

# Tinjauan Efektifitas Penggunaan Algoritma Machine Learning pada Sistem Pengolahan Data AWOS di Badan Meteorologi, Klimatologi, dan Geofisika (BMKG)

Mochammad Fariz Zulfiannur<sup>1</sup>, Erica Nur Rahmatina<sup>2</sup>, M Anwar Syaefudin<sup>3</sup>, Alfian Alfarissy<sup>4</sup>

Politeknik Elektronika Negeri Surabaya

Kampus PENS, Jalan Raya ITS Sukolilo, Surabaya 60111

## Abstrak

Kondisi meteorologi, khususnya suhu udara di sekitar landasan pacu, memainkan peran penting dalam operasional penerbangan, mempengaruhi keselamatan dan efisiensi penerbangan. Monitoring suhu dengan akurat di landasan pacu menjadi kunci untuk memastikan kondisi yang aman dan optimal bagi pesawat yang lepas landas dan mendarat. Automatic Weather Observing System (AWOS), sebagai solusi teknologi terkini, secara otomatis mengumpulkan data suhu untuk memberikan informasi cuaca kepada layanan penerbangan. Studi ini menggali prinsip kerja, karakteristik peralatan, dan metode pemeliharaan AWOS, dengan fokus pada parameter cuaca seperti suhu udara, kelembapan, tekanan udara, arah dan kecepatan angin, curah hujan, dan intensitas penyinaran matahari. Kondisi cuaca memainkan peran krusial dalam penentuan rute penerbangan dan mengatasi penundaan pesawat, menjadikannya variabel kunci dalam keselamatan penerbangan. Kerusakan pada sistem AWOS dapat menghentikan pengukuran parameter cuaca, berpotensi menyebabkan penundaan penerbangan. Oleh karena itu, kebutuhan untuk mengembangkan sistem prediksi suhu yang akurat muncul. Penelitian ini bertujuan memprediksi parameter cuaca yang diukur oleh AWOS, khususnya suhu udara di landasan pacu, menggunakan pendekatan machine learning. Studi ini menggunakan algoritma machine learning, termasuk Linear Regression, Random Forest, Decision Tree, dan K-Nearest Neighbors, untuk mengolah data suhu dari AWOS. Variabel sampel melibatkan data suhu dari Runway 10 dan Runway 28 untuk memprediksi suhu pada Runway 1028. Hasil pengujian menggunakan *Root Mean Square Error* (RMSE) didapatkan bahwa model algoritma *Random Forest* sangat cocok digunakan untuk memprediksi suhu dibandingkan ketiga model algoritma yang telah diuji.

**Kata Kunci :** knn, linear regression, random forest, decision tree, klasifikasi, suhu, AWOS

## Abstract

Meteorological conditions, especially the air temperature around the runway, play an important role in flight operations, affecting flight safety and efficiency. Accurate temperature monitoring on the runway is key to ensuring safe and optimal conditions for aircraft taking off and landing. The Automatic Weather Observing System (AWOS), as the latest technology solution, automatically collects temperature data to provide weather information to aviation services. This study explores the working principles, equipment characteristics, and maintenance methods of AWOS, with a focus on weather parameters such as air temperature, humidity, air pressure, wind direction and speed, rainfall, and solar radiation intensity. Weather conditions play a crucial role in determining flight routes and overcoming aircraft delays, making them a key variable in flight safety. A malfunction in the AWOS system could stop measuring weather parameters, potentially causing flight delays. Therefore, the need to develop an accurate temperature prediction system arises. This research aims to predict weather parameters measured by AWOS, especially air temperature on the runway, using a machine learning approach. This study uses machine learning algorithms, including Linear Regression, Random Forest, Decision Tree, and K-Nearest Neighbors, to process temperature data from AWOS. The variable sample includes temperature data from Runway 10 and Runway 28 to predict the temperature on Runway 1028. Test results using *Root Mean Square Error* (RMSE) show that the Random Forest algorithm model is very suitable for predicting temperature compared to the three algorithm models that have been tested.

**Keywords :** knn, linear regression, random forest, decision tree, classification, temperature, AWOS

## 1. PENDAHULUAN

Dalam operasional penerbangan, kondisi meteorologi seperti salah satunya suhu udara di sekitar landasan pacu (Runway), memiliki dampak yang signifikan terhadap keselamatan dan efisiensi penerbangan. Monitoring suhu secara akurat dan efisien di landasan pacu menjadi kunci dalam memastikan kondisi yang aman dan optimal bagi pesawat yang lepas landas dan mendarat. Seiring dengan perkembangan teknologi, *Automatic Weather Observing System* (AWOS) menjadi salah satu solusi untuk mengumpulkan data suhu secara otomatis.

