

Nama : Qotrunnada Nabila
NIM : A11.2021.13613
Pertemuan : 2
Kelompok : A11.442U

Latihan Soal (Kuis)

- Carilah data yang dapat Anda gunakan untuk **Proses Data Mining**, kemudian gunakan **CRISP – DM** untuk menyelesaikan masalah tersebut.
- Jelaskan setiap langkahnya seperti **Studi Kasus Data Kelulusan Mahasiswa** diatas menggunakan **proses CRISP-DM** dimulai dari **Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation** dan **Deployment**.

1. Pemahaman Proses Bisnis (Business Understanding)

Problems:

- Sani merupakan seorang data analist yang tertarik dan berkerja dalam bidang perubahan iklim di dunia.
- Sani mendapat kabar dan keluhan mengenai masalah/kejadian perubahan iklim yang tidak bisa ditebak beberapa tahun terakhir ini. Perubahan tersebut membawa dampak terjadinya kelangkaan air, kegagalan panen dan meningkatnya harga panen sehingga mengakibatkan menurunnya produksi pangan dan kualitas hasil produksi pangan lokal di beberapa wilayah.
- Ia mengambil data yang memiliki kondisi iklim banyak yaitu, grimis, hujan, berkabut, terik, dan bersalju sehingga dapat menghasilkan pola prediksi iklim yang luas dan akurat.
- Menggunakan pola tersebut para petani dan nelayan dapat memprediksi iklim yang dapat mempengaruhi musim tanam/panen dan hasil produksi secara akurat.

Objective:

- Menemukan pola dalam memprediksi iklim atau cuaca guna membantu hasil panen.

2. Pemahaman Data (Data Understanding)

- Untuk menyelesaikan masalah Sani mengambil data iklim dalam bentuk kuantitatif yang telah dikumpulkan beberapa tahun terakhir pada daerah yang terkena dampak iklim tak tentu.
- Data yang dikumpulkan memiliki atribut seperti dibawah ini.
Precipitation (Curah Hujan) : data curah hujan dalam sehari
Temp_max : suhu maksimal dalam sehari
Temp_min : suhu maksimal dalam sehari
Wind (Tekanan Udara) : tekanan udara dalam sehari
Weather : drizzle (grimis), rain (hujan), fog (berkabut), sun (terik), atau snow (bersalju)

3. Persiapan Data (Data Preparation)

- Dataset : prediksiCuaca.csv

	date	precipitation	temp_max	temp_min	wind	weather
0	1/1/2012	0.0	12.8	5.0	4.7	drizzle
1	1/2/2012	10.9	10.6	2.8	4.5	rain
2	1/3/2012	0.8	11.7	7.2	2.3	rain
3	1/4/2012	20.3	12.2	5.6	4.7	rain
4	1/5/2012	1.3	8.9	2.8	6.1	rain
...
1456	12/27/2015	8.6	4.4	1.7	2.9	rain
1457	12/28/2015	1.5	5.0	1.7	1.3	rain
1458	12/29/2015	0.0	7.2	0.6	2.6	fog
1459	12/30/2015	0.0	5.6	-1.0	3.4	sun
1460	12/31/2015	0.0	5.6	-2.1	3.5	sun

1461 rows × 6 columns

- Terdapat 1461 data iklim dengan 6 atribut.
- Tidak terdapat missing value.

```
dataframe.isnull().sum()
```

```
date          0
precipitation 0
temp_max      0
temp_min      0
wind          0
weather       0
dtype: int64
```

