Data Mining and Warehousing

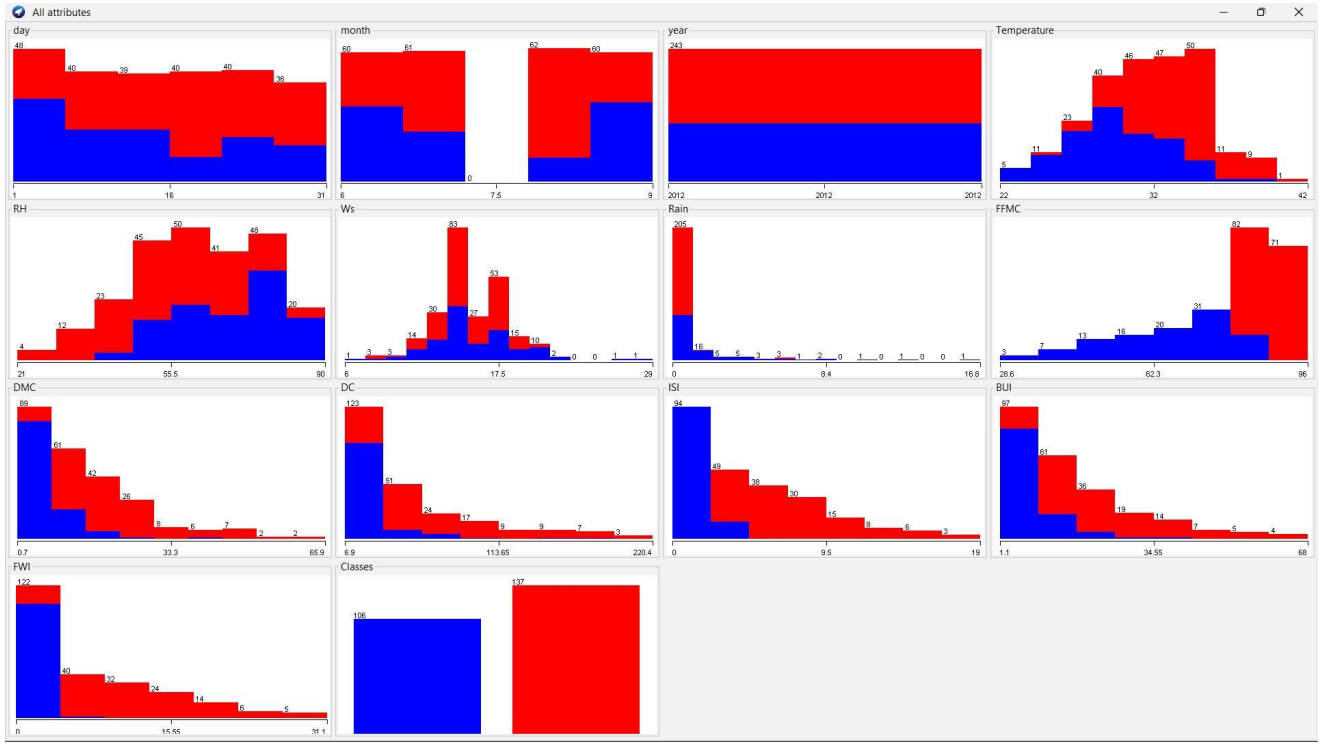
FILE USED - Algerian_forest_fires_dataset_CLEANED.arff

1. Numeric Cleaner

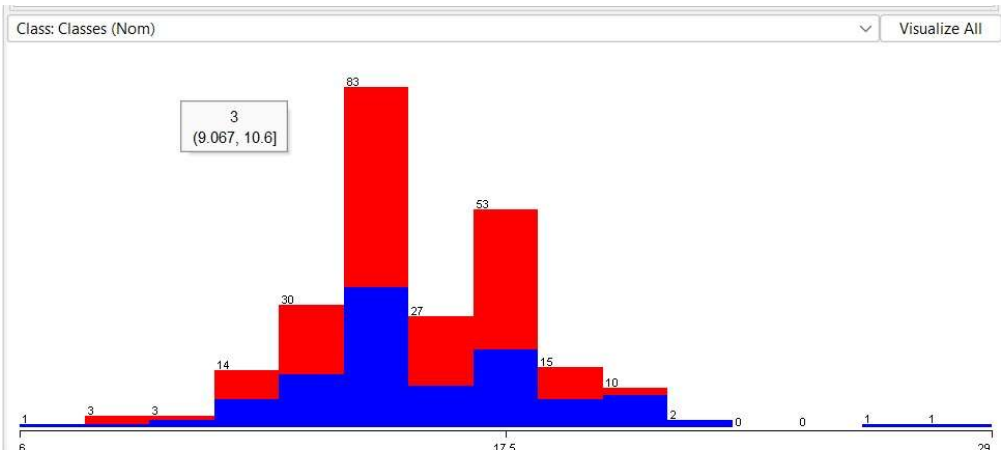
- Open Algerian_forest_fires_dataset_CLEANED.arff. Visualize all, attributes .
- Attributes months, Ws, Rain have missing values.
- Weka, choose, filter, unsupervised, attribute,
- numeric cleaner, click, attribute Indices: 2, minDefault: NaN, MinThreshold: 0.1E-7, Ok, Apply, Edit: you can see missing values.

