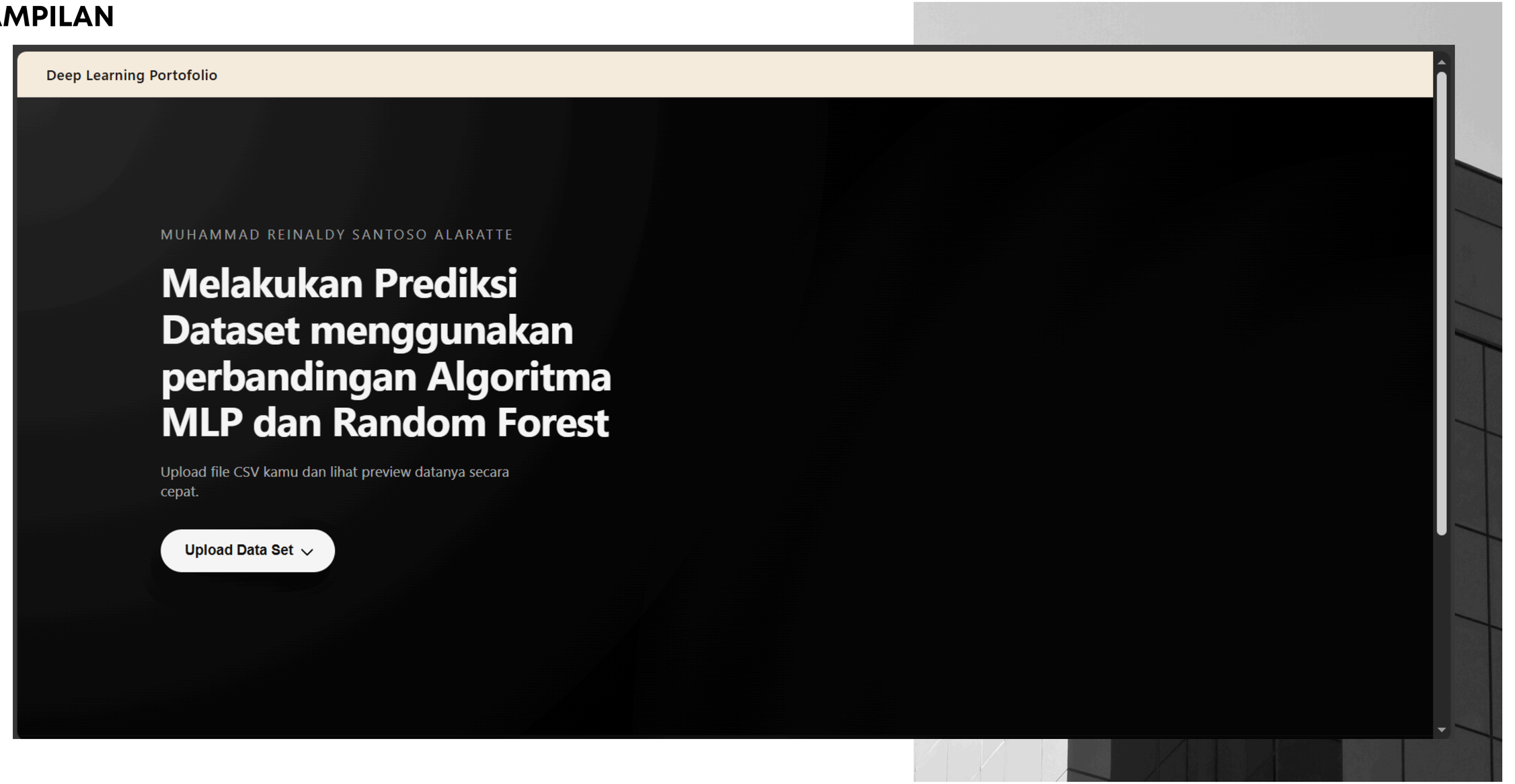
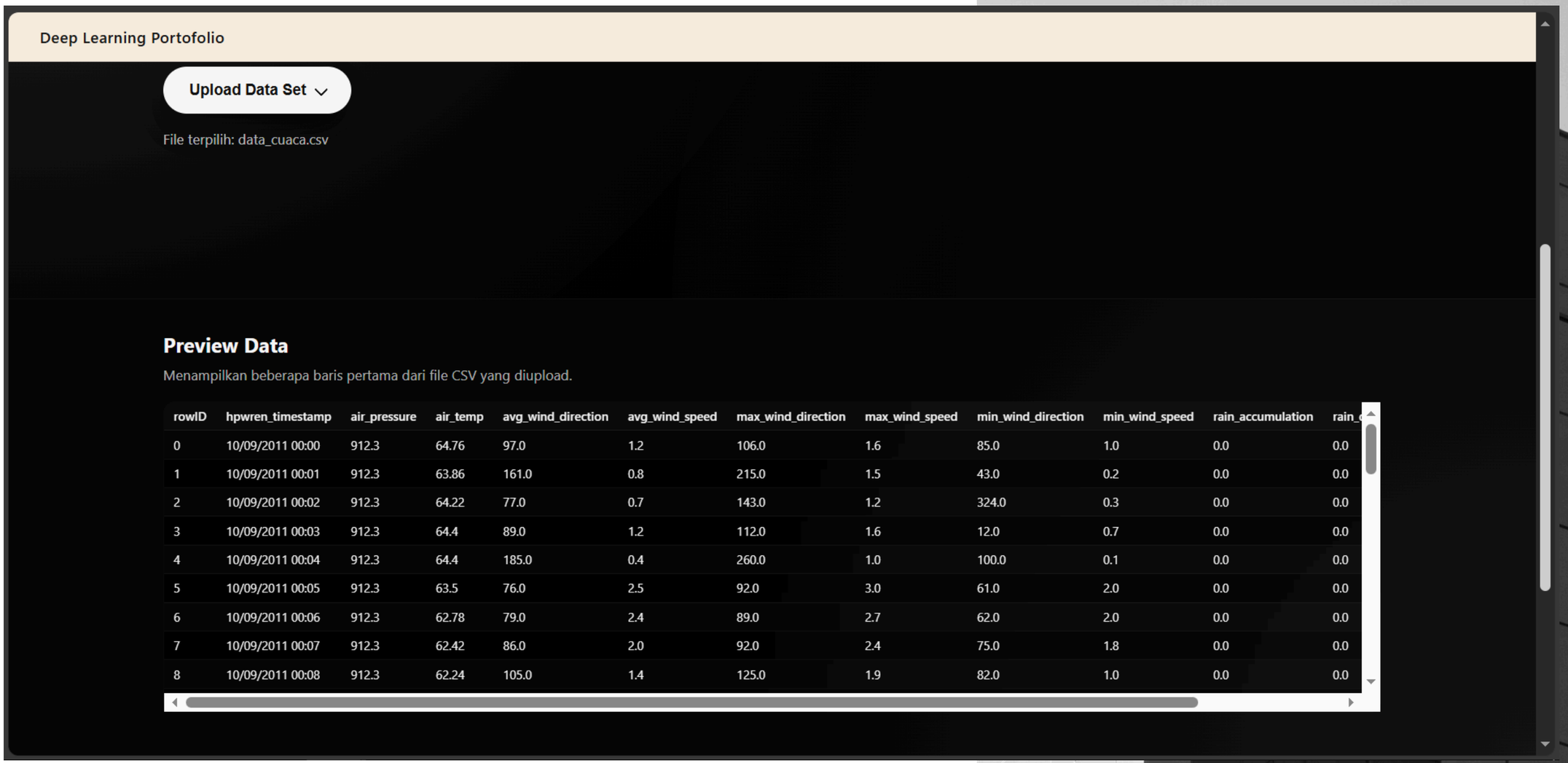
MELAKUKAN PREDIKSI DATASET MENGGUNAKAN PERBANDINGAN ALGORITMA MLP DAN RANDOM FOREST

