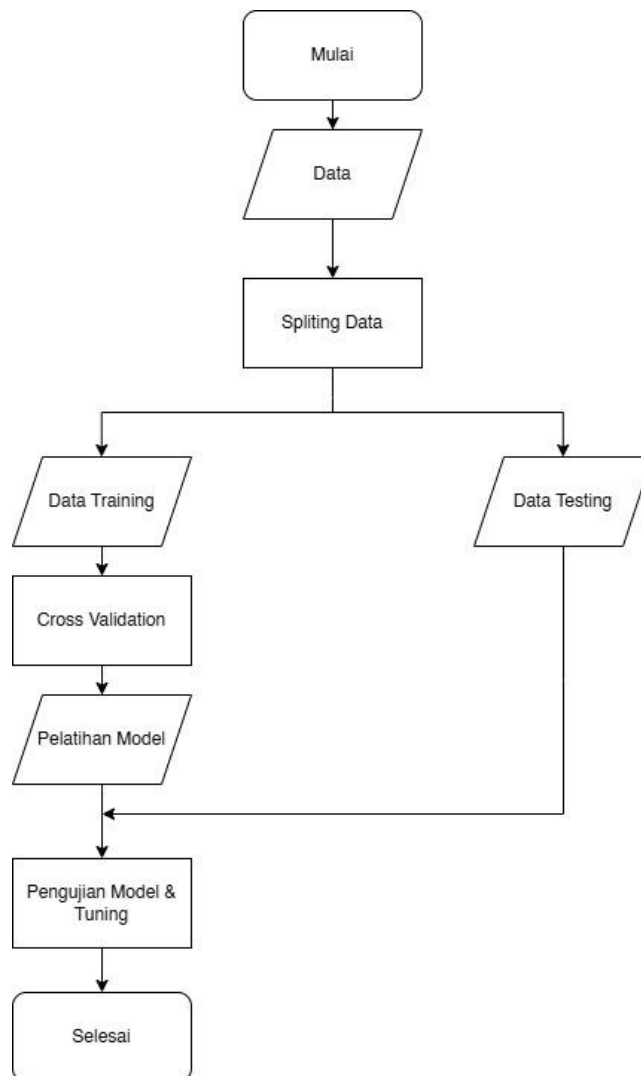
Tahapan *data transformation* yaitu mengubah format atau skala data agar sesuai dengan kebutuhan model. Strategi yang digunakan dalam tahapan ini ialah kolom yang bertipe numerik akan menggunakan pendekatan *minmax scaler*, sedangkan kolom yang bertipe kategori akan dilakukan *encoding* menggunakan *one hot encoding*.

### 3. *Feature Engineering*

Pada tahapan ini, memilih fitur dan target yang paling relevan untuk meningkatkan performa model.

#### 4.4.3 Perancangan Pemodelan

Pada perancangan ini dijelaskan tahapan pembobotan kata yang digunakan untuk penelitian ini.



Gambar 4. 4 Perancangan Pemodelan

Gambar 4.4 menjelaskan tahapan-tahapan dalam pemodelan. Adapun penjelasan dari tahapan-tahapan pada gambar 4.4 sebagai berikut:



### 1. *Splitting Data*

Data pada tahap ini ialah data yang telah melalui tahapan *preprocessing*. Adapun hasil dalam pembagian data latih dan uji pada tahap ini ialah 70% untuk data latih, dan 30% untuk data uji dengan bantuan fungsi `train_test_split` dari pustaka `sklearn`.

### 2. *Cross Validation*

Setelah data dibagi menjadi data latih (70%) dan data uji (30%), proses *cross-validation* dilakukan pada data latih menggunakan teknik *k-fold* dengan nilai  $k = 3$ .

### 3. *Pengujian Model Dan Tuning*

Model yang digunakan dalam penelitian ini adalah *Random Forest Regressor*. Untuk mendapatkan performa model yang optimal, dilakukan proses tuning hyperparameter menggunakan metode *Grid Search Cross Validation (GridSearchCV)*. Teknik ini mencoba berbagai kombinasi parameter.

## 4.4.4 Evaluasi Pemodelan

Pada perancangan akan dilakukan evaluasi dari hasil pemodelan. Evaluasi menggunakan metrik regresi yaitu:

### 1. MSE (Mean Squared Error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Keterangan:

$y_i$  = nilai aktual

$\hat{y}_i$  = nilai prediksi

$n$  = jumlah data

### 2. RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

3. MAE (Mean Absolute Error)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

4.  $R^2$  Score (Coefficient of Determination)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Keterangan:

$\bar{y}$  = rata-rata dari nilai aktual

#### 4.4.5 Implementasi Website

Pertama, model yang telah dibangun dapat disimpan dalam format file.pkl (Pickle File). File.pkl ini berisi parameter dan konfigurasi model yang telah dilatih sebelumnya. Selanjutnya, dalam implementasi menggunakan kerangka kerja *Streamlit*, file.pkl dapat dipanggil dan dimuat ke dalam aplikasi web. Data uji dapat dimasukkan melalui antarmuka aplikasi, dan kemudian model akan melakukan prediksi cuaca berdasarkan data tersebut.

