PRESENTATION

MULTINOMIAL LOGISTIC REGRESSION

- Hajar Hanifah

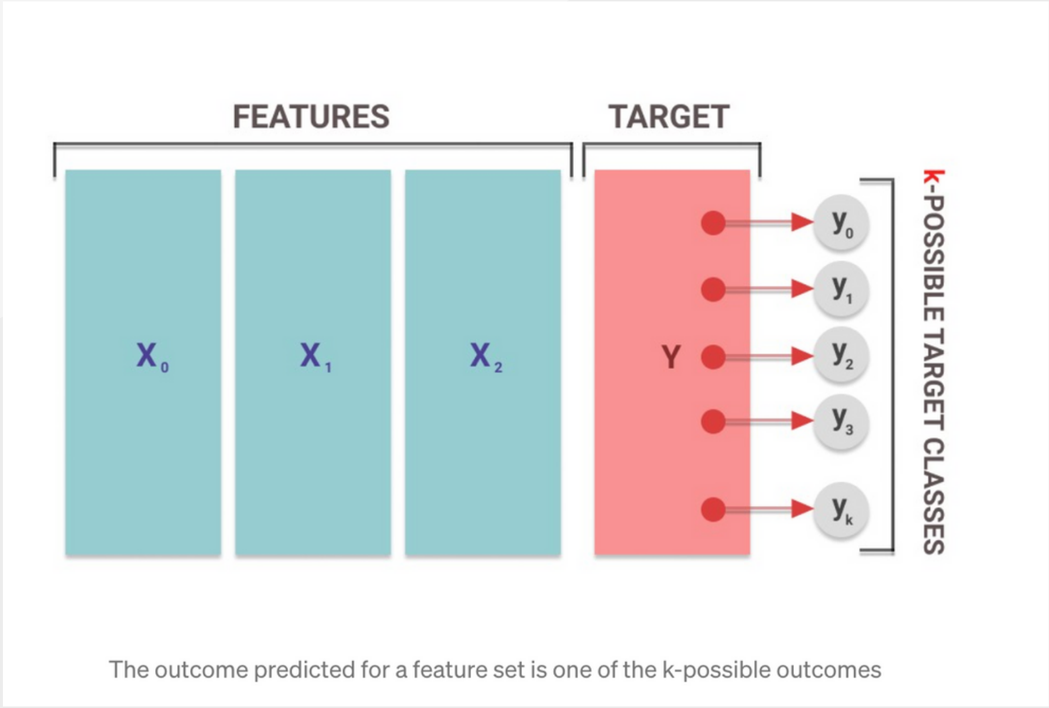
TABLE OF CONTENTS

MULTINOMIAL LOGISTIC REGRESSION

Regresi logistik multinomial adalah perpanjangan sederhana dari regresi logistik biner yang memungkinkan lebih dari dua kategori variabel dependen atau hasil. Seperti regresi logistik biner, regresi logistik multinomial menggunakan estimasi kemungkinan maksimum untuk mengevaluasi probabilitas keanggotaan kategoris.

sumber : https://it.unt.edu/sites/default/files/mlr_jds_aug2011.pdf

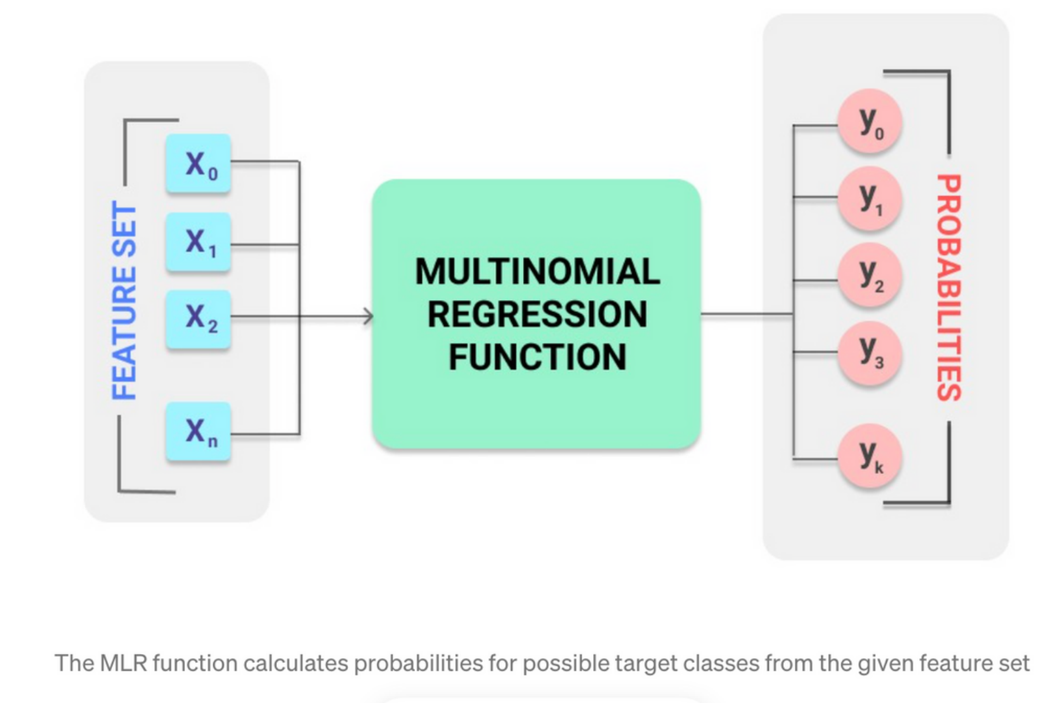
MULTINOMIAL LOGISTIC REGRESSION BASIC



The outcome predicted for a feature set is one of the k-possible outcomes

MLR dapat memprediksi satu dari k-kemungkinan hasil, di mana k dapat berupa sembarang bilangan bulat positif.

