

Tugas Besar Tahap Kedua Machine Learning

Clustering

oleh:

Muhammad Hafidh Raditya (NIM 1301184079)

Edgar Vigo (NIM 1301180149)

IF-42-03



**Program Studi S1 Informatika
Fakultas Informatika
Universitas Telkom
Bandung
2021**

A. Formulasi Masalah

Pada tugas besar tahap kedua ini kami mendapatkan dataset nomor 2 yang berisi tentang berbagai macam atribut yang memengaruhi apakah pada hari ini dan hari esok akan turun salju atau tidak.

Tugas kami adalah untuk membuat model klasifikasi dan melakukan eksperimen terhadap dua atribut, yaitu atribut BersaljuHariIni untuk memprediksi apakah hari ini bersalju atau tidak, dan atribut BersaljuBesok untuk memprediksi apakah besok bersalju atau tidak. Yang akan kami pertimbangkan dalam eksperimen adalah metode data pre-processing apa dan model klasifikasi apa yang menghasilkan nilai akurasi yang paling baik. Detail dari eksperimen akan kami jelaskan pada bagian berikutnya.

Masalah yang dapat diselesaikan menggunakan hasil dari tugas besar tahap dua ini adalah masalah tentang prakiraan salju, apakah bersalju atau tidak. Hasil prediksi dapat dihasilkan dari model klasifikasi yang sudah dibuat.

B. Data Pre-Processing

Pada tugas besar tahap kedua ini kami menggunakan dua dataset yang sudah diberikan oleh dosen, yaitu dataset train dan dataset test. Dataset train digunakan untuk membuat model klasifikasi, sedangkan dataset test dilakukan untuk melakukan prediksi dan menghitung nilai akurasi. Pre-processing yang kami lakukan kepada kedua jenis dataset tersebut akan sama persis, maka dari itu kami hanya akan menyertakan screenshot untuk pre-processing pada dataset train untuk mempersingkat isi laporan.

Langkah pertama yang kami lakukan adalah melakukan drop pada column/atribut yang tidak diperlukan dalam proses klasifikasi. Atribut yang kami pilih adalah atribut Id, Tanggal, dan KodeLokasi.

```
unscaled_datatrain = datatrain.drop(columns=['id', 'Tanggal', 'KodeLokasi'])
```

Selanjutnya yang kami lakukan adalah mengecek apakah terdapat null value pada dataset. Untuk menjaga distribusi dataset, kami tidak melakukan drop langsung pada null value, melainkan atribut yang tergolong numerical attribute, null valuenya akan

