

IMPLEMENTATION REGRESSION ANALYSIS AND K-MEANS TO PREDICT AND CLASSIFY FIXED DATA ASSET

Handy Noviyarto

ABSTRACT

Regional assets represent regional assets which in essence belong to the respective provincial government. These government assets can play a role as collateral for regional development. The preparation of asset documents aims to safeguard assets from the aspect of regional administration. In this study, the method used are Regression Analysis and K-Means Method. The purpose of this study was to predict and classify fixed data asset using Regression Analysis and K-Means Method. This research was conducted using the Python programming language and the Visual Studio code.

KEYWORDS: Analysis Regression, K-means, Clustering, Data mining

1. INTRODUCTION

Local assets are essentially regional wealth is owned by the provincial government each - each. One is a regional asset is an asset not move. As for which is included in the fixed assets to which such land or land, buildings, and so forth. In this aspect, it can play a role of government assets as collateral development in the region. Preparation of the document aims to secure the assets of the assets of the administrative aspects of the area.

According to the Government Accounting Standards (SAP) (2016) assets are economic resources controlled or owned by the government as a result of past events and from which economic and social benefits in the future depandiharapkan can be obtained either by the government or the public, as well as dapatdiukur in units of money, including non-financial resources needed to provide services to the public and the sources of power in maintained for historical and cultural reasons.

Asset security aims to keep local assets do not change hands illegally and facilitate local authorities in managing further. Absolute asset security is done by completing the assets in question to a legal document. In addition, a regional asset wealth can act as a guarantee of regional development.

A common problem of the government's assets, which is not yet completed the document, even none at all. Not infrequently, the region's assets lost due to various reasons. As yet completed documents are Letter or Certificate of Land Ownership History certain land owned by the provincial government, incomplete documents such as letters leasing, handing over others. Table 1.1 lists the data that show the problem in asset Jakarta Education Agency

Based on the background of the issue and the importance of asset security systems in every activity of the company - the company, so in this study was taken the title "Implementation Regression Analysis And K-Means To Predict And Classify Fixed Asset Data In Education Authorities"

2. PLATFORM THEORY

2.1 Definition of Data Mining

Data mining is the process to obtain useful information from large data base warehouse. Techniques in Data Mining: how to search for the data that is to build a model. The model was used to identify the pattern of other data that are not in the data base stored.

2.2. Regression Analysis

Regression analysis in statistics is one method for determining the causal relationship between one variable and another variable (s). "Cause" variables are referred to by various terms: explanatory variables, explanatory variables, independent variables, or independently, variable X (because it is often depicted on the graph as abscissa, or the X-axis). Variables affected as a result are known as influenced variables, dependent variables, dependent variables, or Y variables. Both of these variables can be random variables (random), but the variables affected must always be random variables.

Regression analysis is one of the most popular and widely used analyzes. Regression analysis is widely used to make predictions and forecasts, with uses that complement each other in the field

Faculty of Computer Science, Mercu Buana University, Indonesia

HOW TO CITE THIS ARTICLE:

Handy Noviyarto (2020). Implementation Regression Analysis And K-Means To Predict And Classify Fixed Data Assetin Education Authorities, International Educational Journal of Science and Engineering (IEJSE), Vol: 3, Issue: 1, 01-05

Copyright© 2020, IEJSE. This open-access article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License which permits Share (copy and redistribute the material) under the Attribution-NonCommercial terms.

of machine learning. This analysis is also used to understand which independent variables are related to the dependent variable, and to find out the forms of the relationship.

2.2 K Means

K-means is a clustering algorithm. The purpose of this algorithm is to divide data into groups. This algorithm accepts input in the form of data without class labels. This is different from supervised learning which accepts input in the form of vectors (x1, y1), (x2, y2), ..., (xi, yi), where xi is the data from a training data and yi is the class label for xi.

In this learning algorithm, the computer groups its own data into input without first knowing the target class. This learning is included in unsupervised learning. The input received is the data or object and the desired group (cluster). This algorithm will group data or objects into these groups. In each cluster there is a center point (centroid) that represents the cluster.