HOW DOES MLR WORK?



The MLR function calculates probabilities for possible target classes from the given feature set

- Fungsi regresi multinomial = algoritma klasifikasi statistik.

TUJUAN PENELITIAN

Tujuan penelitian ini adalah melakukan Multinomial Logistic Regression, mencari NPL Score terhadap data seluruh emiten saham di sektor aneka industri untuk sub sektor Mesin dan Komponen Industri Q4 2021.

DATASET

Data

Data yang diambil merupakan data statictical financial ratio Q4 2021 untuk emiten di sektor Industri

Sumber Dataset

Data di dapatkan dari website IDX - Laporan Statistic Financial Data Ratio. https://www.idx.co.id/data-pasar/laporan-statistik/digital- statistic-beta/financial-data-ratio?

DATA UNDERSTANDING

Emiten	Pihak yang melakukan penawaran umum, yaitu penawaran efek yang dilakukan oleh emiten untuk menjual efek kepada masyarakat berdasarkan tata cara yang diatur dalam peraturan undang-undang yang berlaku.
Sales Growth	Kenaikan jumlah penjualan dari tahun ke tahun atau dari waktu ke waktu.
Retturn of Asset (ROA)	Indikator untuk menunjukkan seberapa untuk sebuah perusahaan dibandingkan dengan total asetnya.
Debt to Equity Ratio (DER)	Rasio hutang terhadap ekuitas atau rasio keuangan yang membandingkan jumlah hutang dengan ekuitas

DATA UNDERSTANDING

Category	Hasil Clustering data Emiten sektor Industri - Aneka Industri , Mesin dan Komponen Industri
ROE (Return Of Equity)	Return on equity atau ROE adalah indikator kinerja perusahaan dengan membandingkan laba bersih dan total modal
NPM (Net Profit Margin)	ingkat keuntungan suatu perusahaan dari penjualan atau pendapatan yang diperoleh.

RUMUS

Sales Growth

$$\frac{\text{(penjualan periode 2021 - penjualan periode 2020)}}{\text{penjualan periode 2020}}$$

02 ROA

$$\frac{\text{Laba setelah pajak}}{\text{Total Asset}}$$

03 DER

$$\frac{\text{Total Uang}}{\text{Total Ekuitas}}$$

DATA UNDERSTANDING

Rasio yang digunakan :

- X1. DER = Total Debt / Total Equity
- X2. Return on Asset = Net Income / total Asset
- X3. Sales Growth = (Sales 2021 - sales 2020)/ sales 2020
- X4. Others
- X5 : others
- Y. Performance / Category = Using Clustering result yang hasilnya didapatkan dari Clustering Sektor Industri - Aneka Industri Q4 2021

DATA UNDERSTANDING

Dengan kelompok kategori

- Y0 / Kategori 0 = Cluster yang terdampak negative / Buruk
- Y1 / Kategori 1 = Cluster yang tidak terdampak Covid-19 / Netral
- Y2 / Kategori 2 = Cluster yang terdampak Positif / Baik
- Y3 / Kategori 3 = Cluster yang terdampak sangat negative / Sangat Buruk
- Y4/ Kategori 4 = Cluster yang terdampak sangat positif / Sangat Baik
-

DATA PREPARATION

Siapkan Dataset

Siapkan dataset berupa data Emiten saham, ROA, DER, dll pada satu file

Multinomial logistic regression with scikit-learn

- assign predictors dan kriteria untuk setiap objek
- split x dan Y
- Y = categori
- buat variabel training dan testing

Standardize the Data

Lakukan standarisasi data, tentukan training dan testing sizenya

Model Validation

cari tahu akurasi dan error ratenya

Probabilitas

kalkulasikan probabilitasnya, lihat predicted dan actual classnya. Serta kalkulasikan akurasinya

DEPLOYMENT

Bahasa pemrograman yang digunakan adalah Phyton, adapun code dapat di akses di :

https://github.com/Hajarhanifah/big-data/blob/main/multinomial-logistic-regression/multinomial-logistic-regression-2-aneka-industri.ipynb

STEP

1. Import Libraries
2. Exploratory Data Analysis
    - Data Understanding
    - Understanding the Table
    - Load Data
    - Modeling
3. Multinomial logistic regression with scikit-learn
    - Data Preparation
    - Fit the Model
    - Standardize the Data
    - Model Validation
    - Probabilitas
4. Conclutiont berupa data Emiten saham, ROA, DER, dll pada satu file

