

# **Laporan Case Based - 1 Machine Learning**



**Disusun oleh :**

**M. Nur Hikmal Afandi**

**1301204363**

**IF - 44 - 10**

**PROGRAM STUDI S1 INFORMATIKA**

**FAKULTAS INFORMATIKA**

**UNIVERSITAS TELKOM**

**2022**

## 1. Data Preparation

Data Preparation adalah tahap pertama ketika akan mengolah data karena data yang sudah tersedia harus dipastikan kembali kualitas datanya sehingga salah satu tujuan data preparation adalah memastikan kualitas data sudah bagus. Pada data yang sudah tersedia mungkin tidak semua data terpakai sehingga perlu diringkas kembali.

Pada tahap data preparation ini telah dilakukan pengisian missing value yang diisi menggunakan mean, menghapus data - data outliers, dan menghapus duplikat data. Gambar dibawah adalah rincian dari datasetnya. Terdapat 1 missing value yang sudah tambahkan oleh mean dari kolom yang mempunyai missing value. Ada duplicate rows yang dihapus.

Dataset statistics

Number of variables	27
Number of observations	776
Missing cells	1
Missing cells (%)	< 0.1%
Duplicate rows	10
Duplicate rows (%)	1.3%
Total size in memory	163.8 KiB
Average record size in memory	216.2 B

Variable types

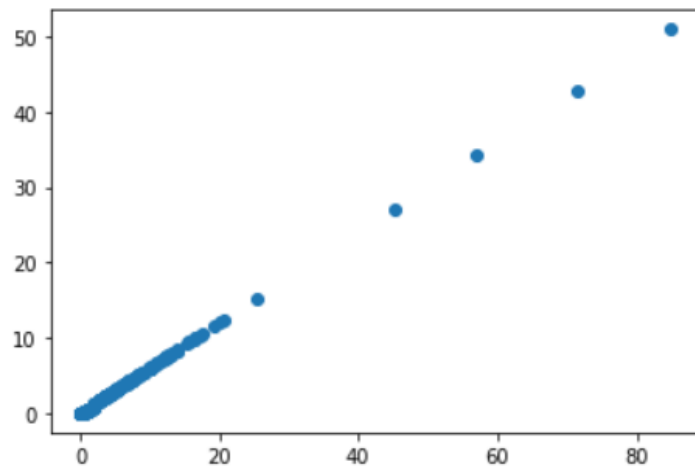
Numeric	14
Categorical	13

Gambar dibawah adalah data hasil ringkasan sehingga tersisa 461 rows dan 26 kolom.

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TOTAL	numbers	...	PROB	Risk_E	History	Prob	Risk_F	Score
0	3.89	23.0	4.18	0.6	2.508	2.50	0.2	0.500	6.68	5.0	...	0.2	0.4	0.0	0.2	0.0	2.4
1	3.89	6.0	0.00	0.2	0.000	4.83	0.2	0.966	4.83	5.0	...	0.2	0.4	0.0	0.2	0.0	2.0
2	3.89	6.0	0.51	0.2	0.102	0.23	0.2	0.046	0.74	5.0	...	0.2	0.4	0.0	0.2	0.0	2.0
4	3.89	6.0	0.00	0.2	0.000	0.08	0.2	0.016	0.08	5.0	...	0.2	0.4	0.0	0.2	0.0	2.0
5	3.89	6.0	0.00	0.2	0.000	0.83	0.2	0.166	0.83	5.0	...	0.2	0.4	0.0	0.2	0.0	2.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
770	55.57	18.0	0.75	0.2	0.150	0.45	0.2	0.090	1.20	5.0	...	0.2	0.4	0.0	0.2	0.0	2.0
772	55.57	16.0	0.47	0.2	0.094	0.37	0.2	0.074	0.84	5.0	...	0.2	0.4	0.0	0.2	0.0	2.0
773	55.57	14.0	0.24	0.2	0.048	0.04	0.2	0.008	0.28	5.0	...	0.2	0.4	0.0	0.2	0.0	2.0
774	55.57	18.0	0.20	0.2	0.040	0.00	0.2	0.000	0.20	5.0	...	0.2	0.4	0.0	0.2	0.0	2.0
775	55.57	15.0	0.00	0.2	0.000	0.00	0.2	0.000	0.00	5.0	...	0.2	0.4	0.0	0.2	0.0	2.0

451 rows × 26 columns

Terdapat contoh Scatter Plot dari salah satu atribut pada dataset yaitu atribut PARA\_A dan atribut Risk\_A



## 2. Prapemrosesan Data

Pada prapemrosesan data telah dilakukan beberapa hal pada dataset yaitu pada missing value yang terdapat pada dataset dilakukan treatment imputasi yaitu menambahkan value pada cell yang missing value dengan menambahkan nilai mean. Pada dataset juga terdapat baris dan kolom yang dihapus. Baris yang dihapus dikarenakan terdapat null atau karena terjadi duplikat data serta terdapat outliers. Kolom dihapus karena terdapat kolom yang memiliki value yang sama semua sehingga tidak akan berpengaruh pada dataset, kolom Detection\_Risk dihapus karena memiliki value yang sama satu kolom tersebut.