*Automatic Weather Observing System* (AWOS) adalah sebuah sistem komputerisasi di bandara yang secara otomatis melakukan pengukuran parameter cuaca untuk memberikan informasi cuaca kepada layanan penerbangan [1]. Pemahaman tentang prinsip kerja sistem, karakteristik peralatan, dan metode pemeliharaan AWOS sangat penting untuk memastikan peralatan beroperasi dengan optimal. Parameter cuaca yang diukur meliputi suhu udara, kelembapan udara, tekanan udara, arah dan kecepatan angin, curah hujan, dan intensitas penyinaran matahari. Kondisi cuaca memainkan peran krusial dalam menentukan rute penerbangan, mengatasi penundaan keberangkatan pesawat, dan bisa menjadi variabel kunci dalam kejadian kecelakaan transportasi udara.

Jika terjadi kerusakan pada sistem AWOS, pengukuran parameter cuaca menjadi terhenti, menyebabkan potensi penundaan keberangkatan pesawat. Dalam konteks operasional penerbangan, suhu udara di sekitar landasan pacu memiliki dampak yang signifikan terhadap keselamatan dan efisiensi penerbangan. Pentingnya pemantauan suhu pada landasan pacu, memunculkan kebutuhan untuk

mengembangkan sistem prediksi nilai suhu yang dapat memberikan estimasi yang akurat.

Oleh karena itu, penelitian ini bertujuan untuk memprediksi parameter cuaca yang umumnya diukur dengan AWOS, yang berfokus pada prediksi suhu udara. Dalam konteks ini, penggunaan algoritma machine learning menjadi suatu pendekatan yang menarik untuk mengolah data suhu dari AWOS dan membuat prediksi nilai suhu yang lebih tepat. Algoritma machine learning yang digunakan yaitu *Linear Regression*, *Random Forest*, *Decision Tree*, *K-Nearest Neighbors*. Dengan variabel sample nya yaitu data suhu pada Runway 10 dan Runway 28 untuk memprediksi data suhu pada Runway 1028.

## 2. DASAR TEORI

### 2.1 Regression Linear

Linear regression adalah salah satu metode statistik yang digunakan untuk memodelkan hubungan linier antara satu variabel dependen (response) dengan satu atau lebih variabel independen (predictor). Metode ini digunakan untuk memahami dan memprediksi hubungan antar variabel-variabel tersebut. Variabel dependen (response) adalah variabel yang ingin di prediksi sedangkan Variabel Independen (Predictor) adalah variabel-variabel yang digunakan untuk memprediksi variabel dependen. Model *Linear Regression* dapat dinyatakan dalam persamaan matematis sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (1)$$

Penjelasan :

- Y adalah variabel dependen
- $X_1, X_2, \dots, X_n$  adalah variabel independen
- $\beta_0$  disebut intercept, merupakan nilai Y ketika semua  $X$  bernilai 0.

- $\beta_1, \beta_2, \dots, \beta_n$  adalah koefisien regresi yang mengukur perubahan dalam Y terkait dengan perubahan satu unit dalam X.
- $\epsilon$  adalah error yang merepresentasikan ketidakpastian atau variabilitas yang tidak dapat dijelaskan oleh model.

## 2.2 Random Forest

*Random Forest* adalah salah satu teknik yang dapat meningkatkan tingkat akurasi saat menghasilkan atribut untuk setiap node dengan pendekatan acak [2]. *Random Forest* terdiri dari sekelompok pohon keputusan, di mana pohon-pohon keputusan tersebut bergabung untuk mengklasifikasikan data ke dalam kelas-kelas tertentu. Pohon keputusan dibangun dengan menentukan suatu node akar dan berakhir pada beberapa node daun untuk menghasilkan hasil akhir. Algoritma metode Random Forest terbagi menjadi dua. Pertama adalah pembentukan pohon 'k' untuk membuat hutan acak. Kedua yaitu melakukan prediksi dengan hutan acak (Random Forest) yang telah terbentuk.