Ans :

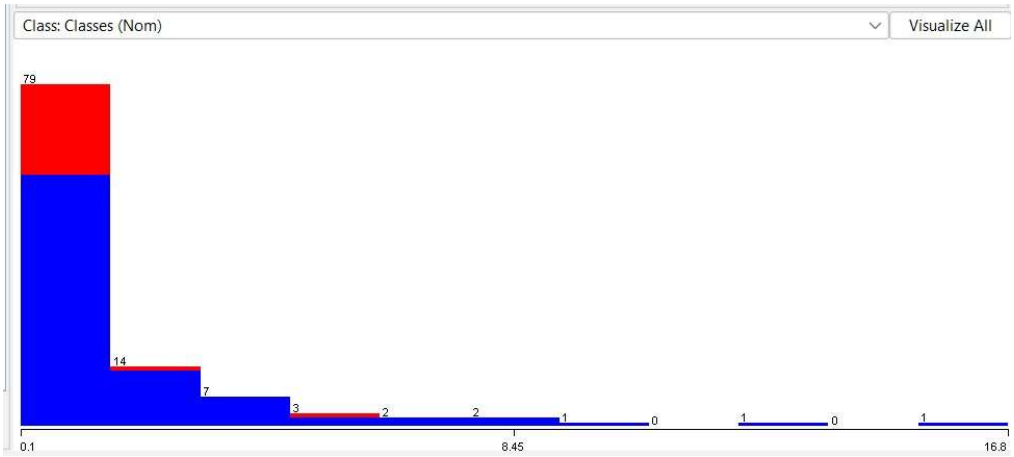
Visualization of all attributes :



