

Pokok Bahasan I

Data Pre-processing

Kode Pokok Bahasan: TIK.RPL03.001.001.01

Deskripsi Pokok Bahasan:

Membahas bagaimana pengolahan data sebelum diproses menggunakan bahasa R. Mempersiapkan sebuah data mentah, agar dapat sesuai dengan spesifikasi untuk tahap processing data.

No	Elemen Kompetensi	Indikator Kinerja	Jml Jam	Hal
1	Memahami data pre-processing.	1.1 Menganalisa data pre-processing dari data yang disediakan	1	5
2	Mengimplementasikan pre-processing data.	1.1 Mampu memproses data dengan Missing Values 1.2 Mampu mengidentifikasi nilai NA disetiap column dan menghapus baris yang memiliki nilai NA 1.3 Mampu memvisualisasikan data dengan histogram 1.4 Mampu membagi data menjadi training dan test dataset. 1.5 Mampu menerapkan feature scaling pada data.	2	6

TUGAS PENDAHULUAN

Hal yang harus dilakukan dan acuan yang harus dibaca sebelum praktikum :

1. Menginstal R pada PC masing-masing praktikan.
2. Menginstal R Studio pada PC masing-masing praktikan.



DAFTAR PERTANYAAN

1. Apa yang dimaksud dengan data pre-processing?

Data preprocessing adalah proses yang mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Proses ini penting dilakukan karena data mentah sering kali tidak memiliki format yang teratur.

2. Mengapa perlu adanya tahapan pre-processing?

Preprocessing data penting untuk dilakukan karena dapat memberikan fungsi atau manfaat pada data mining. Proses ini utamanya dilakukan untuk memastikan kualitas data baik sebelum digunakan saat analisis data.

3. Sebutkan langkah-langkah dari data pre-processing?

1. Data cleaning.
2. Data integration.
3. Data transformation.
4. Data reduction.

TEORI SINGKAT

Data Preparation atau bisa disebut juga dengan data preprocessing adalah suatu proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas(input yang baik untuk data mining tools). Tahapan dalam melakukan data mining salah satunya adalah preprosesing data. Pertanyaannya adalah mengapa data perlu di bersihkan sebelum diproses? Hal ini terjadi karena biasanya data yang akan digunakan belum baik, penyebabnya antara lain :

- **Incomplete** : kekurangan nilai-nilai atribut atau atribut tertentu lainnya.
- **Noisy** : berisi kesalahan atau nilai-nilai outlier yang menyimpang dari yang diharapkan.
- **Inconsisten** : ketidakcocokan dalam penggunaan kode atau nama.

Disini kualitas data yang baik didasarkan oleh keputusan yang baik dan data warehouse memerlukan integrasi kualitas data yang konsisten. Beberapa hal yang perlu diperhatikan untuk mendapatkan data yang baik adalah :

- Accuracy**
- Completeness**
- Consistency**



- Timeliness**
- **Value added**
- Interpretability**
- Accessibility**
- Contextual**
- **Representational**

LAB SETUP

Hal yang harus disiapkan dan dilakukan oleh praktikan untuk menjalankan praktikum modul ini.

1. Menginstall library yang dibutuhkan untuk mengerjakan modul.
2. Menjalankan R Studio.

ELEMEN KOMPETENSI I

Deskripsi:

Memahami data pre-processing.

Kompetensi Dasar:

1. Menganalisa data pre-processing dari data yang telah disediakan.

Latihan 1.1.1

Penjelasan Singkat :

Pada latihan ini anda akan diminta untuk melakukan analisis terhadap data yang telah disediakan.

Apakah data yang disediakan, termasuk data yang perlu di pre-processing? Jika ya, jelaskan alasannya!

-Data yang disediakan adalah data Train,Test,Dataset

-Seluruh data harus di processing karena kita sendiri harus mengisi data-data yang NA(Kosong)

ELEMEN KOMPETENSI 2

Deskripsi:

Mengimplementasikan pre-processing data.

Kompetensi Dasar:

1. Mampu memproses data dengan Missing Values
2. Mampu membagi data menjadi training dan test dataset.
3. Mampu menerapkan feature scaling pada data.



Latihan 1.2.1

Penjelasan Singkat:

Pada latihan ini praktikan diminta untuk memproses missing value dari data yang telah disediakan. Bagaimana cara agar missing value tersebut tidak menjadi kendala dalam tahap processing.