kami ganti dengan nilai mean atau rata-rata dari atribut tersebut. Untuk atribut yang tergolong categorical attribute, null valuenya akan kami ganti dengan nilai modus dari atribut tersebut.

```
dataset.isnull().sum()
```

```
id          0
Tanggal     0
KodeLokasi  0
SuhuMin     1122
SuhuMax     929
Hujan       2431
Penguapan   47024
SinarMatahari 52379
ArahAnginTerkencang 7744
KecepatanAnginTerkencang 7696
ArahAngin9am 7923
ArahAngin3pm 3197
KecepatanAngin9am 1353
KecepatanAngin3pm 2303
Kelembaban9am 2002
Kelembaban3pm 3374
Tekanan9am  11327
Tekanan3pm  11308
Awan9am     41844
Awan3pm     44471
Suhu9am     1340
Suhu3pm     2698
BersaljuHariIni 2431
BersaljuBesok 2431
dtype: int64
```

```
unscaled_datatrain["SuhuMin"].fillna(value=unscaled_datatrain["SuhuMin"].mean(), inplace = True)
unscaled_datatrain["SuhuMax"].fillna(value=unscaled_datatrain["SuhuMax"].mean(), inplace = True)
unscaled_datatrain["Hujan"].fillna(value=unscaled_datatrain["Hujan"].mean(), inplace = True)
unscaled_datatrain["Penguapan"].fillna(value=unscaled_datatrain["Penguapan"].mean(), inplace = True)
unscaled_datatrain["SinarMatahari"].fillna(value=unscaled_datatrain["SinarMatahari"].mean(), inplace = True)
unscaled_datatrain["ArahAnginTerkencang"].fillna(value=unscaled_datatrain["ArahAnginTerkencang"].mode()[0], inplace = True)#Categorical
unscaled_datatrain["KecepatanAnginTerkencang"].fillna(value=unscaled_datatrain["KecepatanAnginTerkencang"].mean(), inplace = True)
unscaled_datatrain["ArahAngin9am"].fillna(value=unscaled_datatrain["ArahAngin9am"].mode()[0], inplace = True)#categorical
unscaled_datatrain["ArahAngin3pm"].fillna(value=unscaled_datatrain["ArahAngin3pm"].mode()[0], inplace = True)#categorical
unscaled_datatrain["KecepatanAngin9am"].fillna(value=unscaled_datatrain["KecepatanAngin9am"].mean(), inplace = True)
unscaled_datatrain["KecepatanAngin3pm"].fillna(value=unscaled_datatrain["KecepatanAngin3pm"].mean(), inplace = True)
unscaled_datatrain["Kelembaban9am"].fillna(value=unscaled_datatrain["Kelembaban9am"].mean(), inplace = True)
unscaled_datatrain["Kelembaban3pm"].fillna(value=unscaled_datatrain["Kelembaban3pm"].mean(), inplace = True)
unscaled_datatrain["Tekanan9am"].fillna(value=unscaled_datatrain["Tekanan9am"].mean(), inplace = True)
unscaled_datatrain["Tekanan3pm"].fillna(value=unscaled_datatrain["Tekanan3pm"].mean(), inplace = True)
unscaled_datatrain["Awan9am"].fillna(value=unscaled_datatrain["Awan9am"].mean(), inplace = True)
unscaled_datatrain["Awan3pm"].fillna(value=unscaled_datatrain["Awan3pm"].mean(), inplace = True)
unscaled_datatrain["Suhu9am"].fillna(value=unscaled_datatrain["Suhu9am"].mean(), inplace = True)
unscaled_datatrain["Suhu3pm"].fillna(value=unscaled_datatrain["Suhu3pm"].mean(), inplace = True)
unscaled_datatrain["BersaljuHariIni"].fillna(value=unscaled_datatrain["BersaljuHariIni"].mode()[0], inplace = True)#Categorical
unscaled_datatrain["BersaljuBesok"].fillna(value=unscaled_datatrain["BersaljuBesok"].mode()[0], inplace = True)#categorical
unscaled_datatrain
```

```
unscaled_datatrain.isnull().sum()
```

```
SuhuMin          0
SuhuMax          0
Hujan            0
Penguapan        0
SinarMatahari    0
ArahAnginTerkencang 0
KecepatanAnginTerkencang 0
ArahAngin9am     0
ArahAngin3pm     0
KecepatanAngin9am 0
KecepatanAngin3pm 0
Kelembaban9am    0
Kelembaban3pm    0
Tekanan9am       0
Tekanan3pm       0
Awan9am          0
Awan3pm          0
Suhu9am          0
Suhu3pm          0
BersaljuHariIni  0
BersaljuBesok    0
dtype: int64
```