DATA

	No	Industri	Company	Emiten	DER	ROA	Sales Growth	ROE	NPM	Category
0	1	Industrial Machinery & Components	Asahimas Flat Glass Tbk	AMFG	1.320000	0.060000	0.330000	0.130000	0.120000	1
1	2	Industrial Machinery & Components	PT Ateliers Mecaniques D Indonesie Tbk.	AMIN	1.140000	-0.050000	0.410000	-0.120000	-0.130000	1
2	3	Industrial Machinery & Components	PT Arita Prima Indonesia Tbk.	APII	0.490000	0.050000	0.030000	0.070000	0.130000	1
3	4	Industrial Machinery & Components	PT Arkha Jayanti Persada Tbk.	ARKA	3.970000	-0.020000	0.220000	-0.080000	-0.160000	1
4	5	Industrial Machinery & Components	Arwana Citramulia Tbk	ARNA	0.510000	0.210000	0.170000	0.320000	0.250000	4
5	6	Industrial Machinery & Components	Cahayaputra Asa Keramik Tbk	CAKK	0.880000	0.030000	0.300000	0.060000	0.070000	1
6	7	Industrial Machinery & Components	Communication Cable Systems Indonesia Tbk	CCSI	0.310000	0.110000	0.790000	0.140000	0.170000	4
7	8	Industrial Machinery & Components	Citatah Tbk	CTTH	2.280000	-0.050000	-0.170000	-0.160000	-0.540000	1
8	9	Industrial Machinery & Components	Hexindo Adiperkasa Tbk	HEXA	1.380000	0.120000	0.680000	0.290000	0.210000	4
9	10	Industrial Machinery & Components	Sumi Indo Kabel Tbk	IKBI	0.670000	0.010000	0.670000	0.010000	0.010000	1
10	11	Industrial Machinery & Components	Impack Pratama Industri Tbk	IMPC	0.720000	0.080000	0.280000	0.130000	0.130000	1
11	12	Industrial Machinery & Components	Intraco Penta Tbk	INTA	-2.860000	-0.320000	-0.220000	0.000000	-1.990000	3
12	13	Industrial Machinery & Components	Jembo Cable Company Tbk	JECC	1.580000	-0.040000	0.160000	-0.100000	-0.050000	1
13	14	Industrial Machinery & Components	KMI Wire & Cable Tbk	KBLI	0.160000	0.020000	-0.170000	0.020000	0.050000	1
14	15	Industrial Machinery & Components	Kabelindo Murni Tbk	KBLM	0.380000	0.000000	0.360000	-0.010000	-0.010000	1
15	16	Industrial Machinery & Components	Keramika Indonesia Assosiasi Tbk	KIAS	0.180000	-0.010000	0.410000	-0.020000	-0.030000	1
16	17	Industrial Machinery & Components	Kobexindo Tractors Tbk	KOBX	2.790000	0.060000	1.440000	0.240000	0.070000	2
17	18	Industrial Machinery & Components	Kokoh Inti Arebama Tbk	KOIN	8.530000	-0.020000	0.990000	-0.210000	-0.010000	2
18	19	Industrial Machinery & Components	Steadfast Marine Tbk	KPAL	3.150000	-0.020000	0.000000	-0.080000	-0.500000	1
19	20	Industrial Machinery & Components	PT Grand Kartech Tbk	KRAH	16.330000	-0.050000	0.000000	-0.930000	-0.180000	0
20	21	Industrial Machinery & Components	PT Mark Dynamics Indonesia Tbk.	MARK	0.670000	0.310000	1.420000	0.520000	0.400000	4
21	22	Industrial Machinery & Components	Mulia Industrindo Tbk	MLIA	0.910000	0.080000	0.160000	0.150000	0.150000	1
22	23	Industrial Machinery & Components	Supreme Cable Manufacturing & Commerce Tbk	SCCO	0.080000	0.040000	0.160000	0.050000	0.040000	1
23	24	Industrial Machinery & Components	Singaraja Putra Tbk	SINI	4.090000	0.020000	0.290000	0.100000	0.010000	1
24	25	Industrial Machinery & Components	Superkrane Mitra Utama Tbk	SKRN	1.680000	0.010000	-0.280000	0.010000	0.030000	1
25	26	Industrial Machinery & Components	Surya Pertiwi Tbk	SPTO	0.540000	0.060000	0.170000	0.080000	0.110000	1
26	27	Industrial Machinery & Components	Surya Toto Indonesia Tbk	TOTO	0.650000	0.020000	0.140000	0.040000	0.060000	1
27	28	Industrial Machinery & Components	United Tractors Tbk	UNTR	0.590000	0.090000	0.240000	0.150000	0.180000	1
28	29	Industrial Machinery & Components	Voksel Electric Tbk	VOKS	1.910000	-0.060000	-0.150000	-0.180000	-0.140000	1