Kode dibawah ini berfungsi untuk melakukan proses imputasi atau menambah mean pada data yang kosong.

```
#Mengganti data null pada kolom Money_Value oleh rata - rata dari kolom Money_Value
data.Money_Value = data.Money_Value.fillna(data.Money_Value.mean())
```

Pada data dibawah ini dilakukan pengecekan duplicate data dan dilkuakn drop pada duplicate data tersebut.

```
#Cek Duplikat Data
data_new.duplicated().any()
```

True

```
hasil_data_drop = data_new.drop_duplicates()
hasil_data_drop
```

Kode dibawah ini yaitu melakukan drop terhadap outliers yang ada pada dataset

Outliers

```
[24] #Q1, Q3 dan IQR
      Q1 = data.quantile(0.25)
      Q3 = data.quantile(0.75)
      IQR = Q3 - Q1

      #Check ukuran (baris dan kolom) sebelum data yang outliers dibuang
      print('Shape awal : ', data.shape)

      #Removing Outliers
      data = data[~((data < (Q1 - 1.5 * IQR)) | (data > (Q3 + 1.5 * IQR))) ]

      print('Shape AKHIR : ', data.shape)
      data
```

### 3. Penerapan Algoritma ANN

Algoritma ANN (Artificial Neural Network) adalah salah satu algoritma supervised learning yang memiliki konsep mensimulasikan jaringan syaraf biologis pada manusia. Pada algoritma ANN akan ada 3 input, setiap input akan dikalikan dengan terlebih dahulu dengan variable weight setelah itu ketiganya dijumlahkan. Setiap koneksi neuron - neuron memiliki weight nya masing - masing, nilainya akan berubah saat proses learning sampai model yang dihasilkan ANN mendekati target output yang diinginkan.

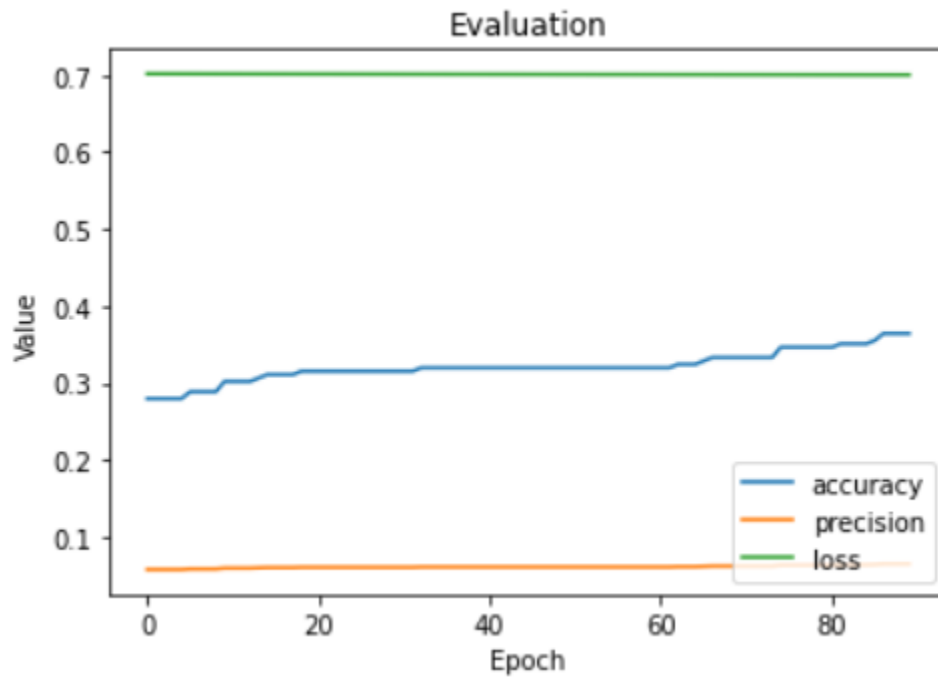
Dibawah ini terdapat implementasi algoritma ANN pada implementasi menggunakan hyperparameter relu dan hyperparameter sigmoid. Adaptive Momentum Estimation (ADAM ) yang digunakan dengan learning rate = 0.01.

```
model = Sequential([
    Dense(64, activation = 'relu'),
    Dense(32, activation = 'relu'),
    Dense(16, activation = 'relu'),
    Dense(8, activation = 'relu'),
    Dense(4, activation = 'relu'),
    Dense(2, activation = 'relu'),
    Dense(1, activation = 'sigmoid'),
])
model.compile(loss = 'binary_crossentropy', optimizer = tf.optimizers.Adam(learning_rate=0.01), metrics = ['accuracy', 'Precision', 'Recall'])
```

[https://colab.research.google.com/drive/1D5R6B5M0sHG4-5udA\\_1S5T2S3VZfWJ25?usp=sharing](https://colab.research.google.com/drive/1D5R6B5M0sHG4-5udA_1S5T2S3VZfWJ25?usp=sharing)

#### 4. Evaluasi Hasil

Berdasarkan hasil implementasi didapatkan bahwa akurasi adalah 36.73 % dengan presisi 8.33 %



akurasi : 36.73%  
presisi : 8.33%

## 5. Presentasi

- Link video presentasi :

<https://youtu.be/7hThpdsLpbM>

- Link Slide presentasi :

[https://docs.google.com/presentation/d/1cg-iwRgZEpiotW2E1RL3\\_cHIcVhidSgwLp50MLtAJbc/edit?usp=sharing](https://docs.google.com/presentation/d/1cg-iwRgZEpiotW2E1RL3_cHIcVhidSgwLp50MLtAJbc/edit?usp=sharing)

- Link Google Collab :

[https://colab.research.google.com/drive/1D5R6B5M0sHG4-5udA\\_lS5T2S3VZfWJ25?usp=sharing](https://colab.research.google.com/drive/1D5R6B5M0sHG4-5udA_lS5T2S3VZfWJ25?usp=sharing)

- Link Laporan :

<https://docs.google.com/document/d/1wviDosBwOK73sXn1IdWswIY6WD9ijUqQNeuQXu10NyA/edit?usp=sharing>

