Feature Selection : Missing Value Ratio

Common Dimensionality Reduction Techniques

- Missing value ratio
- Low variance
- High correlation
- Backward feature elimination
- Forward feature selection

Common Dimensionality Reduction Techniques

- Missing value ratio
- Low variance
- High correlation
- Backward feature elimination
- Forward feature selection

Feature Selection : Missing Value Ratio

ID	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	count
AB101	1.0	0.0	0.0	1.0	9.84	14.395	81.0	NaN	16
AB102	1.0	NaN	0.0	NaN	9.02	13.635	80.0	NaN	40
AB103	1.0	0.0	NaN	1.0	9.02	13.635	80.0	NaN	32
AB104	NaN	0.0	NaN	1.0	9.84	14.395	75.0	NaN	13
AB105	1.0	NaN	0.0	NaN	9.84	14.395	NaN	16.9979	1
AB106	1.0	0.0	NaN	2.0	9.84	12.880	75.0	NaN	1
AB107	1.0	0.0	0.0	1.0	9.02	13.635	80.0	NaN	2
AB108	1.0	NaN	0.0	1.0	8.20	12.880	86.0	NaN	3
AB109	NaN	0.0	0.0	NaN	9.84	14.395	NaN	NaN	8
AB110	1.0	0.0	0.0	1.0	13.12	17.425	76.0	NaN	14

Analytics Vidhya
Learn everything about analytics

Feature Selection : Missing Value Ratio

ID	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	count
AB101	1.0	0.0	0.0	1.0	9.84	14.395	81.0	NaN	16
AB102	1.0	NaN	0.0	NaN	9.02	13.635	80.0	NaN	40
AB103	1.0	0.0	NaN	1.0	9.02	13.635	80.0	NaN	32
AB104	NaN	0.0	NaN	1.0	9.84	14.395	75.0	NaN	13
AB105	1.0	NaN	0.0	NaN	9.84	14.395	NaN	16.9979	1
AB106	1.0	0.0	NaN	2.0	9.84	12.880	75.0	NaN	1
AB107	1.0	0.0	0.0	1.0	9.02	13.635	80.0	NaN	2
AB108	1.0	NaN	0.0	1.0	8.20	12.880	86.0	NaN	3
AB109	NaN	0.0	0.0	NaN	9.84	14.395	NaN	NaN	8
AB110	1.0	0.0	0.0	1.0	13.12	17.425	76.0	NaN	14

Analytics Vidhya
Learn everything about analytics

Feature Selection : Missing Value Ratio

Ratio of missing values

Feature Selection : Missing Value Ratio

$$\text{Ratio of missing values} = \frac{\text{Number of missing values}}{\text{Total number of observations}} * 100$$

Feature Selection : Missing Value Ratio

Variable	Missing value ratio
ID	0%
season	20%
holiday	30%
workingday	30%
weather	30%
temp	0%
atemp	0%
humidity	20%
windspeed	90%
count	0%

Feature Selection : Missing Value Ratio

Analytics Vidhya
Learn everything about analytics

Feature Selection : Missing Value Ratio

Decide a threshold

Feature Selection : Missing Value Ratio

Decide a threshold 70%

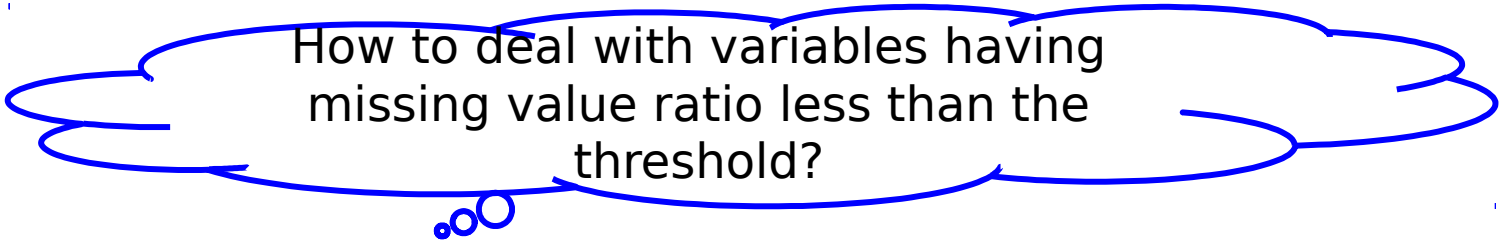
Feature Selection : Missing Value Ratio

Variable	Missing value ratio
ID	0%
season	20%
holiday	30%
workingday	30%
weather	30%
temp	0%
atemp	0%
humidity	20%
windspeed	90%
count	0%

Feature Selection : Missing Value Ratio

Guideline : Can drop a variable having missing value ratio more than 60 - 70%

Feature Selection : Missing Value Ratio



Analytics Vidhya
Learn everything about analytics

Feature Selection : Missing Value Ratio

- Find out the reason for these missing values
  - Non-response
  - Error in data collection
  - Error in Reading Data

Feature Selection : Missing Value Ratio

- Find out the reason for these missing values
- Impute missing values
  - Mean
  - Median
  - Mode
  - Model

Thank
You!

Analytics Vidhya
Learn everything about analytics