DATA

	No	Industri	Company	Emiten	DER	ROA	Sales Growth	ROE	NPM	Category
0	1	Industrial Machinery & Components	Asahimas Flat Glass Tbk	AMFG	1.320000	0.060000	0.330000	0.130000	0.120000	1
1	2	Industrial Machinery & Components	PT Ateliers Mecaniques D Indonesie Tbk.	AMIN	1.140000	-0.050000	0.410000	-0.120000	-0.130000	1
2	3	Industrial Machinery & Components	PT Arita Prima Indonesia Tbk.	APII	0.490000	0.050000	0.030000	0.070000	0.130000	1
3	4	Industrial Machinery & Components	PT Arkha Jayanti Persada Tbk.	ARKA	3.970000	-0.020000	0.220000	-0.080000	-0.160000	1
4	5	Industrial Machinery & Components	Arwana Citramulia Tbk	ARNA	0.510000	0.210000	0.170000	0.320000	0.250000	4
5	6	Industrial Machinery & Components	Cahayaputra Asa Keramik Tbk	CAKK	0.880000	0.030000	0.300000	0.060000	0.070000	1
6	7	Industrial Machinery & Components	Communication Cable Systems Indonesia Tbk	CCSI	0.310000	0.110000	0.790000	0.140000	0.170000	4
7	8	Industrial Machinery & Components	Citatah Tbk	CTTH	2.280000	-0.050000	-0.170000	-0.160000	-0.540000	1
8	9	Industrial Machinery & Components	Hexindo Adiperkasa Tbk	HEXA	1.380000	0.120000	0.680000	0.290000	0.210000	4
9	10	Industrial Machinery & Components	Sumi Indo Kabel Tbk	IKBI	0.670000	0.010000	0.670000	0.010000	0.010000	1
10	11	Industrial Machinery & Components	Impack Pratama Industri Tbk	IMPC	0.720000	0.080000	0.280000	0.130000	0.130000	1
11	12	Industrial Machinery & Components	Intraco Penta Tbk	INTA	-2.860000	-0.320000	-0.220000	0.000000	-1.990000	3
12	13	Industrial Machinery & Components	Jembo Cable Company Tbk	JECC	1.580000	-0.040000	0.160000	-0.100000	-0.050000	1
13	14	Industrial Machinery & Components	KMI Wire & Cable Tbk	KBLI	0.160000	0.020000	-0.170000	0.020000	0.050000	1
14	15	Industrial Machinery & Components	Kabelindo Murni Tbk	KBLM	0.380000	0.000000	0.360000	-0.010000	-0.010000	1
15	16	Industrial Machinery & Components	Keramika Indonesia Assosiasi Tbk	KIAS	0.180000	-0.010000	0.410000	-0.020000	-0.030000	1
16	17	Industrial Machinery & Components	Kobexindo Tractors Tbk	KOBX	2.790000	0.060000	1.440000	0.240000	0.070000	2
17	18	Industrial Machinery & Components	Kokoh Inti Arebama Tbk	KOIN	8.530000	-0.020000	0.990000	-0.210000	-0.010000	2
18	19	Industrial Machinery & Components	Steadfast Marine Tbk	KPAL	3.150000	-0.020000	0.000000	-0.080000	-0.500000	1
19	20	Industrial Machinery & Components	PT Grand Kartech Tbk	KRAH	16.330000	-0.050000	0.000000	-0.930000	-0.180000	0
20	21	Industrial Machinery & Components	PT Mark Dynamics Indonesia Tbk.	MARK	0.670000	0.310000	1.420000	0.520000	0.400000	4
21	22	Industrial Machinery & Components	Mulia Industrindo Tbk	MLIA	0.910000	0.080000	0.160000	0.150000	0.150000	1
22	23	Industrial Machinery & Components	Supreme Cable Manufacturing & Commerce Tbk	SCCO	0.080000	0.040000	0.160000	0.050000	0.040000	1
23	24	Industrial Machinery & Components	Singaraja Putra Tbk	SINI	4.090000	0.020000	0.290000	0.100000	0.010000	1
24	25	Industrial Machinery & Components	Superkrane Mitra Utama Tbk	SKRN	1.680000	0.010000	-0.280000	0.010000	0.030000	1
25	26	Industrial Machinery & Components	Surya Pertiwi Tbk	SPTO	0.540000	0.060000	0.170000	0.080000	0.110000	1
26	27	Industrial Machinery & Components	Surya Toto Indonesia Tbk	TOTO	0.650000	0.020000	0.140000	0.040000	0.060000	1
27	28	Industrial Machinery & Components	United Tractors Tbk	UNTR	0.590000	0.090000	0.240000	0.150000	0.180000	1
28	29	Industrial Machinery & Components	Voksel Electric Tbk	VOKS	1.910000	-0.060000	-0.150000	-0.180000	-0.140000	1

CATEGORY 4 : SANGAT POSITIF

```
display(df[df['category'] == 4])
```

	emiten	stock name	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER	ROA
4	ARNA	Arwana Citramulia Tbk	4	-0.317212	-0.419440	1.806855	0.17	0.51	0.21
6	CCSI	Communication Cable Systems Indonesia Tbk	4	1.145218	-0.479896	0.827438	0.79	0.31	0.11
8	HEXA	Hexindo Adiperkasa Tbk	4	0.885755	-0.156456	0.925380	0.68	1.38	0.12
20	MARK	PT Mark Dynamics Indonesia Tbk.	4	2.631236	-0.371075	2.786272	1.42	0.67	0.31

CATEGORY 2 : POSITIF

```
display(df[df['category'] == 2])
```

	emiten	stock name	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER	ROA
16	KOBX	Kobexindo Tractors Tbk	2	2.678412	0.269759	0.337730	1.44	2.79	0.06
17	KOIN	Kokoh Inti Arebama Tbk	2	1.616970	2.004850	-0.445804	0.99	8.53	-0.02

CATEGORY 1 : NETRAL

```
display(df[df['category'] == 1])
```

	emiten	stock name	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER	ROA
0	AMFG	Asahimas Flat Glass Tbk	1	0.060189	-0.174593	0.337730	0.33	1.32	0.06
1	AMIN	PT Ateliers Mecaniques D Indonesie Tbk.	1	0.248890	-0.229004	-0.739629	0.41	1.14	-0.05
2	APII	PT Arita Prima Indonesia Tbk.	1	-0.647439	-0.425486	0.239788	0.03	0.49	0.05
3	ARKA	PT Arkha Jayanti Persada Tbk.	1	-0.199274	0.626450	-0.445804	0.22	3.97	-0.02
5	CAKK	Cahayaputra Asa Keramik Tbk	1	-0.010574	-0.307597	0.043905	0.30	0.88	0.03
7	CTTH	Citatah Tbk	1	-1.119190	0.115596	-0.739629	-0.17	2.28	-0.05
9	IKBI	Sumi Indo Kabel Tbk	1	0.862167	-0.371075	-0.151978	0.67	0.67	0.01
10	IMPC	Impack Pratama Industri Tbk	1	-0.057749	-0.355961	0.533613	0.28	0.72	0.08
12	JECC	Jembo Cable Company Tbk	1	-0.340800	-0.096000	-0.641687	0.16	1.58	-0.04
13	KBLI	KMI Wire & Cable Tbk	1	-1.119190	-0.525239	-0.054037	-0.17	0.16	0.02
14	KBLM	Kabelindo Murni Tbk	1	0.130952	-0.458737	-0.249920	0.36	0.38	0.00
15	KIAS	Keramika Indonesia Assosiasi Tbk	1	0.248890	-0.519193	-0.347862	0.41	0.18	-0.01
18	KPAL	Steadfast Marine Tbk	1	-0.718201	0.378580	-0.445804	0.00	3.15	-0.02
21	MLIA	Mulia Industrindo Tbk	1	-0.340800	-0.298528	0.533613	0.16	0.91	0.08
22	SCCO	Supreme Cable Manufacturing & Commerce Tbk	1	-0.340800	-0.549421	0.141847	0.16	0.08	0.04
23	SINI	Singaraja Putra Tbk	1	-0.034161	0.662724	-0.054037	0.29	4.09	0.02
24	SKRN	Superkrane Mitra Utama Tbk	1	-1.378654	-0.065772	-0.151978	-0.28	1.68	0.01
25	SPTO	Surya Pertiwi Tbk	1	-0.317212	-0.410372	0.337730	0.17	0.54	0.06
26	TOTO	Surya Toto Indonesia Tbk	1	-0.387975	-0.377121	-0.054037	0.14	0.65	0.02
27	UNTR	United Tractors Tbk	1	-0.152099	-0.395258	0.631555	0.24	0.59	0.09
28	VOKS	Voksel Electric Tbk	1	-1.072015	0.003752	-0.837570	-0.15	1.91	-0.06