GO

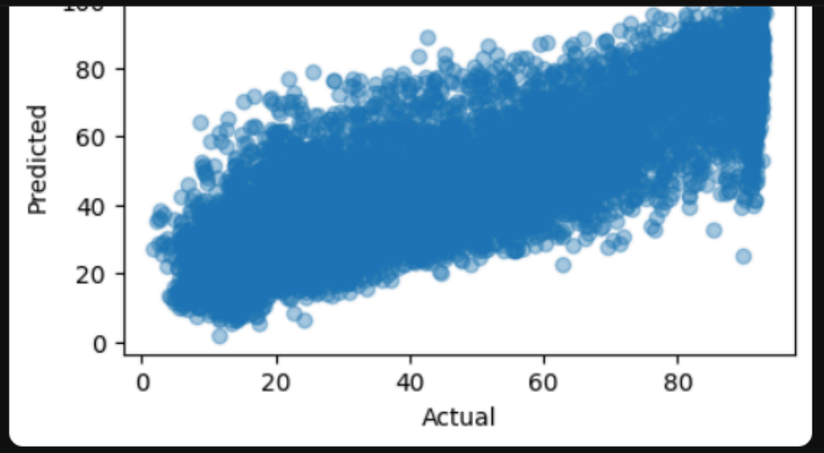
TAMPILAN


Deep Learning Portofolio
MUHAMMAD REINALDY SANTOSO ALARATTE
Melakukan Prediksi Dataset menggunakan perbandingan Algoritma MLP dan Random Forest
Upload file CSV kamu dan lihat preview datanya secara cepat.
Upload Data Set

TAMPILAN


Deep Learning Portofolio
Upload Data Set
File terpilih: data_cuaca.csv
Preview Data
Menampilkan beberapa baris pertama dari file CSV yang diupload.
rowID hpwren_timestamp air_pressure air_temp avg_wind_direction avg_wind_speed max_wind_direction max_wind_speed min_wind_direction min_wind_speed rain_accumulation rain_
0 10/09/2011 00:00 912.3 64.76 97.0 1.2 106.0 1.6 85.0 1.0 0.0 0.0
1 10/09/2011 00:01 912.3 63.86 161.0 0.8 215.0 1.5 43.0 0.2 0.0 0.0
2 10/09/2011 00:02 912.3 64.22 77.0 0.7 143.0 1.2 324.0 0.3 0.0 0.0
3 10/09/2011 00:03 912.3 64.4 89.0 1.2 112.0 1.6 12.0 0.7 0.0 0.0
4 10/09/2011 00:04 912.3 64.4 185.0 0.4 260.0 1.0 100.0 0.1 0.0 0.0
5 10/09/2011 00:05 912.3 63.5 76.0 2.5 92.0 3.0 61.0 2.0 0.0 0.0
6 10/09/2011 00:06 912.3 62.78 79.0 2.4 89.0 2.7 62.0 2.0 0.0 0.0
7 10/09/2011 00:07 912.3 62.42 86.0 2.0 92.0 2.4 75.0 1.8 0.0 0.0
8 10/09/2011 00:08 912.3 62.24 105.0 1.4 125.0 1.9 82.0 1.0 0.0 0.0

Hasil Algoritma MLP



Kesimpulan

Dari hasil pengujian, nilai MSE train sebesar 206.6091 dan MSE test sebesar 203.6829, dengan selisih sekitar 1.4% relatif terhadap MSE train. Perbedaan MSE antara data train dan test relatif kecil (<20%), sehingga model cukup mampu melakukan generalisasi ke data baru.

Berdasarkan kurva loss yang menurun dan sebaran titik pada grafik prediksi vs aktual, model sudah mampu menangkap pola umum dari data target. Namun performa masih dapat ditingkatkan dengan melakukan tuning hyperparameter (jumlah neuron, learning rate, jumlah epoch) atau penambahan teknik regularisasi agar error pada data test dapat diturunkan lebih lanjut.