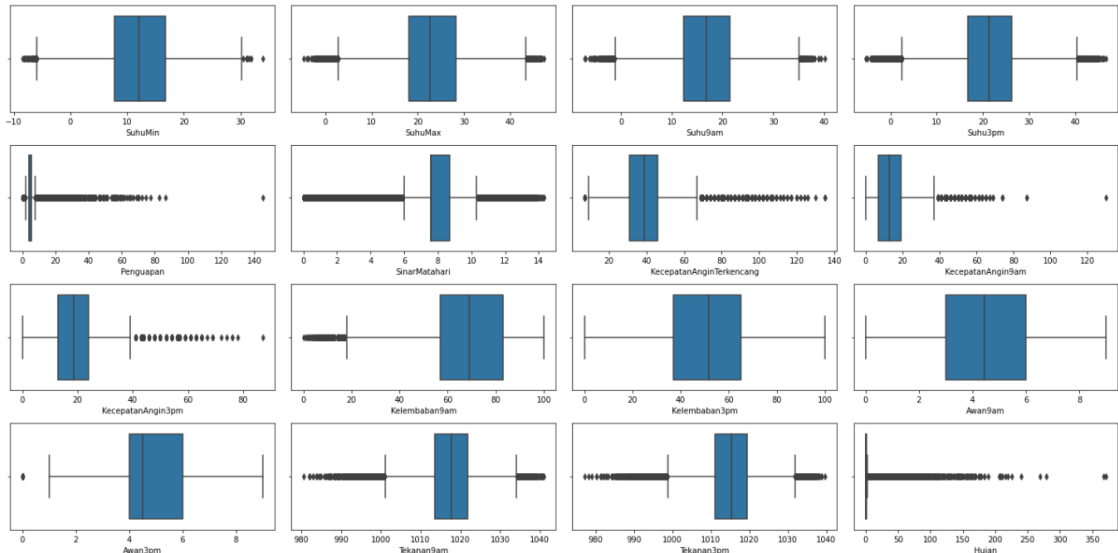
Setelah dataset bersih dari record yang berisi null value, kami akan melakukan konversi atribut yang bernilai kategorikal, menjadi atribut yang bernilai numerikal yang bertujuan untuk mempermudah proses perbandingan.

```
categorical = unscaled_datatrain.dtypes==object
categorical_cols = unscaled_datatrain.columns[categorical].tolist()
unscaled_datatrain[categorical_cols] = unscaled_datatrain[categorical_cols].apply(lambda col: LabelEncoder().fit_transform(col))
unscaled_datatrain[categorical_cols].head()
unscaled_datatrain
```







	SuhuMin	SuhuMax	Hujan	Penguapan	SinarMatahari	ArahAnginTerkencang	KecepatanAnginTerkencang	ArahAngin9am	ArahAngin3pm	KecepatanAn
0	10.4	15.5	4.8	5.46244	7.599527	15	24.000000	3	15	
1	9.0	17.0	8.0	2.60000	7.400000	13	40.032002	12	14	
2	18.2	32.0	0.0	5.46244	7.599527	2	44.000000	9	9	
3	7.3	24.5	0.0	8.40000	10.400000	11	54.000000	3	12	
4	5.9	20.3	0.0	3.60000	12.600000	3	37.000000	6	2	
...
109090	20.1	23.7	0.0	7.20000	8.900000	2	43.000000	9	2	
109091	15.7	25.2	0.0	5.46244	7.599527	10	37.000000	10	0	
109092	7.5	20.4	1.6	5.46244	7.599527	7	33.000000	3	7	
109093	10.8	29.8	0.0	7.80000	11.200000	0	48.000000	2	9	
109094	12.3	27.4	9.0	5.46244	7.599527	14	35.000000	5	4	

109095 rows × 21 columns

Selanjutnya kami mengecek data yang tergolong outlier menggunakan boxplot.



Data yang tergolong outlier ini tidak akan kami drop semua selama nilainya masih dalam skala yang normal.

Scale	Wind speed estimate ^[6]		Potential damage ^[6]	
	mph	km/h		
F0	40–72	64–116	Light damage. Some damage to chimneys; branches broken off trees; shallow-rooted trees pushed over; sign boards damaged.	
F1	73–112	117–180	Moderate damage. The lower limit is the beginning of hurricane wind speed; peels surface off roofs; mobile homes pushed off foundations or overturned; moving vehicles pushed off the roads; attached garages may be destroyed.	
F2	113–157	181–253	Significant damage. Roofs torn off frame houses; mobile homes demolished; boxcars overturned; large trees snapped or uprooted; high-rise windows broken and blown in; light-object missiles generated.	
F3	158–206	254–332	Severe damage. Roofs and some walls torn off well-constructed houses; trains overturned; most trees in forests uprooted; heavy cars lifted off the ground and thrown.	
F4	207–260	333–418	Devastating damage. Well-constructed houses leveled; structures with weak foundations blown away some distance; cars thrown and large missiles generated.	
F5	261–318	419–512	Incredible damage. Strong frame houses lifted off foundations and carried considerable distances to disintegrate; automobile-sized missiles fly through the air farther than 100 meters (110 yards); trees debarked; steel-reinforced concrete structures badly damaged and skyscrapers toppled	