Langkah-Langkah Praktikum:

1. Impor data train.csv dan test.csv dari <https://www.kaggle.com/c/titanic/data>.
2. Baca file excel dengan perintah :

```
> dataku_namapraktikan=read.csv("E:/Kuliah Data Mining gasal 1920/train.csv").
```

3. Lihat 5 data teratas dengan perintah :

```
> head(dataku_namapraktikan)
```

Output :

```
> head(dataku_ibnu)
  PassengerId Survived Pclass      Name Sex Age SibSp Parch
1          1         0       3 Braund, Mr. Owen Harris male 22     1     0
2          2         1       1 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female 38     1     0
3          3         1       3 Heikkinen, Miss. Laina female 26     0     0
4          4         1       1 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35     1     0
5          5         0       3 Allen, Mr. William Henry male 35     0     0
6          6         0       3 Moran, Mr. James male NA     0     0
  Ticket   Fare Cabin Embarked
1   A/5 21171  7.2500      S
2   PC 17599 71.2833   C85      C
3 STON/O2. 3101282 7.9250      S
4  113803 53.1000   C123      S
5  373450  8.0500      S
6  330877  8.4583      Q
```

4. Jalankan perintah berikut :

```
> library("readxl")
> dataku_namapraktikan <- read_excel("E:/Kuliah Data Mining gasal 1920/train.xlsx")
> head(dataku_namapraktikan)
```



5. Apakah baris ke-6 mengandung missing value pada variabel Age ? Y/T

```
> str(dataku_namapraktikan)
```

Output :

```
> str(dataku_ibnu)
tibble [891 x 12] (S3: tbl_df/tbl/data.frame)
 $ PassengerId: num [1:891] 1 2 3 4 5 6 7 8 9 10 ...
 $ Survived   : num [1:891] 0 1 1 1 0 0 0 0 1 1 ...
 $ Pclass     : num [1:891] 3 1 3 1 3 3 1 3 3 2 ...
 $ Name       : chr [1:891] "Braund, Mr. Owen Harris" "Cumings, Mrs. John Bradley (Florence Briggs Thayer)" "Heikkinen
Miss. Laina" "Futrelle, Mrs. Jacques Heath (Lily May Peel)" ...
 $ Sex        : chr [1:891] "male" "female" "female" "female" ...
 $ Age        : num [1:891] 22 38 26 35 35 NA 54 2 27 14 ...
 $ SibSp      : num [1:891] 1 1 0 1 0 0 0 3 0 1 ...
 $ Parch      : num [1:891] 0 0 0 0 0 0 0 1 2 0 ...
 $ Ticket     : chr [1:891] "A/5 21171" "PC 17599" "STON/O2. 3101282" "113803.0" ...
 $ Fare       : num [1:891] 7.25 71.28 7.92 53.1 8.05 ...
 $ Cabin      : chr [1:891] NA "C85" NA "C123" ...
 $ Embarked   : chr [1:891] "S" "C" "S" "S" ...
```

6. Membuat data frame dengan logika nilai True and False untuk mengindikasi Missing Value

```
> is.na(dataku_namapraktikan)
```

output :



```
Console Terminal Jobs
R 4.1.1 ~ /
[57,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[58,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[59,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[60,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[61,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[62,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE
[63,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[64,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[65,] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE
[66,] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE
[67,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[68,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[69,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[70,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[71,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[72,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[73,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[74,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[75,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[76,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[77,] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE
[78,] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE
[79,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[80,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[81,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[82,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
[83,] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE
[ reached getOption("max.print") -- omitted 808 rows ]
>
```

> apply(is.na(dataku_namapraktikan), 2, which)

output :

```
> apply(is.na(dataku_ibnu), 2, which)
$PassengerId
integer(0)

$Survived
integer(0)

$Pclass
integer(0)

$Name
integer(0)

$Sex
integer(0)

$Age
[1] 6 18 20 27 29 30 32 33 37 43 46 47 48 49 56 65 66 77 78 83 88 96 102 108 110 122 127
[28] 129 141 155 159 160 167 169 177 181 182 186 187 197 199 202 215 224 230 236 241 242 251 257 261 265 271 275
[55] 278 285 296 299 301 302 304 305 307 325 331 335 336 348 352 355 359 360 365 368 369 376 385 389 410 411 412
[82] 414 416 421 426 429 432 445 452 455 458 460 465 467 469 471 476 482 486 491 496 498 503 508 512 518 523 525
[109] 528 532 534 539 548 553 558 561 564 565 569 574 579 585 590 594 597 599 602 603 612 613 614 630 634 640 644
[136] 649 651 654 657 668 670 675 681 693 698 710 712 719 728 733 739 740 741 761 767 769 774 777 779 784 791 793
[163] 794 816 826 827 829 833 838 840 847 850 860 864 869 879 889
```