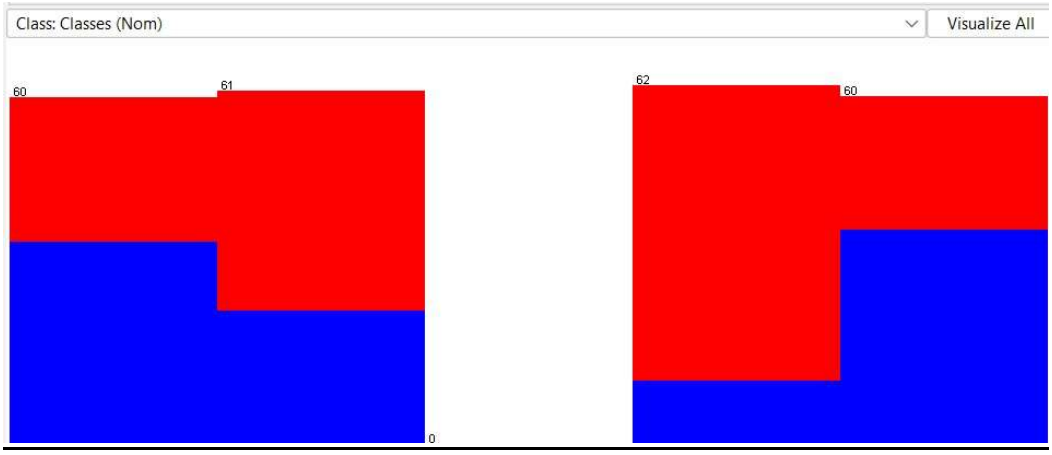
After Applying Numeric Cleaner to Ws



After Applying Numeric Cleaner to Rain



After Applying Numeric Cleaner to Month



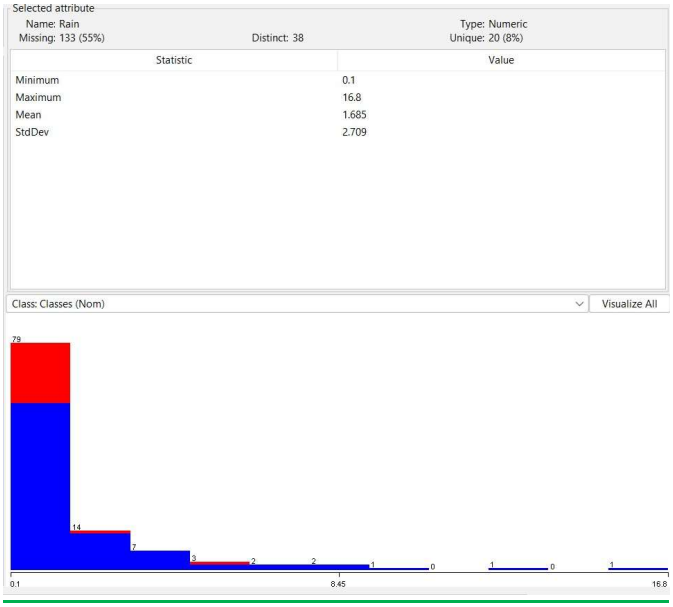
2.

Remove Missing Value

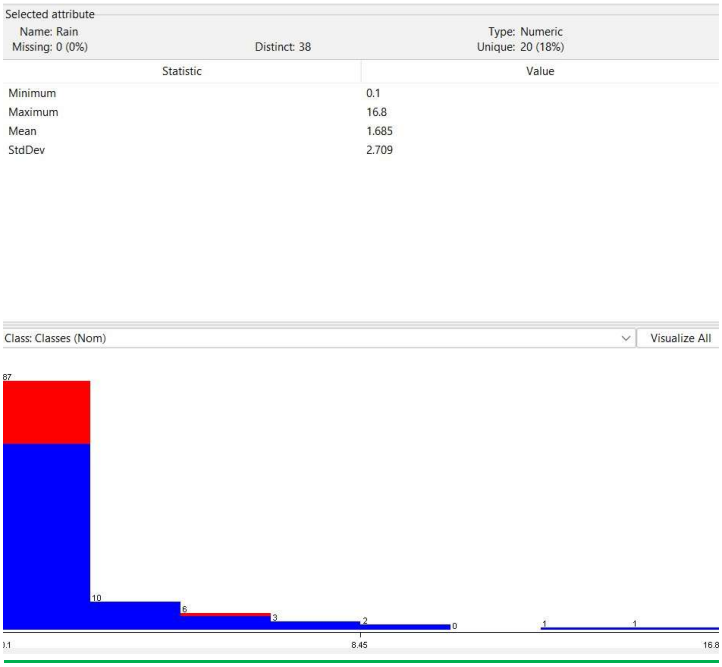
- Go to filter, go to weka, filters, unsupervised, instance, , removeWithvalues, attributeIndex: 7 matchMissingvalues: True, OK, Apply All missing value records removed.
- Undo

Ans

Before Applying RemoveWithValue filter for Rain



After Applying RemoveWithValue filter for Rain



4.

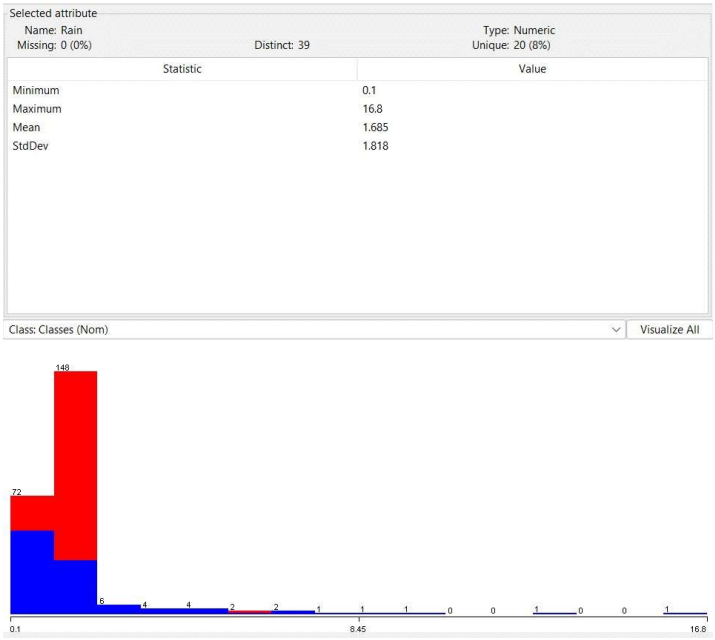
Impute Missing Value

- Weka, filter, Choose, unsupervised, attribute, replaceMissingValues, Ok, Apply.
- Ckeck mass. There are no missing values.