- Membuat atribut baru untuk mengubah data yang berupa String menjadi int supaya dapat dibaca oleh program.

```
for i in range(len(dataframe)):
    if dataframe.loc[i,'weather']== 'drizzle':
        dataframe.loc[i,'prediksi']=1
    elif dataframe.loc[i,'weather']== 'rain':
        dataframe.loc[i,'prediksi']=2
    elif dataframe.loc[i,'weather']== 'sun':
        dataframe.loc[i,'prediksi']=3
    elif dataframe.loc[i,'weather']== 'fog':
        dataframe.loc[i,'prediksi']=4
    elif dataframe.loc[i,'weather']== 'snow':
        dataframe.loc[i,'prediksi']=5

dataframe['prediksi']=dataframe['prediksi'].astype(int)
```

	date	precipitation	temp_max	temp_min	wind	weather	prediksi
0	1/1/2012	0.0	12.8	5.0	4.7	drizzle	1
1	1/2/2012	10.9	10.6	2.8	4.5	rain	2
2	1/3/2012	0.8	11.7	7.2	2.3	rain	2
3	1/4/2012	20.3	12.2	5.6	4.7	rain	2
4	1/5/2012	1.3	8.9	2.8	6.1	rain	2
...
1456	12/27/2015	8.6	4.4	1.7	2.9	rain	2
1457	12/28/2015	1.5	5.0	1.7	1.3	rain	2
1458	12/29/2015	0.0	7.2	0.6	2.6	fog	4
1459	12/30/2015	0.0	5.6	-1.0	3.4	sun	3
1460	12/31/2015	0.0	5.6	-2.1	3.5	sun	3

4. Pemodelan (Modeling)

- Modelkan dataset dengan algoritma KNN dan memperoleh akurasi sebesar 76%

```

from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split

#membagi data set menjadi data latih dan data uji
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.30)

# Train the model
model = KNeighborsClassifier()
model.fit(x_train, y_train)

# Predict the test set results
y_pred = model.predict(x_test)

```

```
knn= KNeighborsClassifier(n_neighbors=7)
```

```
knn.fit(x_train, y_train)
```

```
KNeighborsClassifier(n_neighbors=7)
```

```
knn.score(x_test, y_test)*100
```

```
D:\anaconda3\lib\site-packages\sklearn\neighbors\_classification.py:228: Future
Warning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default
behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.
0, this behavior will change: the default value of `keepdims` will become Fals
e, the `axis` over which the statistic is taken will be eliminated, and the val
ue None will no longer be accepted. Set `keepdims` to True or False to avoid th
is warning.
```

```
mode, _ = stats.mode(_y[neigh_ind, k], axis=1)
```

```
76.5375854214123
```

5. Evaluasi (Evaluation)

Menghasilkan klasifikasi berupa prediksi iklim berdasarkan komponen-komponen yang dapat mempengaruhi perubahan iklim. Output yang dikeluarkan sesuai dengan pola prediksi dari data yang diambil.

6. Penerapan (Deployment)

Sani menerapkan hasil klasifikasi pada beberapa daerah dengan memasukkan inputan data yang diperlukan kemudian akan muncul hasil prediksi iklim yang akan terjadi pada daerah tersebut. Sehingga membantu para warga dalam memprediksi cuaca yang akan datang serta dapat meningkatkan produksi pangan lokal dengan kualitas yang baik dan menghindari petani mengalami keadaan gagal panen maupun nelayan dalam memprediksi cuaca sebelum melaut dan memperkirakan gelombang/angin laut yang aman untuk berlayar.

```
#memasukkan data baru (3)
```

```
new_precipitation = input("Masukan curah hujan : ")
```

```
new_temp_max = input("Masukan Temperature Maximal : ")
```

```
new_temp_min = input("Masukan Temperature Minimal : ")
```

```
new_wind = input("Masukan Tekanan angin : ")
```

```
prediksi_cuaca = knn.predict([[new_precipitation,new_temp_max,new_temp_min,new_wind]])
```

```
print("Hasil Prediksi Cuaca dari data diatas adalah "+nama_cuaca[prediksi_cuaca[0]])
```

```
Masukan curah hujan : 1.3
```

```
Masukan Temperature Maximal : 10
```

```
Masukan Temperature Minimal : 3.6
```

```
Masukan Tekanan angin : 2.5
```

```
Hasil Prediksi Cuaca dari data diatas adalah rain
```