3. RESULTS AND DISCUSSION

3.1 Predictive Analysis (Regression Method)

a. Using Data Preprocessing: Before using the code for Data Processing, first input library that will be used:



We have made, then uflood some parts, such as the name of the dataset is loaded:

Out[19]: Kode_Brang KiB Jeris_Brang Lus Statum Almost Tahun 1 010104002 KB Tanah Bangunan Pendidkan Dan Lathan (sekolah) 5.100 M2 J.I.C.Kini Faya No.87 105 2 010104002 KB Tanah Bangunan Pendidkan Dan Lathan (sekolah) 1.800 M2 J.I.C.Kini Faya No.87 105 3 101104002 KB Tanah Bangunan Pendidkan Dan Lathan (sekolah) 1.800 M2 J.I.Penica No.10 R.D.02 / Ru.00 105 4 101104002 KB Tanah Bangunan Pendidkan Dan Lathan (sekolah) 6.700 M2 J.I.Penica No.10 R.D.02 / Ru.00 105 5 1011104002 KB Tanah Bangunan Pendidkan Dan Lathan (sekolah) 7.000 M2 J.I. Masjid Nor No.33 R.D.20 / Flu.00 105 6 1011104002 KB Tanah Bangunan Pendidkan Dan Lathan (sekolah) 2.400 M2 J.I. Masjid Nor No.33 R.D.20 / Flu.00 105 7 101104002 KB Tanah Bangunan Pendidkan Dan Lathan (sekolah) 2.400 M2 J.I. Main Masia Nor No.33 R.D.20 / Flu.000 105 <th>In [19]: Ŋ</th> <th>data: data:</th> <th colspan="7"><pre>dataset = pd.read_csv("C:/Users/HP/Documents/Big Data/Project II/Data_aset_II_2.csv", sep = "; dataset.head(18)</pre></th> <th>")</th>	In [19]: Ŋ	data: data:	<pre>dataset = pd.read_csv("C:/Users/HP/Documents/Big Data/Project II/Data_aset_II_2.csv", sep = "; dataset.head(18)</pre>							")		
0 101110-4002 K18 D Tanah Bangunan Pendidkan Dan Lathan (sekotah) 5.190 MQ J. Cikini Raya No 57 192 1 101110-4002 K19 D Tanah Bangunan Pendidkan Dan Lathan (sekotah) 1.800 MQ J.L.Cikini Raya No 57 192 2 101110-4002 K19 D Tanah Bangunan Pendidkan Dan Lathan (sekotah) 4.80 MQ J.L.PalukerAH BARKT No.50 192 3 101110-4002 K19 D Tanah Bangunan Pendidkan Dan Lathan (sekotah) 9.700 MQ J.L.PalukerAH BARKT No.50 192 4 101110-4002 K19 D Tanah Bangunan Pendidkan Dan Lathan (sekotah) 9.700 MQ J.L.PalukerAH BARKT No.50 192 5 101110-4002 K19 D Tanah Bangunan Pendidkan Dan Lathan (sekotah) 2.803 MQ J.M. Medah Merdela Timur 1/7 Mender 198 6 101110-4002 K19 D Tanah Bangunan Pendidkan Dan Lathan (sekotah) 2.800 MQ J.I. F. No.1 Kebon Baru 192 7 101110-4002 K19 D Tanah Bangunan Pendidkan Dan Lathan (sekotah) 2.800 MQ J.F. No.1 Kebon Baru	Out[19]:	К	ide_Barang	KIB	3		Jenis_Ba	rang	Luas	Satuan	Alamat	Tahun
1 101110-002 K18 0 Tanah Banguran Pendidkan Dan Lathan (sekolah) 1.800 M2 JALAN PERUK N022 193 2 101110-002 K18 D Tanah Banguran Pendidkan Dan Lathan (sekolah) 4.180 M2 J.F.Penira No.10 Pt.002 / Pw.008 196 3 101110-002 K18 D Tanah Banguran Pendidkan Dan Lathan (sekolah) 9.700 M2 J.F.Penira No.10 Pt.002 / Pw.008 196 4 101110-002 K18 D Tanah Banguran Pendidkan Dan Lathan (sekolah) 9.700 M2 J.M.beidi Mur Ho.33 R.2017 1011 6 101110-002 K18 D Tanah Banguran Pendidkan Dan Lathan (sekolah) 2.400 M2 J.M.beidi Mur Ho.33 R.2018 198 7 101110-002 K18 D Tanah Banguran Pendidkan Dan Lathan (sekolah) 2.400 M2 J.R.Markan Merdeka Timur 1/7 198 7 101110-002 K18 D Tanah Banguran Pendidkan Dan Lathan (sekolah) 2.400 M2 J.R.Horikan Merdeka Timur 1/7 198 10 101110-002 K18 D <td></td> <td>0</td> <td>1011104002</td> <td>KIB D</td> <td>) Tanah Bar</td> <td>igunan Pendidik</td> <td>an Dan Latihan (sek</td> <td>olah)</td> <td>5.190</td> <td>M2</td> <td>JI. Cikini Raya No 87</td> <td>1950</td>		0	1011104002	KIB D) Tanah Bar	igunan Pendidik	an Dan Latihan (sek	olah)	5.190	M2	JI. Cikini Raya No 87	1950