\$Cabin	1	3	5	6	8	9	10	13	14	15	16	17	18	19	20	21	23	25	26	27	29	30	31	33	34	35	36
[1]	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	54	57	58	59	60	61	64	65	66	68	69
[55]	70	71	72	73	74	75	77	78	79	80	81	82	83	84	85	86	87	88	90	91	92	94	95	96	99	100	101
[82]	102	104	105	106	107	108	109	110	112	113	114	115	116	117	118	120	121	122	123	126	127	128	130	131	132	133	134
[109]	135	136	139	141	142	143	144	145	146	147	148	150	151	153	154	155	156	157	158	159	160	161	162	163	164	165	166
[136]	168	169	170	172	173	174	176	177	179	180	181	182	183	185	187	188	189	190	191	192	193	197	198	199	200	201	202
[163]	203	204	205	207	208	209	211	212	213	214	215	217	218	220	221	222	223	224	226	227	228	229	230	232	233	234	235
[190]	236	237	238	239	240	241	242	243	244	245	247	248	250	251	254	255	256	257	259	260	261	262	265	266	267	268	271
[217]	272	273	275	277	278	279	280	281	282	283	284	286	287	288	289	290	291	294	295	296	297	301	302	303	305	307	309
[244]	313	314	315	316	317	318	321	322	323	324	325	327	329	331	334	335	336	339	343	344	345	347	348	349	350	351	353
[271]	354	355	356	358	359	360	361	362	363	364	365	366	368	369	372	373	374	375	376	377	379	380	381	382	383	384	385
[298]	386	387	388	389	390	392	393	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	414	415	416
[325]	417	418	419	420	421	422	423	424	425	426	427	428	429	432	433	434	437	438	440	441	442	443	444	445	447	448	449
[352]	451	452	455	456	459	460	462	464	465	466	467	468	469	470	471	472	473	475	477	478	479	480	481	482	483	484	486
[379]	489	490	491	492	494	495	496	498	500	501	502	503	504	507	508	509	510	511	512	514	515	518	519	520	522	523	525
[406]	526	527	529	530	531	532	533	534	535	536	538	539	542	543	544	546	547	548	549	550	552	553	554	555	556	558	560
[433]	561	562	563	564	565	566	567	568	569	570	571	574	575	576	577	579	580	581	583	585	587	589	590	591	593	594	595
[460]	596	597	598	599	601	602	603	604	605	606	607	608	609	611	612	613	614	615	616	617	618	620	621	623	624	625	627
[487]	629	630	632	634	635	636	637	638	639	640	641	643	644	645	647	649	650	651	652	653	654	655	656	657	658	659	661
[514]	662	664	665	666	667	668	669	671	673	674	675	676	677	678	679	681	683	684	685	686	687	688	689	692	693	694	695
[541]	696	697	698	703	704	705	706	707	709	710	714	715	719	720	721	722	723	724	726	727	728	729	730	732	733	734	735
[568]	736	737	739	740	744	745	747	748	750	751	753	754	755	756	757	758	759	761	762	763	765	767	768	769	770	771	772
[595]	774	775	776	778	779	781	784	785	786	787	788	789	791	792	793	794	795	796	798	799	800	801	802	804	805	806	808
[622]	809	811	812	813	814	815	817	818	819	820	822	823	825	826	827	828	829	831	832	833	834	835	837	838	839	841	842
[649]	843	844	845	846	847	848	849	851	852	853	855	856	857	859	860	861	862	864	865	866	867	869	870	871	874	875	876
[676]	877	878	879	881	882	883	884	885	886	887	889	891															
\$Embarked																											
[1]	62	830																									

7. Selanjutnya Mengatasi missing value dengan cara menghapus baris yang yang terdapat missing value

```
> data_drop = na.omit(dataku_namapraktikan)
```

output :

```
> data_drop = na.omit(dataku_ibnu)
> view(data_drop)
```

```
> view(data_drop)
```

output :



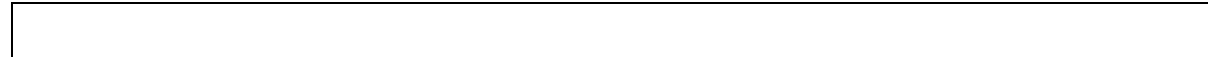
PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.00	1	0	PC 17599	71.2833
2	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.00	1	0	113803.0	53.1000
3	7	0	McCarthy, Mr. Timothy J	male	54.00	0	0	17463.0	51.8625
4	11	1	Sandstrom, Miss. Marguerite Rut	female	4.00	1	1	PP 9549	16.7000
5	12	1	Bonnell, Miss. Elizabeth	female	58.00	0	0	113783.0	26.5500
6	22	1	Beesley, Mr. Lawrence	male	34.00	0	0	248698.0	13.0000
7	24	1	Sloper, Mr. William Thompson	male	28.00	0	0	113788.0	35.5000
8	28	0	Fortune, Mr. Charles Alexander	male	19.00	3	2	19950.0	263.0000
9	53	1	Harper, Mrs. Henry Sleeper (Myna Haxtun)	female	49.00	1	0	PC 17572	76.7292
10	55	0	Ostby, Mr. Engelhart Cornelius	male	65.00	0	1	113509.0	61.9792
11	63	0	Harris, Mr. Henry Birkhardt	male	45.00	1	0	36973.0	83.4750
12	67	1	Nye, Mrs. (Elizabeth Ramell)	female	29.00	0	0	C.A. 29395	10.5000
13	76	0	Moen, Mr. Sigurd Hansen	male	25.00	0	0	348123.0	7.6500
14	89	1	Fortune, Miss. Mabel Helen	female	23.00	3	2	19950.0	263.0000
15	93	0	Chaffee, Mr. Herbert Fuller	male	46.00	1	0	W.E.P. 5734	61.1750