Langkah-langkah dalam penerapan metode Random Forest antara lain:

1. Membuat data sampel dengan cara pengambilan acak dengan pengembalian dari dataset.
2. Gunakan sampel data untuk membangun pohon ke  $i$  ( $i=1, 2, 3, \dots, k$ )
3. Ulangi langkah 1 dan 2 sebanyak  $k$  kali

## 2.3 Decision Tree

*Decision Tree* adalah struktur data yang terbentuk oleh simpul (node) dan sisi-sisinya (edge). Dalam konteks pohon, simpul dapat diklasifikasikan menjadi tiga jenis, yaitu simpul akar (root/node), simpul

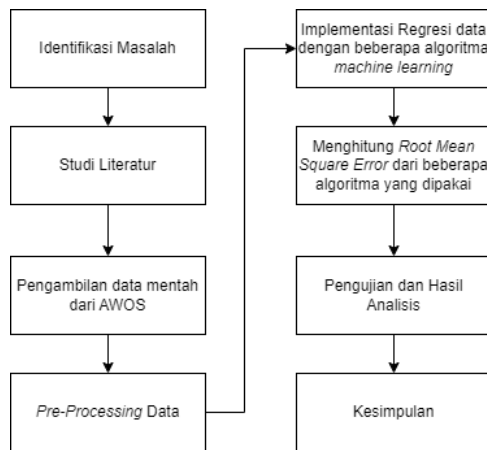
percabangan/internal (branch/internal node), dan simpul daun (leaf node) [3]. Pohon keputusan adalah representasi sederhana dari teknik klasifikasi untuk kelas-kelas yang berbeda, di mana atribut-atribut pada simpul internal dan simpul akar diberi nama, sementara sisi-sisi diberi label dengan nilai atribut yang mungkin. Selain itu, simpul daun ditandai dengan kelas-kelas yang berbeda [3].

## 2.4 K-Nearest Neighbors

*K-Nearest Neighbors* (KNN) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. Ini adalah algoritma pembelajaran non-parametrik dan malas yang bekerja dengan menemukan  $k$  tetangga terdekat ke titik data tertentu berdasarkan metrik jarak, dan kemudian mengklasifikasikan titik data berdasarkan kelas mayoritas tetangganya. Algoritmenya sederhana dan efektif, serta dapat digunakan dalam berbagai aplikasi seperti pemeringkatan kredit, sistem rekomendasi, dan klasifikasi gambar. KNN mempunyai beberapa kelebihan seperti mampu menangani data yang noise dan dataset yang besar, namun juga mempunyai beberapa kekurangan seperti perlunya menentukan nilai parameter  $k$  dan pembelajaran yang tidak jelas berdasarkan jarak.

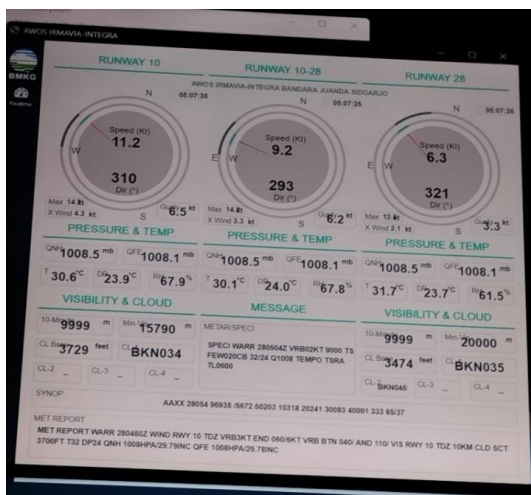
## 3. METODE PENELITIAN

Tahapan yang dilakukan untuk menyelesaikan penelitian ini ditunjukkan oleh Gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan penelitian dimulai dari proses identifikasi masalah yang dilakukan dengan berdiskusi kepada para pegawai Meteorologi BMKG Juanda. Kemudian melakukan studi literatur mengenai machine learning dengan algoritma yang dipakai yaitu *Linear Regression*, *Random Forest*, *Decision Tree*, *K-Nearest Neighbors*. Tahap selanjutnya yaitu pengambilan data mentah dari AWOS. Disini dilakukan secara *real-time* dan data dapat terlihat di aplikasi yang kemudian data-data tersebut dimasukkan ke dalam file dengan format .csv ditunjukkan pada Gambar 2.