2 1011104002 VB 0 Tanah Banguran Pendidan Dan Lathan (sekolah) 4.180 M2 JL Penkin No.10 Pendod 192 3 1011104002 VB D Tanah Banguran Pendidan Dan Lathan (sekolah) 9.700 M2 JL PENLERAH APART NO.50 192 4 1011104002 VB D Tanah Banguran Pendidan Dan Lathan (sekolah) 9.700 M2 JL Madia Method 1480 M2 JL PakuKERAH APART NO.50 192 5 1011104002 VB Tanah Banguran Pendidan Dan Lathan (sekolah) 7.7000 M2 JL Madia Method Method 18 6 1011104002 VB Tanah Banguran Pendidan Dan Lathan (sekolah) 2.400 M2 JL Bukin Turur / 71 Method 18 6 1011104002 VB Tanah Banguran Pendidkan Dan Lathan (sekolah) 2.300 M2 JL Bukin Turur / 71 Method 18 7 10110402 VB Danah Banguran Pendidkan Dan Lathan (sekolah) <td></td> <td>1</td> <td>1011104002</td> <td>KIB D</td> <td>) Tanah Bar</td> <td>ıgunan Pendidik</td> <td>an Dan Latihan (sek</td> <td>olah)</td> <td>1.800</td> <td>M2</td> <td>JALAN PERUK NO. 32</td> <td>1951</td>		1	1011104002	KIB D) Tanah Bar	ıgunan Pendidik	an Dan Latihan (sek	olah)	1.800	M2	JALAN PERUK NO. 32	1951
3 1011104002 148 0 Tanah Bangunan Pendidkan Dan Lathan (sekolah) 9.700 M2 JL PALMERAH BARAT N0.59 193 4 101104002 108 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 775.000 M2 JL PALMERAH BARAT N0.59 193 5 1011104002 108 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 775.000 M2 JL PALMERAH BARAT N0.59 193 6 1011104002 108 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 72.000 M2 JL PALMERAH BARAT N0.59 193 8 1011104002 108 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 2.400 M2 JL PALMERAH BARAT N0.59 193 9 1011104002 108 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 2.000 M2 JL F. No.1 Kebon Ban 193 01102 101104002 108 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 1.700 M2 JL Banah Bangunan Pendidkan Dan Lathan (sekolah) 1.700 M2 JL Banah Bangunan Pendidkan Dan Lathan (sekolah) 1.700 M2 JL Banah Bangunan Pendidkan Dan Lathan (sekolah) 1.700 M2 <td></td> <td>2</td> <td>1011104002</td> <td>KIB D</td> <td>) Tanah Bar</td> <td>igunan Pendidik</td> <td>an Dan Latihan (sek</td> <td>olah)</td> <td>4.180</td> <td>M2</td> <td>JI.Penuira No.10 Rt.002 / Rw.008</td> <td>1951</td>		2	1011104002	KIB D) Tanah Bar	igunan Pendidik	an Dan Latihan (sek	olah)	4.180	M2	JI.Penuira No.10 Rt.002 / Rw.008	1951
4 1011104002 108 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 75.000 M2 Jil. Mesjid Nur No.33 Ft.002 / 101 196 5 1011104002 108 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.400 M2 Jil. Mesjid Nur No.33 Ft.002 / 101 196 6 1011104002 108 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.400 M2 Jil. Mesjid Nur No.33 Ft.002 / 101 196 7 1011104002 108 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.300 M2 Jil. Bulak Timur I/ 1 196 9 1011104002 108 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.300 M2 Jil. Bulak Timur I/ 1 196 011104002 108 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 1.700 M2 Jil. Utama Raya No.41 196 01110402 108 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 1.700 M2 Jil. Utama Raya No.41 196 011014420 100000000 1000.000000 1000.00000 1000.00000 13472094 8.008085 13472094 8.008000		3	1011104002	KIB D) Tanah Bar	igunan Pendidik	an Dan Latihan (sek	olah)	9.700	M2	JL PALMERAH BARAT NO. 59	1959