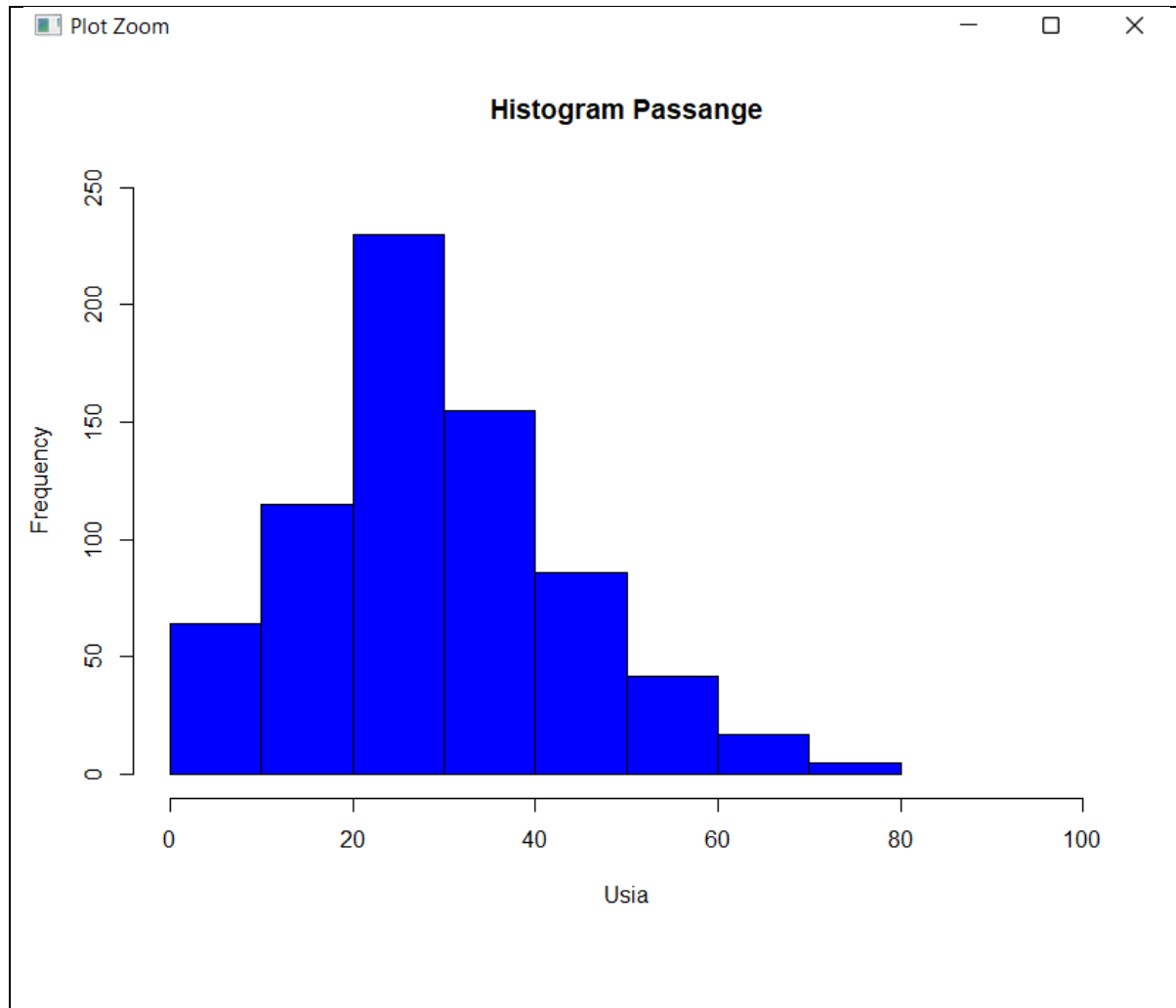
Pastika sudah tidak ada missingvalue pada table

8. Buat Histogram berdasarkan Usia

```
> hist(dataku_namapraktikan $Age, main = "Histogram Passange",
+      xlab = "Usia",xlim=c(0,100), ylim = c(0,250), col="blue")
```

output :