CATEGORY 0 : NEGATIVE

```
]: display(df[df['category'] == 0])
```

	emiten	stock name	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER	ROA
19	KRAH	PT Grand Kartech Tbk	0	-0.718201	4.362638	-0.739629	0.0	16.33	-0.05

CATEGORY 3 : SANGAT NEGATIVE

```
display(df[df['category'] == 3])
```

	emiten	stock name	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER	ROA
11	INTA	Intraco Penta Tbk	3	-1.237128	-1.438126	-3.384054	-0.22	-2.86	-0.32

MULTINOMIAL LOGISTIC REGRESSION WITH SCIKIT-LEARN

Tetapkan prediktor dan kriteria untuk setiap objek dan membagi datensatz menjadi bagian pelatihan dan pengujian

X : DER, ROA, SALES GROWTH, NPM
Y : CATEGORY

```
[800]: x = df.drop(['Category','No','Industri','Company','Emiten'], axis=1)
       x.head()
```

[80…

	DER	ROA	Sales Growth	ROE	NPM
0	1.320000	0.060000	0.330000	0.130000	0.120000
1	1.140000	-0.050000	0.410000	-0.120000	-0.130000
2	0.490000	0.050000	0.030000	0.070000	0.130000
3	3.970000	-0.020000	0.220000	-0.080000	-0.160000
4	0.510000	0.210000	0.170000	0.320000	0.250000

```
y = df['Category']
y.head()
```

[80…

```
0    1
1    1
2    1
3    1
4    4
Name: Category, dtype: int64
```

+ Code + Markdown

STANDARDIZE THE DATA

Tujuan :

- agar dapat mengoptimalkan parameter model,
- penskalaan dapat dilakukan secara signifikan,
- meningkatkan kecepatan dan akurasi pengoptimalan data.

$$X' = \frac{X - \mu}{\sigma}$$

Here,

- μ = Mean of all the values within a column
- σ = Standard deviation of the column

STANDARDIZE THE DATA

3.2 Standardize the Data

[2562]:
```
from sklearn.preprocessing import StandardScaler
from sklearn import preprocessing
x_scaled = StandardScaler().fit_transform(x)
data_scaled = pd.DataFrame(x_scaled, columns=df.columns.drop(['Industri','Company','Emiten','Category','ROE']))

data_scaled
```

[25…

	No	DER	ROA	Sales Growth	NPM
0	-0.174593	0.337730	0.060189	0.457607	0.419982
1	-0.229004	-0.739629	0.248890	-0.595616	-0.185360
2	-0.425486	0.239788	-0.647439	0.204834	0.444196
3	0.626450	-0.445804	-0.199274	-0.427100	-0.258001
4	-0.419440	1.806855	-0.317212	1.258057	0.734760
5	-0.307597	0.043905	-0.010574	0.162705	0.298914
6	-0.479896	0.827438	1.145218	0.499736	0.541050
7	0.115596	-0.739629	-1.119190	-0.764131	-1.178121
8	-0.156456	0.925380	0.885755	1.131670	0.637905
9	-0.371075	-0.151978	0.862167	-0.047940	0.153632

SPLIT DATA, TENTUKAN X TRAIN, Y TRAIN, X TEST DAN Y TEST

```
[804]: trainX, testX, trainY, testY = train_test_split(x, y, test_size = 0.3)
```

Diketahui :