5 1011104002 K8 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.400 M2 Jin. Medan Merdela Timur 14 198 6 1011104002 K8 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 4.808 M2 Jin. Medan Merdela Timur 14 198 7 1011104002 K8 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.303 M2 Jil Balai Timur 17 Viender 198 8 101110402 K8 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.800 M2 Jil Balai Timur 17 Viender 198 9 101110402 K8 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.800 M2 Jil Balai Timur 17 Viender 198 01110402 K8 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 1.200 M2 Jil Balai Timur 17 Viender 198 01110402 K8 D Tanah Banguran Pendiskan Dan Lathan (sekolah) 1.20 M2 Jil Balai Timur 14 198 0112(32): M dataset.describe() Jil Balai Timur 14 198 198 Jil Balai Timur 14 198 0110812		4	1011104002	KIB D) Tanah Bar	igunan Pendidik	an Dan Latihan (sek	olah)	775.000	M2	JI. Masjid Nur No.33 Rt.002 / 010	1962
6 101110-4002 K8 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 4.808 M2 jl Pertanian Klender 196 7 101110-4002 K8 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 2.383 M2 Jl Budak Timuri I/7 Nender 196 8 101110-4002 K8 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 2.383 M2 Jl Budak Timuri I/7 Nender 196 9 101110-4002 K8 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 2.900 M2 JL F. No.1 Nabon Baru 196 10 10110-4002 K8 D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 1.730 M2 JL tama Raya No.41 196 Cont (201) Mataset. describe() Kode_Barang Luss Tahun add 1.407096e405 113.472064 8.00885 min 1.010101e-409 1.957000 1967.000000 25% 1.01110-e-409 2.052000 1968.000000 1968.000000 197.57100 197.1104-e-409 2.15900 198.1104-e-40 198.1104-e-40 198.1104-e-40 <td></td> <td>5</td> <td>1011104002</td> <td>KIB D</td> <td>) Tanah Bar</td> <td>igunan Pendidik</td> <td>an Dan Latihan (sek</td> <td>olah)</td> <td>2.400</td> <td>M2</td> <td>Jin. Medan Merdeka Timur 14</td> <td>1962</td>		5	1011104002	KIB D) Tanah Bar	igunan Pendidik	an Dan Latihan (sek	olah)	2.400	M2	Jin. Medan Merdeka Timur 14	1962
7 1011104002 KB D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.383 M2 JE Bulak Timur I/7 Viender 198 8 1011104002 KB D Tanah Banguran Pendiskan Dan Lathan (sekolah) 2.000 M2 J.E F. No. 1 Kebon Baru 198 9 1011104002 KB D Tanah Banguran Pendiskan Dan Lathan (sekolah) 1.730 M2 J.U tama Raya No. 41 198 0 101104002 KB D Tanah Banguran Pendiskan Dan Lathan (sekolah) 1.730 M2 J.U tama Raya No. 41 198 0 101104002 KB D Tanah Banguran Pendiskan Dan Lathan (sekolah) 1.730 M2 J.U tama Raya No. 41 198 0 041104400 1000.00000 1000.00000 1000.00000 1000.00000 1000.00000 1000.00000 1000.00000 1000.00000 1000.00000 10101411 10101114409 11010402 1000.00000 1000.00000 1000.00000 1011111111111111111111111111111111111		6	1011104002	KIB D) Tanah Bar	igunan Pendidik	an Dan Latihan (sek	olah)	4.808	M2	jl.Pertanian Klender	1963