Gambar 2. Software AWOS IRMAVIA

Lalu melakukan *pre-processing* pada data mentah AWOS. Disini menggunakan data AWOS yang digunakan yaitu 3 bulan dari

Desember 2022 sampai dengan Februari 2023. *Pre-processing* data dilakukan dengan tiga tahapan. Tahap pertama yaitu integrasi data. Integrasi data dilakukan untuk menggabungkan data dari beberapa tabel yang berbeda. Pengambilan data dari AWOS menjadi tabel dalam format .csv yaitu setiap satu hari sehingga memerlukan penggabungan data dari banyak hari di bulan Desember hingga bulan Februari. Penggabungan file csv ini menggunakan library pandas agar lebih cepat dan efisien. Berikut hasil pemrograman untuk menggabungkan beberapa file tersebut ditunjukkan pada Gambar 3.

```

import pandas as pd
import os

folderpath = 'PKL_data'

path10 = []
path1028 = []
path28 = []

# Include data for December 2022
for i in range(1, 32):
    path10.append(f'{folderpath}/2022/12/{i}/10/kal.csv')
    path1028.append(f'{folderpath}/2022/12/{i}/10-28/kal.csv')
    path28.append(f'{folderpath}/2022/12/{i}/28/kal.csv')
# Include data for January 2023
for i in range(1, 32):
    path10.append(f'{folderpath}/2023/1/{i}/10/kal.csv')
    path1028.append(f'{folderpath}/2023/1/{i}/10-28/kal.csv')
    path28.append(f'{folderpath}/2023/1/{i}/28/kal.csv')
# Include data for February 2023
for i in range(1, 29): # Assuming February has up to 28 days
    path10.append(f'{folderpath}/2023/2/{i}/10/kal.csv')
    path1028.append(f'{folderpath}/2023/2/{i}/10-28/kal.csv')
    path28.append(f'{folderpath}/2023/2/{i}/28/kal.csv')

merged_data = pd.DataFrame()
for file_path in path1028:
    if os.path.exists(file_path): # Check if the file exists
        df = pd.read_csv(file_path)
        merged_data = pd.concat([merged_data, df], ignore_index=True)
# Save the merged data to a new CSV file
merged_data.to_csv('merged_data_path1028.csv', index=False)
  
```

Gambar 3. Program menggabungkan file

Tahap kedua yaitu pembersihan data. Pembersihan data ini dilakukan dengan cara mengisi *missing value* yang terdapat pada kolom dengan metode *interpolasi*. Kemudian mengubah tipe data kolom yang memiliki tipe data *object* menjadi *float* agar dapat diolah. Kemudian menghapus kolom yang tidak diperlukan. Tahap ketiga yaitu transformasi data. Transformasi data dilakukan dengan cara mengubah nilai atribut tertentu yang sebelumnya berupa

huruf menjadi angka agar dapat mudah diolah. Suhu pada kolom tabel diberi nama AT.

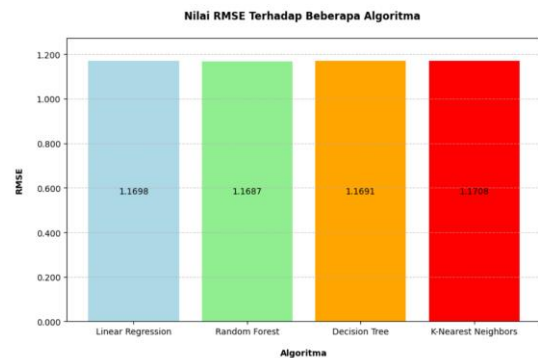
Setelah melakukan *pre-processing*, kemudian pengimplementasian regresi data dengan algoritma *Linear Regression*, *Random Forest*, *Decision Tree*, dan *K-Nearest Neighbors*. Lalu kolom suhu di runway 10 dan 28 sebagai fitur dan kolom suhu di runway 1028 sebagai target. Melatih data yaitu data pada bulan Desember 2022 dan bulan Januari 2023. Untuk bulan Februari 2023, digunakan untuk melakukan prediksi pada suhu runway 1028 dengan fitur runway 10 dan runway 28. Kemudian dibuatkan kolom baru untuk suhu dengan algoritma *Linear Regression*, suhu dengan algoritma *Random Forest*, suhu dengan algoritma *Decision Tree*, dan suhu dengan algoritma *K-Nearest Neighbors*. Hasil gambaran tabel ditunjukkan pada Gambar 4.