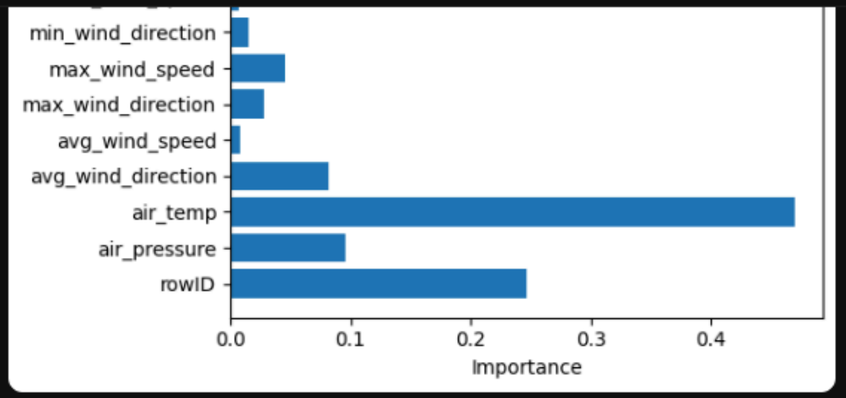
Tutup

```
C:\Windows\system32\cmd.exe - uvicorn  mlp_server:app --reload --host 0.0.0.0 --port 8000
Microsoft Windows [Version 10.0.19042.631]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\Users\rei alt>cd C:\\Users\rei alt\Videos\DL

C:\Users\rei alt\Videos\DL>uvicorn mlp_server:app --reload --host 0.0.0.0 --port 8000
←[32mINFO←[0m:     Will watch for changes in these directories: ['C:\\Users\\rei alt\\Videos\\DL']
←[32mINFO←[0m:     Uvicorn running on ←[1mhttp://0.0.0.0:8000←[0m (Press CTRL+C to quit)
←[32mINFO←[0m:     Started reloader process [←[36m←[1m932←[0m] using ←[36m←[1mStatReload←[0m
←[32mINFO←[0m:     Started server process [←[36m3328←[0m]
←[32mINFO←[0m:     Waiting for application startup.
←[32mINFO←[0m:     Application startup complete.

=== REQUEST MLP DITERIMA ===
Ukuran file diterima: 82612977
CSV terbaca. Shape asli: (1048575, 13)
Kolom numerik: (1048575, 12)
Setelah hapus NaN awal: (1048341, 11)
Sampling 5.0% data → (52417, 12)
Setelah hapus NaN setelah sampling: (52417, 11)
Split train-test...
Train: (41933, 11) Test: (10484, 11)
Mulai training MLP...
Iteration 1, loss = 990.98847690
Iteration 2, loss = 223.50922102
Iteration 3, loss = 172.94091651
Iteration 4, loss = 154.24621162
Iteration 5, loss = 144.75099630
Iteration 6, loss = 139.01874405
Iteration 7, loss = 135.13359916
```

C:\Windows\system32\cmd.exe - uvicorn rf_server:app --reload --host 0.0.0.0 --port 8001

```
Microsoft Windows [Version 10.0.19042.631]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\Users\rei alt>cd C:\\Users\rei alt\Videos\DL

C:\Users\rei alt\Videos\DL>uvicorn rf_server:app --reload --host 0.0.0.0 --port 8001
←[32mINFO←[0m:     Will watch for changes in these directories: ['C:\\Users\\rei alt\\Videos\\DL']
←[32mINFO←[0m:     Uvicorn running on ←[1mhttp://0.0.0.0:8001←[0m (Press CTRL+C to quit)
←[32mINFO←[0m:     Started reloader process [←[36m←[1m8476←[0m] using ←[36m←[1mStatReload←[0m
←[32mINFO←[0m:     Started server process [←[36m3944←[0m]
←[32mINFO←[0m:     Waiting for application startup.
←[32mINFO←[0m:     Application startup complete.

=== REQUEST RANDOM FOREST DITERIMA ===
CSV terbaca: (1048575, 13)
Kolom numerik: (1048575, 12)
Setelah hapus NaN: (1048341, 11)
Setelah sampling: (52417, 11)
Train: (41933, 11) Test: (10484, 11)
←[32mINFO←[0m:     127.0.0.1:53177 - "←[1mPOST /rf HTTP/1.1←[0m" ←[32m200 OK←[0m

=== REQUEST RANDOM FOREST DITERIMA ===
CSV terbaca: (1048575, 13)
Kolom numerik: (1048575, 12)
Setelah hapus NaN: (1048341, 11)
Setelah sampling: (52417, 11)
Train: (41933, 11) Test: (10484, 11)
←[32mINFO←[0m:     127.0.0.1:53241 - "←[1mPOST /rf HTTP/1.1←[0m" ←[32m200 OK←[0m
Iteration 6, loss = 139.61874465
Iteration 7, loss = 135.13359916
```