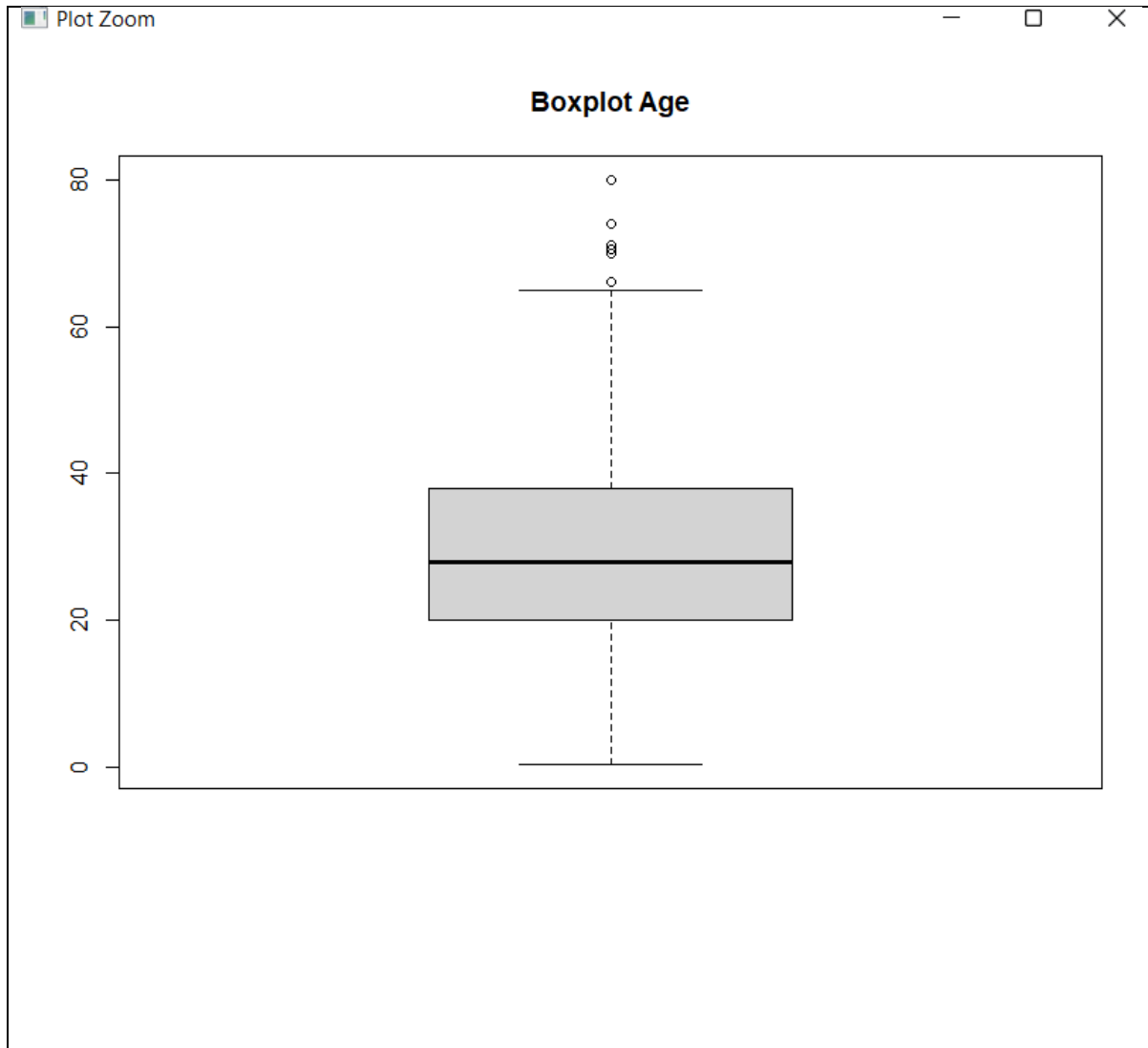




```
> boxplot(dataku$Age, main = "Boxplot Age")
```

output :





9. Mengatasi missing value dengan cara mengganti nilai missing value dengan mean

```
> dataku_namapraktikan=read.csv("E:/Kuliah Data Mining gasal 1920/train.csv").
```

output :



```
> mean(dataku_namapraktikan$Age)
```

Output :



```
> mean(dataku_ibnu$Age)
[1] NA
```

```
> mean(dataku_namapraktikan$Age, na.rm = 'TRUE')
```

Output :

```
> mean(dataku_ibnu$Age, na.rm = 'TRUE')
[1] 29.69912
```

```
> dataku_namapraktikan$Age = ifelse(is.na(dataku_namapraktikan$Age), ave(dataku_namapraktikan$Age, FUN = function(x) mean(x, na.rm = 'TRUE')), dataku_namapraktikan$Age)
> head(dataku_namapraktikan)
```