8 101110-4002 KlB D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 2.000 M2 J. F. No. 1 Kabon Baru 196 9 101110-4002 KlB D Tanah Bangunan Pendidkan Dan Lathan (sekolah) 1.730 M2 J. E. No. 1 Kabon Baru 196 Int (20): M dataset.describe() Kode_Barang Luss Thrun		7	1011104002	KIB D) Tanah Bar	igunan Pendidik	an Dan Latihan (sek	olah)	2.363	M2	JI Bulak Timur I/7 Klender	1963
9 1011104002 KB 0 Tanah Bangunan Pendidkan Dan Lathan (sekolah) 1.730 M2 J. Utama Raya No. 41 190 In [20]: M dataset.describe() J. Utama Raya No. 41 190 0ut[30]: M dataset.describe() <		8	1011104002	KIB D) Tanah Bar	igunan Pendidik	an Dan Latihan (sek	olah)	2.900	M2	JI. F. No.1 Kebon Baru	1969
In [20]: N dataset.describe() Out[20]: Kode_Burang Luss Tahun oourt 1.000000e-03 1000.00000 1000.00000 mean 1.01108+e-09 21.910009 1987.657000 sid 1.407098+e-05 113.727204 80.00895 min 1.010101e-09 1.05700 1980.00000 25% 1.01104+e-09 2.12500 1984.00000 50% 1.01104+e-09 3.05300 1985.00000 75% 1.01104+e-09 4.49175 1988.00000		9	1011104002	KIB D) Tanah Bar	ıgunan Pendidik	an Dan Latihan (sek	olah)	1.730	M2	JI. Utama Raya No. 41	1969
Kode_Barang Luas Tahun oourt 1.00000e403 1000.000000 1000.000000 mean 1.011084e00 21.919009 1987.857000 rad 1.407986e400 11.9742964 80.08885 min 1.011014e409 1.957000 1980.00000 25% 1.011014e409 2.129500 1986.00000 50% 1.011014e409 3.035000 1985.00000 75% 1.01104e409 4.431759 1980.00000	In [20]: 🕅	dataset.describe()										
count 1.000000+403 1000.00000 1000.00000 mean 1.011084+09 21.919609 1987.657000 std 1.407980+405 113.472894 8.008855 min 1.011010+409 1.057000 1980.00000 25% 1.011104+409 2.12500 1984.00000 50% 1.011104+409 3.053000 1985.000000 75% 1.011104+409 3.053000 1988.000000 75% 1.011104+409 3.052000 1989.000000	Out[20]:		Kode Bara	ina	Luas	Tabun						
mean 1.01108-4-09 21.919009 1987.857000 std 1.4079696-405 113.472204 8.608685 min 1.010101+-09 1.057000 1950.00000 25% 1.011104-409 2.12500 1984.00000 50% 1.011104-409 3.057000 1985.00000 75% 1.01104-409 4.451700 1989.00000		count	1.000000e	+03 11	000.000000	1000.000000						
add 1.407080e+e05 113.4720804 8.008085 min 1.010101e+e09 1.057000 1980.00000 25 % 1.011104e+e09 2.125000 1986.00000 50 % 1.011104e+e09 3.035000 1985.00000 75 % 1.011104e+e09 3.035000 1985.00000 75 % 0.01104e+e09 3.035000 1980.00000		mean	1.011084	+09	21,916909	1987.657000						
min 1.010101e+409 1.057000 1950.000000 25% 1.011104+409 2.123500 1984.000000 50% 1.011104+409 3.035000 1985.00000 75% 1.011104+409 4.431750 1985.000000		std	1.407996e	+05	113.472694	8.608685						
25% 1.011104++09 2.129500 1984.000000 50% 1.011104++09 3.035000 1985.000000 75% 1.011104++09 4.431750 1988.000000 75% 1.011104++09 4.431750 1988.000000		min	1.010101e	+09	1.057000	1950.000000						
50% 1.011104++09 3.035000 1985.000000 75% 1.011104++09 4.431780 1988.00000 mm 1.011104++09 0.68.00000 0.0000		25%	1.011104e	+09	2.129500	1984.000000						
75% 1.0.11104e+09 4.431750 1988.000000		50%	1.011104e	+09	3.035000	1985.000000						
max 1.0112016-00. 078.000000. 2017.000000		75%	1.011104	+09	4 431750	1988 000000						
		max	1 0113014	 	976 000000	2017 000000						