- training : 70%
- testing 30%

Diketahui :
- training size : 70%
- testing size : 30%

semakin besar data training, semakin besar akurasi

CONVERTING LOGIT

```
mnlogit_mod = sm.MNLogit(y,sm.add_constant(data_scaled))
mnlogit_fit = mnlogit_mod.fit(method='bfgs')

print (mnlogit_fit.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.000013
         Iterations: 19
         Function evaluations: 20
         Gradient evaluations: 20
                          MNLogit Regression Results
==============================================================================
Dep. Variable:                Category   No. Observations:                   29
Model:                         MNLogit   Df Residuals:                       5
Method:                            MLE   Df Model:                          20
Date:                 Wed, 06 Jul 2022   Pseudo R-squ.:                  1.000
Time:                         05:09:43   Log-Likelihood:            -0.00037979
converged:                       True   LL-Null:                       -26.785
Covariance Type:             nonrobust   LLR p-value:                6.699e-05
==============================================================================
  Category=1       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const           19.4679   6890.967      0.003      0.998   -1.35e+04    1.35e+04
No              -9.4023   6071.701     -0.002      0.999   -1.19e+04    1.19e+04
DER             -4.8520   2.98e+04     -0.000      1.000   -5.84e+04    5.84e+04
ROA             -9.3511   8984.133     -0.001      0.999   -1.76e+04    1.76e+04
Sales Growth     0.5490   5966.315    9.2e-05      1.000   -1.17e+04    1.17e+04
NPM              4.3875   2.62e+04      0.000      1.000   -5.13e+04    5.13e+04
------------------------------------------------------------------------------
  Category=2       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const           -4.1715   1.29e+04     -0.000      1.000   -2.54e+04    2.54e+04
No               1.8280   3682.085      0.000      1.000   -7214.927    7218.583
DER             -0.2063   3.65e+04  -5.66e-06      1.000   -7.15e+04    7.15e+04
ROA             11.8315   7416.284      0.002      0.999   -1.45e+04    1.45e+04
Sales Growth     3.7040   9225.151      0.000      1.000   -1.81e+04    1.81e+04
NPM              0.4250   7.86e+04    5.4e-06      1.000   -1.54e+05    1.54e+05
------------------------------------------------------------------------------
  Category=3       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const           -6.6830   3.71e+04     -0.000      1.000   -7.28e+04    7.28e+04
No              -3.7460   6.89e+04  -5.44e-05      1.000   -1.35e+05    1.35e+05
DER             -4.8789    9.8e+04  -4.98e-05      1.000   -1.92e+05    1.92e+05
ROA             -1.3672   8.01e+04  -1.71e-05      1.000   -1.57e+05    1.57e+05
Sales Growth     1.6032    9.2e+04   1.74e-05      1.000    -1.8e+05     1.8e+05
NPM             -7.8324   7.58e+04     -0.000      1.000   -1.49e+05    1.49e+05
------------------------------------------------------------------------------
  Category=4       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const           -0.5430   6841.773  -7.94e-05      1.000   -1.34e+04    1.34e+04
No              -3.6925   6928.272     -0.001      1.000   -1.36e+04    1.36e+04
DER             12.5186   2.98e+04      0.000      1.000   -5.84e+04    5.84e+04
ROA              2.6967   8938.080      0.000      1.000   -1.75e+04    1.75e+04
Sales Growth     8.4361   6355.567      0.001      0.999   -1.24e+04    1.25e+04
NPM              4.7752   2.62e+04      0.000      1.000   -5.13e+04    5.13e+04
==============================================================================
```

LOGISTIC REGRESSION

$$\log\left(\frac{Y}{1-Y}\right) = C + B1X1 + B2X2 + \ldots.$$

- Y is the probability of an event to happen which you are trying to predict
- x1, x2 are the independent variables which determine the occurrence of an event i.e. Y
- C is the constant term which will be the probability of the event happening when no other factors are considered

Y Prediction

[2566]:
```
logistic_regresion = LogisticRegression(multi_class='auto').fit(x,y)
y_pred = logistic_regresion.predict(testX)
y_pred
```

[25…] `array([1, 1, 1, 2, 1, 0, 1, 1, 1])`

multiclass = auto , agar dapat mengenerate secara otomatis Y nya apa saja (categori 0,1, 2, 3 4) sesuai dengan nilai Y- nya

Intercept

+ Code + Markdown

[2567]:
```
print('Intercept :', logistic_regresion.intercept_)
```

```
Intercept : [-5.43616632  3.93807952 -0.82916441  0.42773238  1.89951884]
```

+ Code + Markdown

Koefisien

```
print(logistic_regresion.coef_)
```

```
[[ 0.96676318 -0.00402755 -0.09958339 -0.08469312 -0.01798341]
 [ 0.03128648 -0.22798664 -1.20875198 -0.58734021  0.08282493]
 [ 0.5682376  -0.00301388  0.85484969  0.13319329  0.04347885]
 [-1.09240544 -0.11833112 -0.17970715 -0.02731978 -0.64963306]
 [-0.47388182  0.35335919  0.63319283  0.56615983  0.54131269]]
```

+ Code + Markdown

MODEL VALIDATION / TESTING THE MODEL

Mendapatkan Prediction test menggunakan model

[2569]:
```
prediction_test = logistic_regresion.predict(testX)
```

Print akurasi dan error rate

[2570]:
```
print('Accuracy: {:.2f}'.format(accuracy_score(testY, y_pred)))
print('Error rate: {:.2f}'.format(1 - accuracy_score(testY, y_pred)))
```

```
Accuracy: 0.89
Error rate: 0.11
```

Matrix

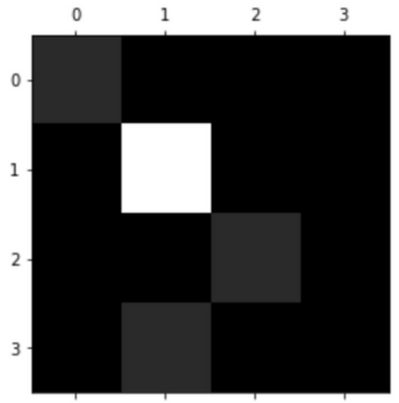
[2571]:
```
confusion_matrix = confusion_matrix(testY, y_pred)
print(confusion_matrix)
```

```
[[1 0 0 0]
 [0 6 0 0]
 [0 0 1 0]
 [0 1 0 0]]
```

[2571]:
```
confusion_matrix = confusion_matrix(testY, y_pred)
print(confusion_matrix)
```

```
[[1 0 0 0]
 [0 6 0 0]
 [0 0 1 0]
 [0 1 0 0]]
```