Hasil Algoritma Random Forest



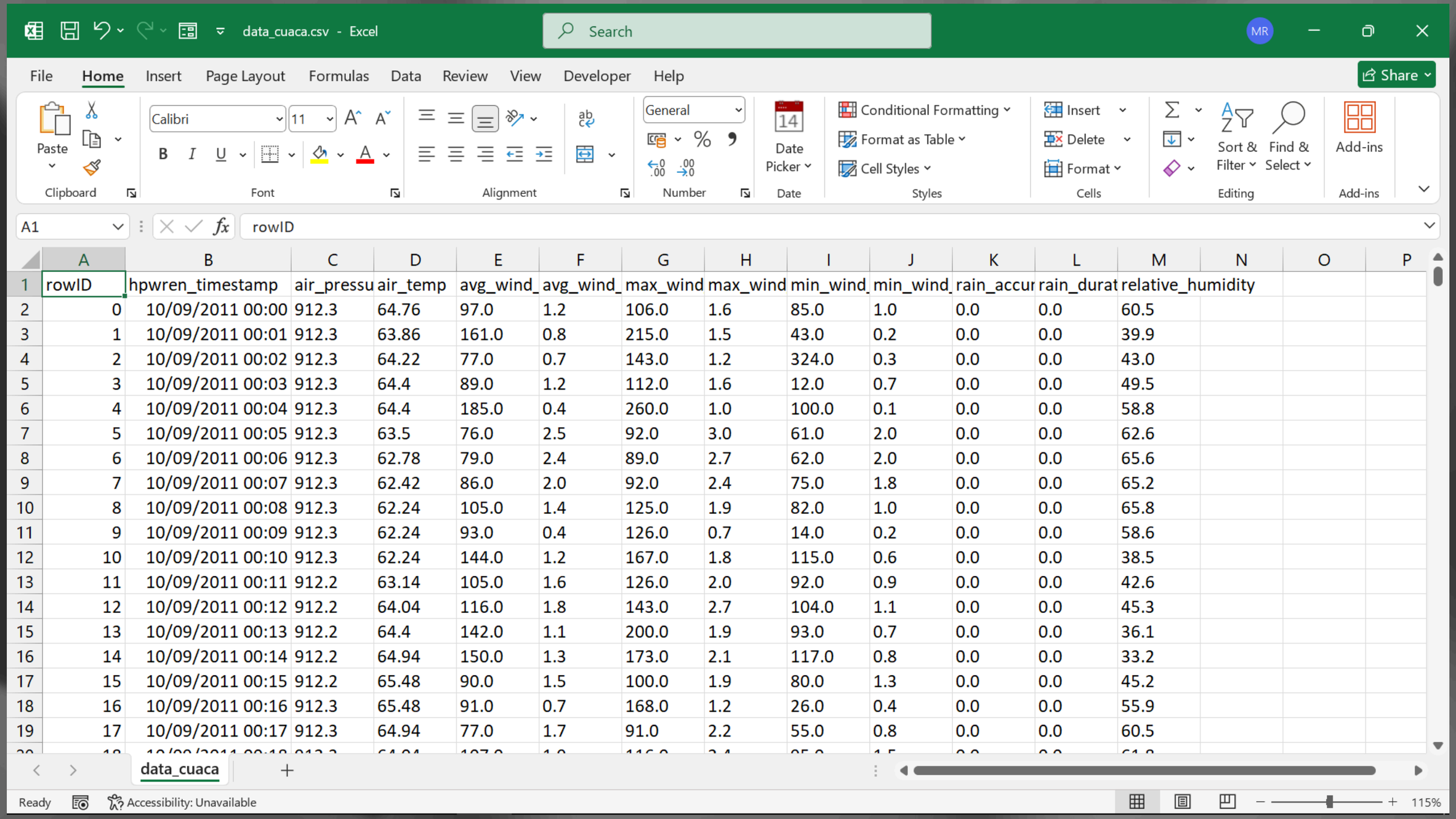
Kesimpulan

Hasil evaluasi menunjukkan MSE train 4.3439 dan MSE test 29.9955 (RMSE test 5.4768). Selisih error sebesar 590.5% mengindikasikan bahwa Model cenderung overfitting atau belum stabil..

Dari hasil perhitungan feature importance, variabel yang paling berpengaruh dalam prediksi adalah 'air_temp'. Hal ini menunjukkan bahwa variabel tersebut memiliki kontribusi terbesar terhadap target yang diprediksi.

Secara keseluruhan, Random Forest mampu menangkap pola non-linear pada data cuaca dan memberikan hasil prediksi yang cukup akurat. Performa dapat ditingkatkan dengan tuning jumlah tree, max_depth, atau penyesuaian sampling data.