Figure 1: Data Pre Processing Regression Analysis

b. Fitting Simple Linear Regression in Training Set:

3.	Library Train - Test Dataset
In [38]: 🕅	<pre>msk = np.random.rand(len(df)) < 0.8 train = asset[msk] test = asset[~msk]</pre>
4.	Fitting Simple Linear Regression pada Training-Set
In [72]: 🕅	<pre>from sklearn import linear_model regr = linear_model.LinearRegression() train_x = np.asanyarey(train['Tahun']]) train_y = np.asanyarey(train[['Kode_Bararg']]) regr.fit (train_x, train_y)</pre>
Out[72]:	LinearRepression(copy_X=True, fit_intercept=True, n_jobs=Wone, normalize=False)
In [73]: 🕅	<pre>plt.scatter(train.Tahun, train.Kode_Barang, color='green') plt.plot(train_X, regr.coef_[0][0]'train_X + regr.intercept_[0], '-r') plt.slabel("Tahun") plt.slabel("Kode_Barang") plt.title("#nalisa Data Train")</pre>
Out[73]:	Text(0.5, 1.0, 'Analisa Data Train')
	1e9 Analisa Data Train
	10112
	10110
	E 10108 -
	8, 10106 -
	10104 -
	10102 -
	1950 1960 1970 1980 1990 2000 2010 2020 Tahun

Figure 2: Fitting Simple Linear Regression in Training Set

Creating value coefficient and the intercept on predictive data tabulation.

In [60]: M	<pre>N # The coefficients print ('coefficients: ', regr.coef_) print ('Intercept: ',regr.intercept_)</pre>						
	Coefficients: [[-304,99274936]] Intercept: [1.0116371e+09]						
	Figure 3: Coefficient and Intercept						

Coefficients and interception in a simple linear regression fit the parameters of the line. Given that this is a simple linear regression, with only two parameters, and knowing that the parameter is the intercept and the slope of the line, can sklearn direct estimate of the data.

c. Predicting Results of Test-Set

And to know the quality of the data is carried out also for testing against test data

Figure 4: Predicting Result And included the form of the plot





3.2 K-Means Method:

a. Screening Data: Before using the code for Data Processing, first input library that will be used:

1. Seleksi Data (Selection Data)

Data yang diambil google drive dan menggunakan colab dari google. Data menggunakan authorization code untuk masuk kedalam drive

[12] 1 import pandas as pd 2 import numpy as np

[13] 1 from google.colab import drive 2 drive.mount('/content/drive')

D Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount

[14] 1 data = pd.read_csv('/content/drive/My Drive/Project/Data/Data_aset_II_3.csv', sep=';')

Figure 6: Screening Data

b. Selecting Data

Before the data were used to do the cleansing / disaggregation that will grab

2. Pemilihan Data (Preprocessing/Cleaning)

Proses Preprocessing mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

[15]	1	data.head()						
C•		Kode_Barang	KIB	Jenis_Barang	Ukuran	Satuan	Alamat	Tahun
	0	1011104002	KIB D	Tanah Bangunan Pendidikan Dan Latihan (sekolah)	5.19	M2	JI. Cikini Raya No 87	1950
	1	1011104002	KIB D	Tanah Bangunan Pendidikan Dan Latihan (sekolah)	1.80	M2	JALAN PERUK NO. 32	1951
	2	1011104002	KIB D	Tanah Bangunan Pendidikan Dan Latihan (sekolah)	4.18	M2	JI.Perwira No.10 Rt.002 / Rw.008	1951
	3	1011104002	KIB D	Tanah Bangunan Pendidikan Dan Latihan (sekolah)	9.70	M2	JL PALMERAH BARAT NO. 59	1959
	4	1011104002	KIB D	Tanah Bangunan Pendidikan Dan Latihan (sekolah)	775.00	M2	JI. Masjid Nur No.33 Rt.002 / 010	1962
[57]	1 2	# Mengecek ap data.empty	akah a	da deret yang kosong				
C•	Fal	se						
[58]	1 2	# Melihat uku data.size	iran da	ri data				
C•	700	0						