Output :

```
> dataku_ibnu$Age = ifelse(is.na(dataku_ibnu$Age), ave(dataku_ibnu$Age, FUN = function(x) mean(x, na.rm = 'TRUE')), dataku_ibnu$Age)
> head(dataku_ibnu)
# A tibble: 6 x 12
  PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked
  <dbl> <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl> <chr> <chr>
1 1 0 3 Braund, Mr. Owen Ha~ male 22 1 0 A/5 21171 7.25 NA S
2 2 1 1 Cumings, Mrs. John ~ fema~ 38 1 0 PC 17599 71.3 C85 C
3 3 1 3 Heikkinen, Miss. La~ fema~ 26 0 0 STON/O2.~ 7.92 NA S
4 4 1 1 Futrelle, Mrs. Jacq~ fema~ 35 1 0 113803.0 53.1 C123 S
5 5 0 3 Allen, Mr. William ~ male 35 0 0 373450.0 8.05 NA S
6 6 0 3 Moran, Mr. James male 29.7 0 0 330877.0 8.46 NA Q
```

Apakah variabel Age baris ke-6 masih mengandung missing value ? Y/T Y

Latihan :



	nation	purchased_item	age	salary
1	India	No	25	35000
2	Russia	Yes	NA	40000
3	Germany	No	50	54000
4	Russia	No	35	NA
5	Germany	Yes	40	60000
6	India	Yes	35	58000
7	Russia	No	NA	52000
8	India	Yes	48	NA
9	Germany	No	50	83000
10	India	Yes	37	NA
11	Germany	No	21	24000
12	India	Yes	NA	60000
13	Russia	No	63	70000
14	Germany	yes	26	36000
15	India	No	45	40000

Lakukan penanganan missing value pada variabel age dan salary sesuai langkah-langkah yang telah dipelajari :

Output :

```
> head(dataku_ibnu)
# A tibble: 6 x 4
  nation purchased_item age salary
  <chr>   <chr>          <dbl> <dbl>
1 india   no             25   35000
2 russia  yes             NA   40000
3 germany no             50   54000
4 russia  no             35    NA
5 germany yes             40   60000
6 india   yes             35   58000
>
> str(dataku_ibnu)
tibble [15 x 4] (S3: tbl_df/tbl/data.frame)
 $ nation      : chr [1:15] "india" "russia" "germany" "russia" ...
 $ purchased_item: chr [1:15] "no" "yes" "no" "no" ...
 $ age         : num [1:15] 25 NA 50 35 40 35 NA 48 50 37 ...
 $ salary      : num [1:15] 35000 40000 54000 NA 60000 58000 52000 NA 83000 NA ...
> |
```