[2572]:
```
plt.matshow(confusion_matrix, cmap=plt.cm.gray)
plt.show()
```



+ Code + Markdown

PROBABILITAS

```
probability = logistic_regresion.predict_proba(testX)
probability
```

```
array([[3.13059612e-04, 8.54137469e-01, 2.84586316e-02, 1.10582739e-02,
        1.06032566e-01],
       [2.30551717e-04, 8.29791682e-01, 2.20377127e-02, 1.23828313e-02,
        1.35557223e-01],
       [1.68281115e-04, 5.71358472e-01, 3.29966454e-02, 2.62644864e-02,
        3.69212115e-01],
       [9.32080732e-02, 1.39828128e-01, 7.65792102e-01, 7.11067189e-07,
        1.17098620e-03],
       [1.47926844e-04, 8.34582788e-01, 1.44005568e-02, 1.57295227e-02,
        1.35139206e-01],
       [8.88884217e-01, 3.85122995e-03, 1.07264513e-01, 8.32737254e-13,
        4.01087297e-08],
       [1.20478653e-04, 7.68227954e-01, 1.65120504e-02, 2.88774200e-02,
        1.86262097e-01],
       [5.49338275e-04, 9.61083943e-01, 1.76188080e-02, 2.51230507e-03,
        1.82356058e-02],
       [1.59207540e-04, 8.58184559e-01, 1.42457375e-02, 1.39961881e-02,
        1.13414308e-01]])
```

+ Code + Markdown

- Setiap kolom merepresentasikan class
- class dengan probabilitas paling tinggi = output dari predicted class

[2574]:

```
print(probability.shape[0])
print(testX.shape[0])
```

```
9
9
```

Disini, kita dapat melihat bahwa panjang data probabilitas (9) = panjang data uji (9)

PROBABILITAS

[2576]:
```
data_result = pd.DataFrame(logistic_regresion.predict_proba(testX), columns=logistic_regresion.classes_)
data_result
```

[25...

	0	1	2	3	4
0	0.000313	0.854137	0.028459	0.011058	0.106033
1	0.000231	0.829792	0.022038	0.012383	0.135557
2	0.000168	0.571358	0.032997	0.026264	0.369212
3	0.093208	0.139828	0.765792	0.000001	0.001171
4	0.000148	0.834583	0.014401	0.015730	0.135139
5	0.888884	0.003851	0.107265	0.000000	0.000000
6	0.000120	0.768228	0.016512	0.028877	0.186262
7	0.000549	0.961084	0.017619	0.002512	0.018236
8	0.000159	0.858185	0.014246	0.013996	0.113414

PREDICTED CLASS

[2578]:
```
data_result['predicted_class'] = y_pred
data_result
```

[25...

	0	1	2	3	4	sum	predicted_class
0	0.000313	0.854137	0.028459	0.011058	0.106033	1.000000	1
1	0.000231	0.829792	0.022038	0.012383	0.135557	1.000000	1
2	0.000168	0.571358	0.032997	0.026264	0.369212	1.000000	1
3	0.093208	0.139828	0.765792	0.000001	0.001171	1.000000	2
4	0.000148	0.834583	0.014401	0.015730	0.135139	1.000000	1
5	0.888884	0.003851	0.107265	0.000000	0.000000	1.000000	0
6	0.000120	0.768228	0.016512	0.028877	0.186262	1.000000	1
7	0.000549	0.961084	0.017619	0.002512	0.018236	1.000000	1
8	0.000159	0.858185	0.014246	0.013996	0.113414	1.000000	1

ACTUAL CLASS

```
2579]: data_result['actual_class'] = testY.to_frame().reset_index().drop(columns='index')
       data_result
```

[25...

	0	1	2	3	4	sum	predicted_class	actual_class
0	0.000313	0.854137	0.028459	0.011058	0.106033	1.000000	1	1
1	0.000231	0.829792	0.022038	0.012383	0.135557	1.000000	1	1
2	0.000168	0.571358	0.032997	0.026264	0.369212	1.000000	1	4
3	0.093208	0.139828	0.765792	0.000001	0.001171	1.000000	2	2
4	0.000148	0.834583	0.014401	0.015730	0.135139	1.000000	1	1
5	0.888884	0.003851	0.107265	0.000000	0.000000	1.000000	0	0
6	0.000120	0.768228	0.016512	0.028877	0.186262	1.000000	1	1
7	0.000549	0.961084	0.017619	0.002512	0.018236	1.000000	1	1
8	0.000159	0.858185	0.014246	0.013996	0.113414	1.000000	1	1

Dari data diatas, kita dapat melihat hasil dari prediksi setiap class. Untuk memudahkan dalam membaca prediksi, lakukan konversi kelas prediksi (prediction class) dan actual class menggunakan encoder label dari scikit-learn, lanjutkan dengan nilai numerik.

DI SINI KITA MELIHAT BAHWA DUA VARIABEL (PREDICTED_CLASS & ACTUAL_CLASS) DIBERI KODE YANG SAMA

```
le = preprocessing.LabelEncoder()

data_result['label_pred'] = le.fit_transform(data_result['predicted_class'])
data_result['label_actual'] = le.fit_transform(data_result['actual_class'])
data_result
```

	0	1	2	3	4	sum	predicted_class	actual_class	label_pred	label_actual
0	0.000313	0.854137	0.028459	0.011058	0.106033	1.000000	1	1	1	1
1	0.000231	0.829792	0.022038	0.012383	0.135557	1.000000	1	1	1	1
2	0.000168	0.571358	0.032997	0.026264	0.369212	1.000000	1	4	1	3
3	0.093208	0.139828	0.765792	0.000001	0.001171	1.000000	2	2	2	2
4	0.000148	0.834583	0.014401	0.015730	0.135139	1.000000	1	1	1	1
5	0.888884	0.003851	0.107265	0.000000	0.000000	1.000000	0	0	0	0
6	0.000120	0.768228	0.016512	0.028877	0.186262	1.000000	1	1	1	1
7	0.000549	0.961084	0.017619	0.002512	0.018236	1.000000	1	1	1	1
8	0.000159	0.858185	0.014246	0.013996	0.113414	1.000000	1	1	1	1