Figure 7: Selecting Data

c.Creating a Data Transformation

```
[59] 1 # Menetapkan Variabel inde
     2 x = data.drop(['Tahun', 'KIB', 'Satuan','Jenis_Barang', 'Alamat'], axis = 1)
     3 x.head()
D
       Kode_Barang Ukuran
     0 1011104002 5.19
         1011104002
                      1.80
     2 1011104002 4.18
     3 1011104002 9.70
     4 1011104002 775.00
[60] 1 # Menetapkan Variabel independen
      3 y = data['Tahun']
      4 y.head()
C• °
         1950
          1951
         1959
          1962
     Name: Tahun, dtvpe: int64
```

Figure 8: Creating Data Transformation

Classification Data K-Means Method

```
1. Determining the predictive data
```

```
[31] 1 # Menentukan hasil prediksi dari x_test
2 y_pred = nbtrain.predict(x_test)
3 y_pred
```

r, array([1986, 2010, 1986, 1986, 1986, 1986, 1982, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1982, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1990, 1986, 1986, 1986, 1986, 1990, 1986, 1986, 1986, 1982, 1986, 1990, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1990, 1986, 1982, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1986, 1990, 1986, 1986, 1990, 1986])

Figure 9: Determining Predictive Data

2. Determine the probability of data

[32] 1 # Menentukan probabilitas hasil prediksi 2 nbtrain.predict proba(x test)

C,	array([[0.00501279, 0.01233	699, 0.00127184	·,, (). ,	0.01218608
	0.01033674], [0.00566488_0.01196	87 A AA196A36		1	0 01160236
	0.01085605],	02 , 0.0010000	,, .	,, ,	0.01100050
	[0.00493076, 0.01237	628, 0.00120303	,, (). ,	0.01225602
	0.01027391],				
	[0.00636245, 0.01145 0.01146746],	526, 0.00309225	,, (), ,	0.01093496
	[0.00538868, 0.01213 0.01063127].	673, 0.00163482	,, (). ,	0.01185809
	[0.00725134, 0.01056 0.01239659]])	22 , 0.00567666	,, (). ,	0.00992861
	0.017220223]])				

Figure 10: Determine the probability data

3. Determining Matrix Model:

```
[33] 1 # import confusion_matrix model
2 from sklearn.metrics import confusion_matrix
3 confusion_matrix(y_test, y_pred)
```

```
[→ array([[0, 0, 0, ..., 0, 0, 0, 0],

[0, 0, 0, ..., 0, 0, 0],

[0, 0, 0, ..., 0, 0, 0],

...,

[0, 0, 0, ..., 0, 0, 0],

[0, 0, 0, ..., 0, 0, 0],

[0, 0, 0, ..., 0, 0, 0]])
```

```
[37] 1 # Merapikan hasil confusion matrix
    2 y_actual1 = pd.series([1, 0,1,0,1,0,1,0,1,0,0,1,1,0,0], name ='actual')
    3 y_pred1 = pd.series([1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1], name ='prediction')
    4 df_confusion = pd.crosstab(y_actual1, y_pred1)
    5 df_confusion
    6
    prediction 0 1
    actual
```

0	7	2	
1	1	8	

Figure 11: Determining Matrix Model

d. Specifying Process Data Mining: Apply an algorithm to classify the data.