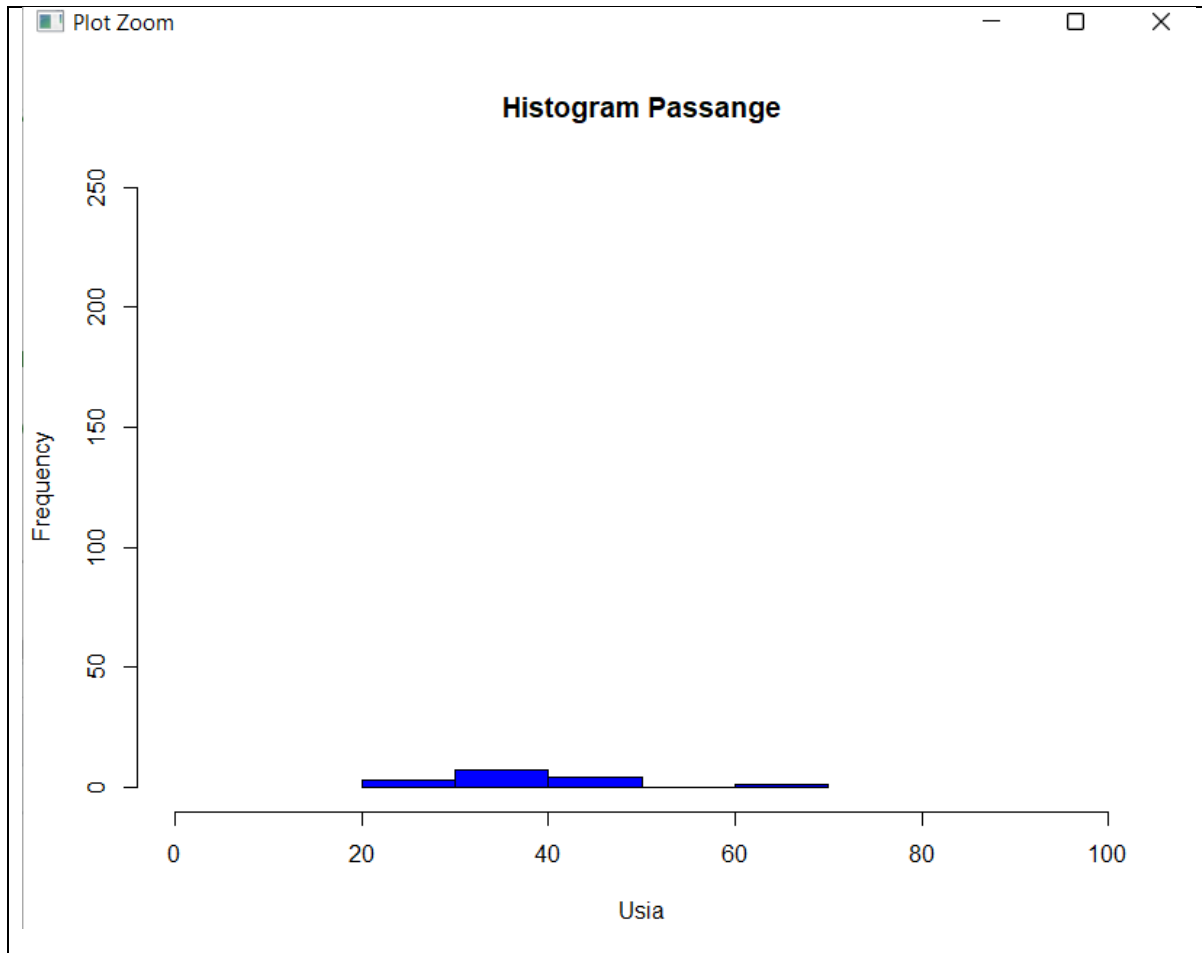
```
> apply(is.na(dataku_ibnu), 2, which)
$nation
integer(0)

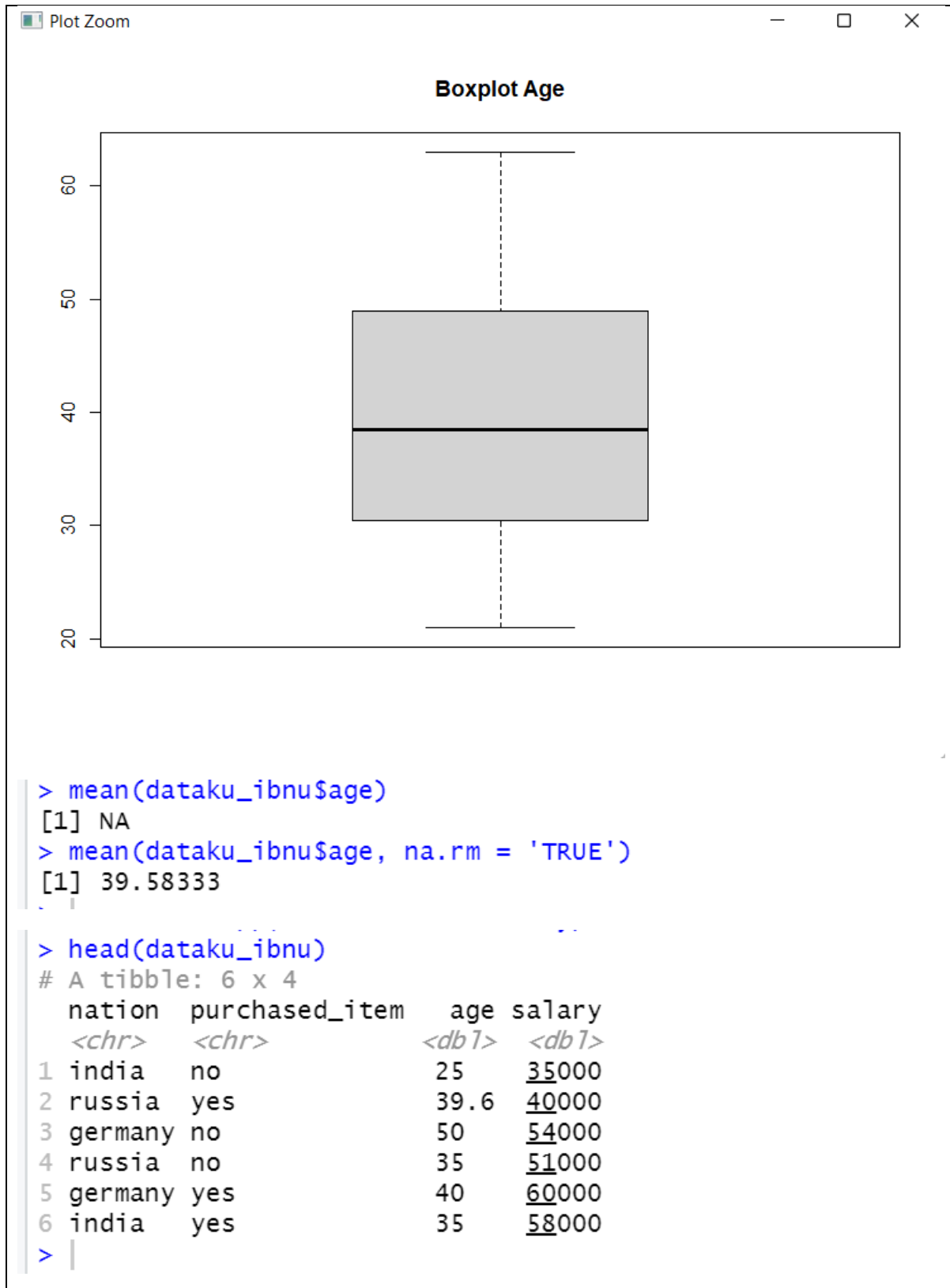
$purchased_item
integer(0)

$age
[1] 2 7 12

$salary
[1] 4 8 10
```