Setelah melakukan penelitian, kami menemukan tentang Fujita Scale yang menjelaskan tentang skala minimum sampai maksimum dari angin topan. Fujita Scale tersebut menjelaskan kalau batas minimum kecepatan angin agar bisa disebut sebagai angin topan adalah sebesar 40 mph. Karena dataset ini menggunakan satuan kmh untuk atribut kecepatan angin, maka 40 mph sama dengan 64 kmh. Maka dari itu kami memutuskan untuk menetapkan jika kecepatan angin sudah diatas 64 kmh, maka data

tersebut sudah tidak dalam skala yang normal karena sudah dapat dikategorikan sebagai bencana angin topan dan akan kami drop dari dataset untuk atribut KecepatanAnginTerkencang, KecepatanAngin9am, dan KecepatanAngin3pm.

Tahapan terakhir yang kami lakukan pada data pre-processing adalah membuat dataset menjadi dua jenis, yaitu dataset yang dilakukan minmaxscaler dan yang tidak dilakukan minmaxscaler. Dua jenis dataset ini yang nantinya akan menjadi bahan eksperimen kami.

Dataset yang tidak dilakukan minmaxscaler

	SuhuMin	SuhuMax	Hujan	Penguapan	SinarMatahari	ArahAnginTerkencang	KecepatanAnginTerkencang	ArahAngin9am	ArahAngin3pm	KecepatanAn
0	10.4	15.5	4.8	5.46244	7.599527	15	24.000000	3	15	
1	9.0	17.0	8.0	2.60000	7.400000	13	40.032002	12	14	
2	18.2	32.0	0.0	5.46244	7.599527	2	44.000000	9	9	
3	7.3	24.5	0.0	8.40000	10.400000	11	54.000000	3	12	
4	5.9	20.3	0.0	3.60000	12.600000	3	37.000000	6	2	
...
109090	20.1	23.7	0.0	7.20000	8.900000	2	43.000000	9	2	
109091	15.7	25.2	0.0	5.46244	7.599527	10	37.000000	10	0	
109092	7.5	20.4	1.6	5.46244	7.599527	7	33.000000	3	7	
109093	10.8	29.8	0.0	7.80000	11.200000	0	48.000000	2	9	
109094	12.3	27.4	9.0	5.46244	7.599527	14	35.000000	5	4	

103374 rows × 21 columns

Dataset yang dilakukan minmaxscaler

	SuhuMin	SuhuMax	Hujan	Penguapan	SinarMatahari	ArahAnginTerkencang	KecepatanAnginTerkencang	ArahAngin9am	ArahAngin3pm	Kecepatan
0	0.445755	0.376471	0.012938	0.037672	0.531435	1.000000	0.303571	0.200000	1.000000	
1	0.412736	0.405882	0.021563	0.017931	0.517483	0.866667	0.589857	0.800000	0.933333	
2	0.629717	0.700000	0.000000	0.037672	0.531435	0.133333	0.660714	0.600000	0.600000	
3	0.372642	0.552941	0.000000	0.057931	0.727273	0.733333	0.839286	0.200000	0.800000	
4	0.339623	0.470588	0.000000	0.024828	0.881119	0.200000	0.535714	0.400000	0.133333	
...
103369	0.674528	0.537255	0.000000	0.049655	0.622378	0.133333	0.642857	0.600000	0.133333	
103370	0.570755	0.566667	0.000000	0.037672	0.531435	0.666667	0.535714	0.666667	0.000000	
103371	0.377358	0.472549	0.004313	0.037672	0.531435	0.466667	0.464286	0.200000	0.466667	
103372	0.455189	0.656863	0.000000	0.053793	0.783217	0.000000	0.732143	0.133333	0.600000	
103373	0.490566	0.609804	0.024259	0.037672	0.531435	0.933333	0.500000	0.333333	0.266667	

103374 rows × 21 columns

Terakhir, kami simpan kedua jenis dataset tersebut kedalam file csv.