[29]	1 # Import train_test 2 from sklearn.model_ 3 x_train, x_test, y_	t_split function selection import trai train, y_test = train	n_test_split _test_split(x, y,	test_size = 0.2,	random_state = 123)
[30]	1 # Import Gaussian H 2 from sklearn.naive 3 4 # Mengaktifkan/mem 5 modelnb = Gaussian 6 7 # Menasukkan data H 8 nbtrain = modelnb. 9 nbtrain.class_count	Naive Bayes model bayes import Gaussian anggil/membuat fungsi N NB() training pada fungsi k fit(x_train, y_train) t_	NB klasifikasi Naive lasifikasi naive b	bayes Iayes	
Ŀ	array([1., 2., 1 5., 6., 5 44., 20., 5 26., 16., 1 1., 5., 5	l., 1., 4., 1., 5., 6., 9., 11., 5., 3., 1., 6., 8., 4., 4., 9., 8., 3., 1., 2.,	2., 2., 2., 10., 83., 234., 7., 4., 1., 6., 3., 2., 2.])	6., 3., 93., 64., 11., 26., 7., 24.,	
	1962 1963 1975 1976 1977 1978 1979 1981	precision 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.	recall f	1-score 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.	support 2 1 3 2 3 3 2 3

		precision	recall	f1-score	support
1	962	0.00	0.00	0.00	2
1	963	0.00	0.00	0.00	1
1	975	0.00	0.00	0.00	1
1	976	0.00	0.00	0.00	3
1	977	0.00	0.00	0.00	2
1	978	0.00	0.00	0.00	3
1	979	0.00	0.00	0.00	3
1	981	0.00	0.00	0.00	2
1	982	0.00	0.00	0.00	1
1	983	0.00	0.00	0.00	27
1	984	0.00	0.00	0.00	56
1	985	0.00	0.00	0.00	11
1	986	0.08	1.00	0.16	16
1	987	0.00	0.00	0.00	10
1	988	0.00	0.00	0.00	5
1	989	0.00	0.00	0.00	3
1	990	0.00	0.00	0.00	0
1	991	0.00	0.00	0.00	2
1	993	0.00	0.00	0.00	1
1	994	0.00	0.00	0.00	4
1	996	0.00	0.00	0.00	5
1	997	0.00	0.00	0.00	6
1	998	0.00	0.00	0.00	9
1	999	0.00	0.00	0.00	2
2	000	0.00	0.00	0.00	2
2	001	0.00	0.00	0.00	1
2	003	0.00	0.00	0.00	3
2	004	0.00	0.00	0.00	8
2	005	0.00	0.00	0.00	2
2	008	0.00	0.00	0.00	5
2	010	0.00	0.00	0.00	3
2	014	0.00	0.00	0.00	1
accur	acy			0.08	200
macro	avg	0.00	0.03	0.00	200
eighted a	avg	0.01	0.08	0.01	200

Figure 12: Specifying Process Data Mining

- [56] 1 plt.figure(figsize = (17,6))
 2 plt.scatter(y_test,y_pred, c='red', s=300, alpha=0.1 , marker="0")
 - 3 plt.xticks(data['Tahun'],rotation=45)
 - 4 plt.xlabel('Tahun',fontsize=18)
 - 5 plt.ylabel('Tahun', fontsize=18)
 - 6 plt.show()



Figure 13: K-Means

e. Application of Interpretation / Evaluation

[38] 1 # Menghitung nilai akurasi dari klasifikasi naive bayes 2 from sklearn.metrics import classification_report

3 print(classification_report(y_test,y_pred))

Figure 14: Interpretation/Evaluation

4. CONCLUSION

The k-means method succeeded in making the accuracy of the data classification of DKI Jakarta Education authorities asset beneficiary percentage accuracy of 71.42%.

REFERENCE

- 1. X. Wu and V. Kumar, eds., The Top Ten Algorithms in Data Mining.Chapman and Hall, 2009.
- S. Russell and P. Norvig, Artificial Intelligence A Modern Approach. Upper Saddle River, New Jersey 07458: Pearson Education, Inc., 3 ed., 2010.
- Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, (First Edition). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.
- O. Maimon and L. Rokach, Data Mining and Knowledge Discovery Handbo- ok. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.
- Nia Rahma Kurnianda & Yunita Sartika Sari. Analysis and Design of Information System for Self-Journal on Food Based Dietary Assessment Record for Diabetes Patients. International Research Journal of Computer Science (IRJCS). Volume 06 Issue 5. 2018
- Ranggadara, Indra & Suhendra. Zachman Framework Approach for Designing Recruitment System Modules in HRIS Application (Case Study in PT. Karya Impian Teknologi Abadi).IJCSMC, Vol 7, issue 2.2018