	▲ nation ▼	purchased_item ▼	age ▼	salary ▼
1	india	no	25	35000
2	germany	no	50	54000
3	germany	yes	40	60000
4	india	yes	35	58000
5	germany	no	50	83000
6	germany	no	21	24000
7	russia	no	63	70000
8	germany	yes	26	36000
9	india	no	45	40000





Latihan 1.2.2

Penjelasan Singkat:

Pada latihan ini praktikan diminta untuk mampu membagi data menjadi training dan test dataset.

Langkah-Langkah Praktikum:

1. Gunakan dataaset seperti berikut :

Dataset

##	Country	Age	Salary	Purchased
## 1	France	44	72000	No
## 2	Spain	27	48000	Yes
## 3	Germany	30	54000	No
## 4	Spain	38	61000	No
## 5	Germany	40	NA	Yes
## 6	France	35	58000	Yes
## 7	Spain	NA	52000	No
## 8	France	48	79000	Yes
## 9	Germany	50	83000	No
## 10	France	37	67000	Yes

2. Ganti nilai NA dengan nilai average dari kolom terkait!

```
> install.packages("catTools")
> library(caTools) #adding caTools to the library
> set.seed(123)
> split = sample.split(dataset$Purchased, SplitRatio = 0.8)
> training_set = subset(dataset, split == TRUE)
> test_set = subset(dataset, split == FALSE)
```

3. Tampilkan 5 data teratas

```
> head (training_set)
```

output :




```
> head(training_set)
# A tibble: 6 x 4
  country    age salary purchased
  <chr>    <dbl> <dbl> <chr>
1 france    44  72000 no
2 spain     27  48000 yes
3 germany   30  54000 no
4 spain     38  61000 no
5 germany   40    NA yes
6 spain     NA  52000 no
```

```
> head(test_set)
```

output :

```
> head(test_set)
# A tibble: 2 x 4
  country    age salary purchased
  <chr>    <dbl> <dbl> <chr>
1 france    35  58000 yes
2 germany   50  83000 no
```

Latihan 1.2.3

Penjelasan Singkat:

Pada latihan ini praktikan diminta untuk mampu melakukan featurig scaling pada data uji.

Langkah-Langkah Praktikum:

1. Lakukan feature scaling pada training_set dan test_set

```
#feature scaling
```



```
training_set[,2:3] = scale(training_set[,2:3])  
test_set[,2:3] = scale(test_set[,2:3])
```

output :

```
> head (training_set)  
# A tibble: 6 x 4  
  country    age  salary purchased  
  <chr>    <dbl>  <dbl> <chr>  
1 france    0.853    0.893   no  
2 spain    -1.45    -1.22   yes  
3 germany -1.05    -0.691  no  
4 spain     0.0388 -0.0754  no  
5 germany  0.310    NA      yes  
6 spain     NA      -0.867  no  
#>  
  
> head(test_set)  
# A tibble: 2 x 4  
  country    age salary purchased  
  <chr>    <dbl>  <dbl> <chr>  
1 france  -0.707 -0.707  yes  
2 germany  0.707  0.707  no
```

2. Ambil dataku_namapraktikan dan lakukan fungsi scale.

```
> training_namapraktikan<-dataku_namapraktikan[,c(2,3)]
```

```
> scale_namapraktikan <- scale(training_namapraktikan)
```

output :



Elemen Kompetensi	No Latihan	Penyelesaian	
		Selesai	Tidak selesai
1	1.1.1	✓	
2	1.2.1	✓	
	1.2.2	✓	
	1.2.3	✓	



FORM UMPAN BALIK

Elemen Kompetensi	Tingkat Kesulitan	Tingkat Ketertarikan	Waktu Penyelesaian dalam menit
Memahami data pre-processing.	<input type="checkbox"/> Sangat Mudah <input type="checkbox"/> Mudah <input checked="" type="checkbox"/> Biasa <input type="checkbox"/> Sulit <input type="checkbox"/> Sangat Sulit	<input type="checkbox"/> Tidak Tertarik <input type="checkbox"/> Cukup Tertarik <input type="checkbox"/> Tertarik <input checked="" type="checkbox"/> Sangat Tertarik	30
Mengimplementasikan pre-processing data.	<input type="checkbox"/> Sangat Mudah <input type="checkbox"/> Mudah <input checked="" type="checkbox"/> Biasa <input type="checkbox"/> Sulit <input type="checkbox"/> Sangat Sulit	<input type="checkbox"/> Tidak Tertarik <input type="checkbox"/> Cukup Tertarik <input type="checkbox"/> Tertarik <input checked="" type="checkbox"/> Sangat Tertarik	30