C. Proses Classification

Pada tahap ini, kami menggunakan dua jenis model klasifikasi, yaitu algoritma Naive Bayes dan algoritma ID3. Kedua jenis algoritma ini juga akan menjadi bahan eksperimen kami selain dua jenis dataset yang sudah kami buat tadi.

Karena diperbolehkan untuk menggunakan library, maka kami membuat model klasifikasinya dengan memanfaatkan library yang sudah ada.

Algoritma Naive Bayes

```
In [218]: # Import Gaussian Naive Bayes model
from sklearn.naive_bayes import GaussianNB

#Buat model Gaussian Classifier
model = GaussianNB()

#Latih model menggunakan data latih
model.fit(x_scaled_train, y_scaled_train)

#Prediksi respons untuk dataset uji
y_pred_NB = model.predict(x_scaled_test)
y_pred_NB

c:\users\msi\appdata\local\programs\python\python38\lib\site
column-vector y was passed when a 1d array was expected. Ple
().
return f(*args, **kwargs)
```

Out[218]: array([0., 0., 0., ..., 0., 0., 0.])

```
In [219]: from sklearn import metrics
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report

y_true = y_scaled_test
results = confusion_matrix(y_true, y_pred_NB)
print('Confusion Matrix :')
print(results)
print('Accuracy Score :',accuracy_score(y_true, y_pred_NB))
print('Report : ')
print(classification_report(y_true, y_pred_NB))
```

Algoritma ID3

```
In [220]: from sklearn.tree import DecisionTreeClassifier
model_dt = DecisionTreeClassifier(criterion='entropy')
model_dt.fit(x_scaled_train, y_scaled_train)
y_pred_ID3 = model_dt.predict(x_scaled_test)
y_pred_ID3
```

Out[220]: array([0., 0., 0., ..., 0., 0., 0.])

```
In [221]: y_true = y_scaled_test
results = confusion_matrix(y_true, y_pred_ID3)
print('Confusion Matrix :')
print(results)
print('Accuracy Score :',accuracy_score(y_true, y_pred_ID3))
print('Report : ')
print(classification_report(y_true, y_pred_ID3))
```

D. Eksperimen dan Analisis

Sebelum masuk ke rincian eksperimen, perlu kami beritahu bahwa pada eksperimen kali ini nilai 0.0 menunjukkan nilai numerikal dari nilai kategorikal “Tidak Bersalju”. Sedangkan nilai 1.0 menunjukkan nilai numerikal dari nilai kategorikal “Bersalju”.

1. Prediksi atribut BersaljuHariIni

- Dataset yang sudah dilakukan minmaxscaler

➤ Algoritma Naive Bayes

Confusion Matrix :

```
[[12663  954]
```

```
 [   27 3598]]
```

Accuracy Score : 0.9431040482542629

Report :

	precision	recall	f1-score	support
0.0	1.00	0.93	0.96	13617
1.0	0.79	0.99	0.88	3625
accuracy			0.94	17242
macro avg	0.89	0.96	0.92	17242
weighted avg	0.95	0.94	0.95	17242

Dengan menggunakan algoritma Naive Bayes dihasilkan akurasi keseluruhan sebesar 94.31040482542629%. Jika mengacu pada nilai precision, prediksi tidak bersalju memiliki tingkat akurasi sebesar 100%, sedangkan untuk prediksi bersalju memiliki tingkat akurasi sebesar 79%.

➤ Algoritma ID3

Confusion Matrix :

```
[[12817  800]
```

```
[    0 3625]]
```

Accuracy Score : 0.9536016703398678

Report :

	precision	recall	f1-score	support
0.0	1.00	0.94	0.97	13617
1.0	0.82	1.00	0.90	3625
accuracy			0.95	17242
macro avg	0.91	0.97	0.94	17242
weighted avg	0.96	0.95	0.96	17242

Dengan menggunakan algoritma ID3 dihasilkan akurasi keseluruhan sebesar 95.36016703398678%. Jika mengacu pada nilai precision, prediksi tidak bersalju memiliki tingkat akurasi sebesar 100%, sedangkan untuk prediksi bersalju memiliki tingkat akurasi sebesar 82%.