id	date	AT_10	AT_28	AT_1028	AT_1028_Linear_Regression	AT_1028_Random_Forest	AT_1028_Decision_Tree	AT_1028_K-Nearest_Neighbors
1	2023-02-01 00:00:15	26.1	26.3	26.3	26.15427073488843	26.12472517125689	26.12588444444444	26.1000000000
2	2023-02-01 00:01:15	26.1	26.3	26.3	26.15427073488843	26.12472517125689	26.12588444444444	26.1000000000
3	2023-02-01 00:02:15	26.1	26.4	26.4	26.19124030304803	26.15902144881982	26.159021447633	26.1399999999
4	2023-02-01 00:03:15	26.2	26.4	26.4	26.23819052444885	26.23819052444885	26.23819052444885	26.2000000000
5	2023-02-01 00:04:15	26.2	26.4	26.4	26.23819052444885	26.23819052444885	26.23819052444885	26.2000000000
6	2023-02-01 00:05:15	26.2	26.4	26.5	26.2611886232012	26.30311288881788	26.30311288881788	26.30311288881788
7	2023-02-01 00:06:15	26.3	26.4	26.5	26.30311288881788	26.41133873506248	26.41133873506248	26.3800000000
8	2023-02-01 00:07:15	26.3	26.4	26.5	26.30311288881788	26.41133873506248	26.41133873506248	26.3800000000
9	2023-02-01 00:08:15	26.3	26.4	26.5	26.30311288881788	26.41133873506248	26.41133873506248	26.3800000000
10	2023-02-01 00:09:15	26.3	26.4	26.5	26.30311288881788	26.41133873506248	26.41133873506248	26.3800000000
11	2023-02-01 00:10:15	26.3	26.4	26.5	26.30311288881788	26.41133873506248	26.41133873506248	26.3800000000
12	2023-02-01 00:11:15	26.4	26.4	26.5	26.428404080519105	26.45833847377835	26.45833847377835	26.45833847377835
13	2023-02-01 00:12:15	26.4	26.4	26.5	26.428404080519105	26.45833847377835	26.45833847377835	26.45833847377835
14	2023-02-01 00:13:15	26.4	26.5	26.5	26.428404080519105	26.45833847377835	26.45833847377835	26.45833847377835
15	2023-02-01 00:14:15	26.4	26.5	26.5	26.428404080519105	26.45833847377835	26.45833847377835	26.45833847377835
16	2023-02-01 00:15:15	26.4	26.5	26.5	26.428404080519105	26.45833847377835	26.45833847377835	26.45833847377835
17	2023-02-01 00:16:15	26.5	26.6	26.5	26.45833847377835	26.47181878018833	26.47181878018833	26.47181878018833
18	2023-02-01 00:17:15	26.5	26.6	26.5	26.45833847377835	26.47181878018833	26.47181878018833	26.47181878018833
19	2023-02-01 00:18:15	26.5	26.7	26.6	26.45833847377835	26.47181878018833	26.47181878018833	26.47181878018833
20	2023-02-01 00:19:15	26.5	26.7	26.6	26.45833847377835	26.47181878018833	26.47181878018833	26.47181878018833
21	2023-02-01 00:20:15	26.5	26.7	26.6	26.45833847377835	26.47181878018833	26.47181878018833	26.47181878018833
22	2023-02-01 00:21:15	26.6	26.8	26.7	26.55523404016833	26.55523404016833	26.55523404016833	26.55523404016833
23	2023-02-01 00:22:15	26.6	26.8	26.7	26.55523404016833	26.55523404016833	26.55523404016833	26.55523404016833
24	2023-02-01 00:23:15	26.6	26.8	26.7	26.55523404016833	26.55523404016833	26.55523404016833	26.55523404016833
25	2023-02-01 00:24:15	26.6	26.8	26.8	26.55523404016833	26.55523404016833	26.55523404016833	26.55523404016833

Gambar 4. Tabel Prediksi dengan beberapa Algoritma

#### 4. HASIL DAN PEMBAHASAN

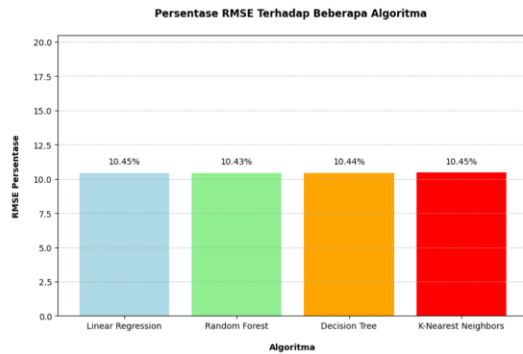
Berdasarkan proses sebelumnya yaitu menerapkan beberapa algoritma dalam memprediksi nilai suhu di runway 1028, kemudian akan melakukan proses analisis hasil berdasarkan beberapa rendah nilai *Root Mean Square Error* (RMSE) pada setiap algoritma yang digunakan. Hasil nilai *Root Mean Square Error* (RMSE) pada setiap algoritma ditunjukkan pada Gambar 5.