Ans

Before Applying ReplaceMissingValues filter for Rain

After Applying ReplaceMissingValues filter for Rain

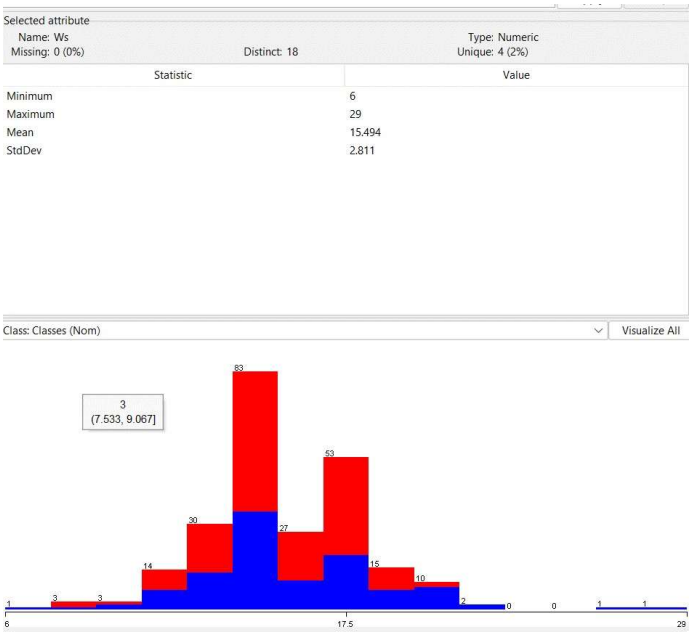


Discretize

- Select attribute age unsupervised, attribute, Discretize, select on the discretize bar, attribute indices 6 (for Ws), bins range precision (for decimal values limit) = 2,bins =3, apply, save as type csv
- Open file in excel replace values with Low speed, Medium speed and High speed, save the file as csv

6.

Ans Before Applying Discretize filter for Ws



After Applying Discretize filter for Ws

Updation in CSV file after changing Ws to nominal

A	B	C	D	E	F	G	H	I	J	K	L	M	N
day	month	year	Temperatu	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes
1	6	2012	29	57	Medium Speed	?	65.7	3.4	7.6	1.3	3.4	0.5	'not fire'
2	6	2012	29	61	Low Speed	1.3	64.4	4.1	7.6	1	3.9	0.4	'not fire'
3	6	2012	26	82	High Speed	13.1	47.1	2.5	7.1	0.3	2.7	0.1	'not fire'
4	6	2012	25	89	Low Speed	2.5	28.6	1.3	6.9	0	1.7	0	'not fire'
5	6	2012	27	77	Medium Speed	?	64.8	3	14.2	1.2	3.9	0.5	'not fire'
6	6	2012	31	67	Medium Speed	?	82.6	5.8	22.2	3.1	7	2.5	fire
7	6	2012	33	54	Low Speed	?	88.2	9.9	30.5	6.4	10.9	7.2	fire
8	6	2012	30	73	Medium Speed	?	86.6	12.1	38.3	5.6	13.5	7.1	fire
9	6	2012	25	88	Low Speed	0.2	52.9	7.9	38.8	0.4	10.5	0.3	'not fire'
10	6	2012	28	79	Low Speed	?	73.2	9.5	46.3	1.3	12.6	0.9	'not fire'
11	6	2012	31	65	Medium Speed	?	84.5	12.5	54.3	4	15.8	5.6	fire
12	6	2012	26	81	Medium Speed	?	84	12.8	61.4	4.8	17.7	7.1	fire

5. Info Gain Attribute Evaluator

- Open csv file Algerian_forest_fires_dataset_CLEANED.arff. in weka
- Select attributes from top bar
- attribute Evaluator
- InfogainAttributeEval
- Alert- yes for ranker
- Start
- Check Results
- Select attributes : (3) YEAR, remove, save

Ans Info Gain Attribute Evaluator for Forest Fire Dataset

```
Ranked attributes:
 0.9454    8 FFMC
 0.9398   11 ISI
 0.8148   13 FWI
 0.483    12 BUI
 0.4726    9 DMC
 0.4319   10 DC
 0.2174    4 Temperature
 0.1662    5 RH
 0.0817    2 month
 0.0444    1 day
 0.0215    6 Ws
 0.0157    7 Rain
 0         3 year

Selected attributes: 8,11,13,12,9,10,4,5,2,1,6,7,3 : 13
```

I removed Year attribute according to Info Gain