- Dataset yang belum dilakukan minmaxscaler

➤ Algoritma Naive Bayes

Confusion Matrix :

```
[[12816  801]
```

```
[  213 3412]]
```

Accuracy Score : 0.9411901171557824

Report :

	precision	recall	f1-score	support
0	0.98	0.94	0.96	13617
1	0.81	0.94	0.87	3625
accuracy			0.94	17242
macro avg	0.90	0.94	0.92	17242
weighted avg	0.95	0.94	0.94	17242

Dengan menggunakan algoritma Naive Bayes dihasilkan akurasi keseluruhan sebesar 94.11901171557824%. Jika mengacu pada nilai precision, prediksi tidak bersalju memiliki tingkat akurasi sebesar 98%, sedangkan untuk prediksi bersalju memiliki tingkat akurasi sebesar 81%.

➤ Algoritma ID3

Confusion Matrix :

```
[[13255  362]
```

```
[    0 3625]]
```

Accuracy Score : 0.9790047558287902

Report :

	precision	recall	f1-score	support
0	1.00	0.97	0.99	13617
1	0.91	1.00	0.95	3625
accuracy			0.98	17242
macro avg	0.95	0.99	0.97	17242
weighted avg	0.98	0.98	0.98	17242

Dengan menggunakan algoritma ID3 dihasilkan akurasi keseluruhan sebesar 97.90047558287902%. Jika mengacu pada nilai precision, prediksi tidak bersalju memiliki tingkat akurasi sebesar 100%, sedangkan untuk prediksi bersalju juga memiliki tingkat akurasi sebesar 91%.

Tabel dibawah ini merupakan hasil dari tingkat akurasi dari 6 percobaan yang sudah kami lakukan diatas.

Akurasi Atribut BersaljuHariIni

	Naive Bayes	ID3	Rata-rata
Dataset yang sudah dilakukan minmaxscaler	94.31%	95.36%	94.83%
Dataset yang belum dilakukan minmaxscaler	94.11%	97.90%	96%
Rata-rata	94.21%	96.63%	

Jika melihat pada tabel tingkat akurasi, dapat disimpulkan kalau untuk kasus ini dan untuk prediksi atribut BersaljuHariIni, dataset yang belum dilakukan minmaxscaler memiliki rata-rata tingkat akurasi yang lebih baik yaitu sebesar 96%, berbeda 1.17% jika dibandingkan dengan dataset yang sudah dilakukan

minmaxscaler yang memiliki rata-rata tingkat akurasi sebesar 94.83%. Sedangkan untuk metode klasifikasinya, algoritma ID3 memiliki rata-rata tingkat akurasi yang paling baik yaitu sebesar 96.63%.

2. Prediksi atribut BersaljuBesok

- Dataset yang sudah dilakukan minmaxscaler

➤ Algoritma Naive Bayes

Confusion Matrix :

```
[[12226 1541]
```

```
[ 1708 1767]]
```

Accuracy Score : 0.811564783667788

Report :

	precision	recall	f1-score	support
0.0	0.88	0.89	0.88	13767
1.0	0.53	0.51	0.52	3475
accuracy			0.81	17242
macro avg	0.71	0.70	0.70	17242
weighted avg	0.81	0.81	0.81	17242

Dengan menggunakan algoritma Naive Bayes dihasilkan akurasi keseluruhan sebesar 81.15647836677880%. Jika mengacu pada nilai precision, prediksi tidak bersalju memiliki tingkat akurasi sebesar 88%, sedangkan untuk prediksi bersalju memiliki tingkat akurasi sebesar 53%.