Gambar 5. Tabel RMSE beberapa algoritma

Pada tabel diatas, hasil perhitungan nilai RMSE untuk model algoritma *Linear Regression* sebesar 1,1698. Selanjutnya, fokus pada evaluasi menggunakan model algoritma *Random Forest*. Model ini memiliki sifat ensemble yang dapat meningkatkan ketepatan prediksi. Hasil perhitungan RMSE untuk *Random Forest* adalah 1,1687. Kemudian model algoritma *Decision Tree* memiliki karakteristik yang lebih sederhana dibandingkan dengan *Random Forest*. Hasil perhitungan RMSE untuk *Decision Tree* adalah 1,1691. Terakhir, evaluasi dilakukan terhadap metode *K-Nearest Neighbors* (KNN). KNN merupakan metode berbasis instansi yang menggunakan data latih terdekat untuk membuat prediksi. Hasil perhitungan RMSE untuk KNN adalah 1,1708. Kemudian jika ingin melihat persentase dari RMSE yaitu dengan cara nilai RMSE dibagi dengan rentang nilai. Rentang nilai didapatkan dari nilai terbesar yang ada pada kolom AT\_1028 dikurangi dengan nilai terendah yang ada pada kolom AT\_1028. Hasil dari persentase nilai RMSE dari setiap algoritma ditunjukkan pada Gambar 6.





Gambar 6. Tabel Persentase RMSE

Dari hasil tabel persentase RMSE diatas, didapatkan bahwa persentase RMSE yang didapatkan dari model algoritma *Linear Regression* yaitu 10,45%. Untuk persentase RMSE dari model algoritma *Random Forest* yaitu 10,43%. Kemudian untuk persentase RMSE dari model algoritma *Decision Tree* yaitu 10,44%. Terakhir, yaitu persentase RMSE dari model algoritma *K-Nearest Neighbors* yaitu 10,45%. Dari hasil nilai dan persentase RMSE dari beberapa algoritma yang dipakai maka model algoritma yang cocok digunakan untuk memprediksi nilai suhu di runway 1028 yaitu algoritma *Random Forest* dikarenakan memiliki nilai dan persentase RMSE yang paling rendah diantara ketiga algoritma lainnya.

## 5. KESIMPULAN

Penggunaan berbagai model algoritma machine learning dalam upaya memprediksi suatu data menawarkan pendekatan yang beragam untuk menangani masalah prediksi. Setiap model memiliki kelebihan dan kelemahan tertentu yang perlu dipertimbangkan sesuai dengan karakteristik data dan tujuan prediksi. Dalam penelitian ini, menerapkan beberapa model algoritma untuk memprediksi nilai suhu di runway 1028. Beberapa model algoritma yang digunakan yaitu *Linear Regression*, *Random Forest*, *Decision Tree*, *K-Nearest Neighbors*.

Hasil pengujian model algoritma tersebut dengan menggunakan *Root Mean Square Error* (RMSE). Dari hasil pengujian

tersebut didapatkan bahwa model algoritma yang cocok dalam memprediksi nilai suhu Runway 1028 yaitu model *Random Forest* dikarenakan memiliki nilai RMSE terendah dibandingkan tiga model algoritma lainnya.

## 6. Daftar Pustaka

- [1] Tri Sutanto, A., Abdhu Mukti, E., & Ade Setiawan, E. P. (2022). Rancang Bangun Portable Automated Weather Observing System (AWOS) Kategori 1 Berbasis Antarmuka Website Dan Android. *Jurnal Instrumentasi dan Teknologi Kebumian*, 1(1), 39-52. ISSN 2829-4734.
- [2] Amaliah, S., Nusrang, M., & Aswi. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konjiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 121-127. ISSN 2684-7590. doi:10.35580/variasiunm31
- [3] Nasrullah, A. H. (2021). Implementasi Algoritma Decision Tree untuk Klasifikasi Produk Laris. *Jurnal Ilmiah Ilmu Komputer*, 7(2), September 2021, ISSN 2442-451X (P), ISSN 2503-3832 (O).
- [4] Bachtiar, F. A., Syahputra, I. K., & Wicaksono, S. A. (2019). Perbandingan Algoritme Machine Learning untuk Memprediksi Pengambil Mata Kuliah. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 6(5), Oktober 2019, hlm. 543-548. doi:10.25126/jtiik.2019611755. p-ISSN: 2355-7699, e-ISSN: 2528-6579.