Tutup



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	rowID	hpwren_timestamp	air_pressu	air_temp	avg_wind_	avg_wind_	max_wind	max_wind	min_wind_	min_wind_	rain_accur	rain_durat	relative_humidity			
2	0	10/09/2011 00:00	912.3	64.76	97.0	1.2	106.0	1.6	85.0	1.0	0.0	0.0	60.5			
3	1	10/09/2011 00:01	912.3	63.86	161.0	0.8	215.0	1.5	43.0	0.2	0.0	0.0	39.9			
4	2	10/09/2011 00:02	912.3	64.22	77.0	0.7	143.0	1.2	324.0	0.3	0.0	0.0	43.0			
5	3	10/09/2011 00:03	912.3	64.4	89.0	1.2	112.0	1.6	12.0	0.7	0.0	0.0	49.5			
6	4	10/09/2011 00:04	912.3	64.4	185.0	0.4	260.0	1.0	100.0	0.1	0.0	0.0	58.8			
7	5	10/09/2011 00:05	912.3	63.5	76.0	2.5	92.0	3.0	61.0	2.0	0.0	0.0	62.6			
8	6	10/09/2011 00:06	912.3	62.78	79.0	2.4	89.0	2.7	62.0	2.0	0.0	0.0	65.6			
9	7	10/09/2011 00:07	912.3	62.42	86.0	2.0	92.0	2.4	75.0	1.8	0.0	0.0	65.2			
10	8	10/09/2011 00:08	912.3	62.24	105.0	1.4	125.0	1.9	82.0	1.0	0.0	0.0	65.8			
11	9	10/09/2011 00:09	912.3	62.24	93.0	0.4	126.0	0.7	14.0	0.2	0.0	0.0	58.6			
12	10	10/09/2011 00:10	912.3	62.24	144.0	1.2	167.0	1.8	115.0	0.6	0.0	0.0	38.5			
13	11	10/09/2011 00:11	912.2	63.14	105.0	1.6	126.0	2.0	92.0	0.9	0.0	0.0	42.6			
14	12	10/09/2011 00:12	912.2	64.04	116.0	1.8	143.0	2.7	104.0	1.1	0.0	0.0	45.3			
15	13	10/09/2011 00:13	912.2	64.4	142.0	1.1	200.0	1.9	93.0	0.7	0.0	0.0	36.1			
16	14	10/09/2011 00:14	912.2	64.94	150.0	1.3	173.0	2.1	117.0	0.8	0.0	0.0	33.2			
17	15	10/09/2011 00:15	912.2	65.48	90.0	1.5	100.0	1.9	80.0	1.3	0.0	0.0	45.2			
18	16	10/09/2011 00:16	912.3	65.48	91.0	0.7	168.0	1.2	26.0	0.4	0.0	0.0	55.9			
19	17	10/09/2011 00:17	912.3	64.94	77.0	1.7	91.0	2.2	55.0	0.8	0.0	0.0	60.5			

Ringkasan Hasil MLP & Random Forest

- MSE test: **29.9955**
- RMSE train: **2.0842**
- RMSE test: **5.4768**
- Fitur paling berpengaruh (RF): **air_temp**

Hasil evaluasi menunjukkan MSE train 4.3439 dan MSE test 29.9955 (RMSE test 5.4768). Selisih error sebesar 590.5% mengindikasikan bahwa Model cenderung overfitting atau belum stabil..

Dari hasil perhitungan feature importance, variabel yang paling berpengaruh dalam prediksi adalah 'air_temp'. Hal ini menunjukkan bahwa variabel tersebut memiliki kontribusi terbesar terhadap target yang diprediksi.

Secara keseluruhan, Random Forest mampu menangkap pola non-linear pada data cuaca dan memberikan hasil prediksi yang cukup akurat. Performa dapat ditingkatkan dengan tuning jumlah tree, max_depth, atau penyesuaian sampling data.

Perbandingan MLP vs Random Forest

Pada dataset ini, Random Forest menghasilkan nilai MSE test yang lebih kecil dibandingkan MLP, sehingga secara umum Random Forest memberikan performa prediksi yang lebih baik.

Secara umum, perbandingan error (MSE/ RMSE) dan karakteristik model ini dapat dijadikan dasar untuk menjelaskan ke dosen mana algoritma yang lebih sesuai untuk kasus prediksi kelembapan udara pada dataset cuaca yang digunakan, serta bagaimana trade-off antara kompleksitas model dan akurasi.