➤ Algoritma ID3

Confusion Matrix :

```
[[11184 2583]
```

```
[ 1577 1898]]
```

Accuracy Score : 0.7587286857673123

Report :

	precision	recall	f1-score	support
0.0	0.88	0.81	0.84	13767
1.0	0.42	0.55	0.48	3475
accuracy			0.76	17242
macro avg	0.65	0.68	0.66	17242
weighted avg	0.79	0.76	0.77	17242

Dengan menggunakan algoritma ID3 dihasilkan akurasi keseluruhan sebesar 75.87286857673123%. Jika mengacu pada nilai precision, prediksi tidak bersalju memiliki tingkat akurasi sebesar 88%, sedangkan untuk prediksi bersalju memiliki tingkat akurasi sebesar 42%.

- Dataset yang belum dilakukan minmaxscaler

➤ Algoritma Naive Bayes

Confusion Matrix :

[[12072 1695]

[1516 1959]]

Accuracy Score : 0.8137687043266443

Report :

	precision	recall	f1-score	support
0	0.89	0.88	0.88	13767
1	0.54	0.56	0.55	3475
accuracy			0.81	17242
macro avg	0.71	0.72	0.72	17242
weighted avg	0.82	0.81	0.82	17242

Dengan menggunakan algoritma Naive Bayes dihasilkan akurasi keseluruhan sebesar 81.37687043266443%. Jika mengacu pada nilai precision, prediksi tidak bersalju memiliki tingkat akurasi sebesar 89%, sedangkan untuk prediksi bersalju memiliki tingkat akurasi sebesar 54%.

➤ Algoritma ID3

Confusion Matrix :

[[11961 1806]

[1690 1785]]

Accuracy Score : 0.7972392993852221

Report :

	precision	recall	f1-score	support
0	0.88	0.87	0.87	13767
1	0.50	0.51	0.51	3475
accuracy			0.80	17242
macro avg	0.69	0.69	0.69	17242
weighted avg	0.80	0.80	0.80	17242

Dengan menggunakan algoritma ID3 dihasilkan akurasi keseluruhan sebesar 79.72392993852221%. Jika mengacu pada nilai precision, prediksi tidak bersalju memiliki tingkat akurasi sebesar 88%, sedangkan untuk prediksi bersalju memiliki tingkat akurasi sebesar 50%.

Tabel dibawah ini merupakan hasil dari tingkat akurasi dari 6 percobaan yang sudah kami lakukan diatas.

Akurasi Atribut BersaljuBesok

	Naive Bayes	ID3	Rata-rata
Dataset yang sudah dilakukan minmaxscaler	81.15%	75.87%	78.51%
Dataset yang belum dilakukan minmaxscaler	81.37%	79.72%	80.54%
Rata-rata	81.26%	77.79%	

Jika melihat pada tabel tingkat akurasi, dapat disimpulkan kalau untuk kasus ini dan untuk prediksi atribut BersaljuBesok, dataset yang belum dilakukan minmaxscaler memiliki rata-rata tingkat akurasi yang lebih baik yaitu sebesar 80.54%, berbeda 2.03% jika dibandingkan dengan dataset yang sudah dilakukan minmaxscaler yang memiliki rata-rata tingkat akurasi sebesar 78.51%. Sedangkan untuk metode klasifikasinya, algoritma Naive Bayes memiliki rata-rata tingkat akurasi yang paling baik yaitu sebesar 81.26%.

E. Kesimpulan

Ada beberapa kesimpulan yang kami dapatkan setelah melakukan eksperimen dan analisis pada bagian sebelumnya, yaitu:

1. Perbedaan metode pada data pre-processing dan perbedaan metode klasifikasi dapat menghasilkan output tingkat akurasi yang berbeda juga.

2. Baik untuk atribut BersaljuHariIni maupun atribut BersaljuBesok, dataset yang belum dilakukan minmaxscaler mutlak menghasilkan prediksi dengan tingkat akurasi yang lebih baik.
3. Terdapat perbedaan algoritma yang lebih baik dalam menentukan atribut BersaljuHariIni dengan atribut BersaljuBesok. Untuk atribut BersaljuHariIni, algoritma ID3 memiliki tingkat akurasi yang paling baik jika dibandingkan dengan algoritma Naive Bayes. Namun untuk atribut BersaljuBesok, algoritma Naive Bayes memiliki tingkat akurasi yang paling baik jika dibandingkan dengan algoritma ID3.