+ Code + Markdown

Di sini kita melihat bahwa dua variabel (predicted_class & actual_class) diberi kode yang sama

MAPPING

[2581]:
```
targets = data_result['predicted_class']
integerEncoded = le.fit_transform(targets)
integerMapping=dict(zip(targets,integerEncoded))
integerMapping
```

[25…] {1: 1, 2: 2, 0: 0}

[2582]:
```
targets = data_result['actual_class']
integerEncoded = le.fit_transform(targets)
integerMapping=dict(zip(targets,integerEncoded))
integerMapping
```

[25…] {1: 1, 4: 3, 2: 2, 0: 0}

MAPPING

Cek apakah prediksi class benar

[2583]:
```
data_result['check'] = data_result['label_actual'] - data_result['label_pred']
data_result
```

[25...

	0	1	2	3	4	sum	predicted_class	actual_class	label_pred	label_actual	check
0	0.000313	0.854137	0.028459	0.011058	0.106033	1.000000	1	1	1	1	0
1	0.000231	0.829792	0.022038	0.012383	0.135557	1.000000	1	1	1	1	0
2	0.000168	0.571358	0.032997	0.026264	0.369212	1.000000	1	4	1	3	2
3	0.093208	0.139828	0.765792	0.000001	0.001171	1.000000	2	2	2	2	0
4	0.000148	0.834583	0.014401	0.015730	0.135139	1.000000	1	1	1	1	0
5	0.888884	0.003851	0.107265	0.000000	0.000000	1.000000	0	0	0	0	0
6	0.000120	0.768228	0.016512	0.028877	0.186262	1.000000	1	1	1	1	0
7	0.000549	0.961084	0.017619	0.002512	0.018236	1.000000	1	1	1	1	0
8	0.000159	0.858185	0.014246	0.013996	0.113414	1.000000	1	1	1	1	0

[1]:
```
data_result['correct_prediction?'] = np.where(data_result['check'] == 0, 'True', 'False')
data_result = data_result.drop(['label_pred', 'label_actual', 'check'], axis=1)
data_result
```

	0	1	2	3	4	sum	predicted_class	actual_class	correct_prediction?
0	0.000313	0.854137	0.028459	0.011058	0.106033	1.000000	1	1	True
1	0.000231	0.829792	0.022038	0.012383	0.135557	1.000000	1	1	True
2	0.000168	0.571358	0.032997	0.026264	0.369212	1.000000	1	4	False
3	0.093208	0.139828	0.765792	0.000001	0.001171	1.000000	2	2	True
4	0.000148	0.834583	0.014401	0.015730	0.135139	1.000000	1	1	True
5	0.888884	0.003851	0.107265	0.000000	0.000000	1.000000	0	0	True
6	0.000120	0.768228	0.016512	0.028877	0.186262	1.000000	1	1	True
7	0.000549	0.961084	0.017619	0.002512	0.018236	1.000000	1	1	True
8	0.000159	0.858185	0.014246	0.013996	0.113414	1.000000	1	1	True

KALAU RESULT SUBSTRACTION == 0 BERARTI PREDIKSII BENAR / TRUE

kalkulasikan value dari akurasi, dan probabilitas kelas yang salah prediksi

[2586]:
```
true_predictions = data_result[(data_result["correct_prediction?"] == 'True')].shape[0]
false_predictions = data_result[(data_result["correct_prediction?"] == 'False')].shape[0]
total = df1["correct_prediction?"].shape[0]

print('manual calculated Accuracy is:', (true_predictions / total * 100))
```

```
manual calculated Accuracy is: 88.88888888888889
```

Probabilitas kelas yang salah prediksi

[2587]:
```
wrong_pred = data_result[(data_result["correct_prediction?"] == 'False')]
wrong_pred
```

[25...

	0	1	2	3	4	sum	predicted_class	actual_class	correct_prediction?
2	0.000168	0.571358	0.032997	0.026264	0.369212	1.000000	1	4	False

KESIMPULAN

- Regresi Logistik Multinomial dapat digunakan untuk memprediksi beberapa kelas
- Dengan data yang ada, dilakukan validasi dengan testing size 30% menghasilkan akurasi data 88,88%
- Panjang probabilitas dan panjang data uji yang dihasilkan dari model MLR sektor insudtri masing- masing = 9
- Nilai akurasi awal dan akhir tidak berubah

KESIMPULAN

- Regresi multinomial dapat digunakan untuk memprediksi beberapa kelas, pada kasus ini, metode digunakan untuk memprediksi terjadinya penurunan secara finansial pada emiten sektor Aneka Industri sub sektor Mesin dan komponen Industri. Adapun rasio yang digunakan adalah Return of Asset (ROA), Ratio Return on Equity (ROE), Sales growth, Debt to Equity Ratio (DER), dan Net Profit Margin (NPM) berdasarkan indikator Kategori hasil dari clustering, dimana kelompok kategori dibagi menjadi 5 yaitu Sangat positif, Positif, Netral, Negatif dan Sangat Negatif. Hasil uji menunjukkan bahwa model yang digunakan sudah sangat baik dengan tingkat akurasi prediksi yaitu 89%.
-

THANK YOU

- Hajar Hanifah

EMAIL : HAJAR.HANIFAH@GMAIL.COM
GITHUB : GITHUB.COM/HAJARHANIFAH
KAGGLE : HAJARHANIFAH