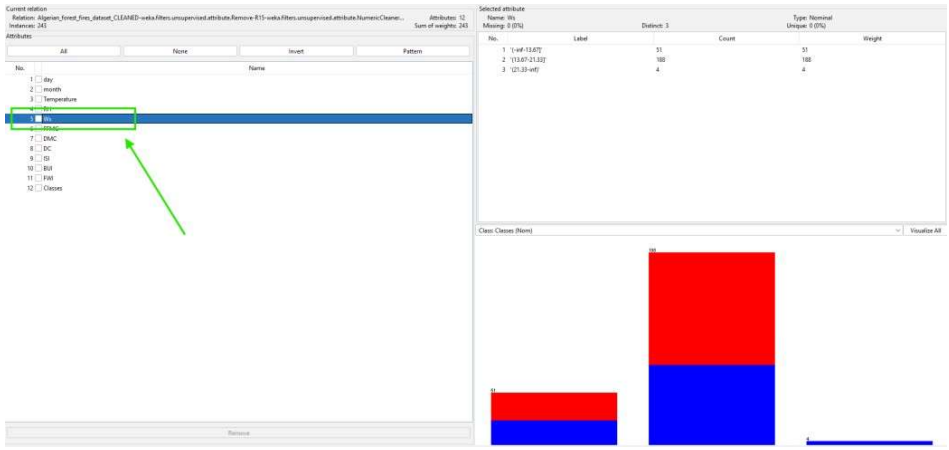
Data Mining and Warehousing

FILE USED - Algerian_forest_fires_dataset_CLEANED.arff

1. Change any attribute as class

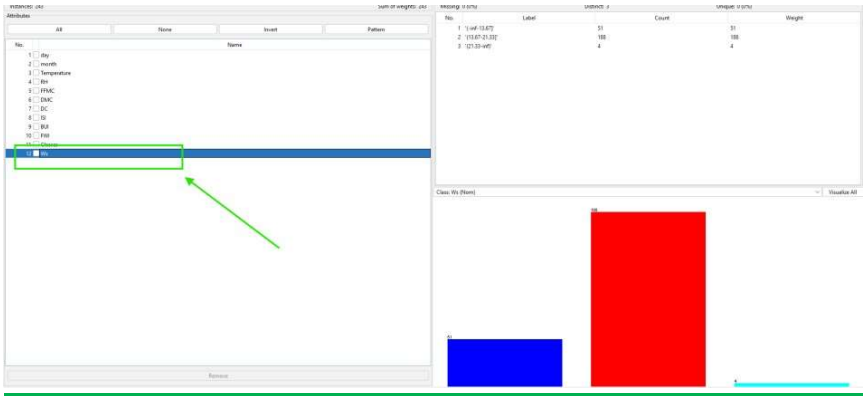
- Open dataset
- Edit
- Select Ws, set attribute as class, ok

Before setting attribute as class



After Setting attribute as class

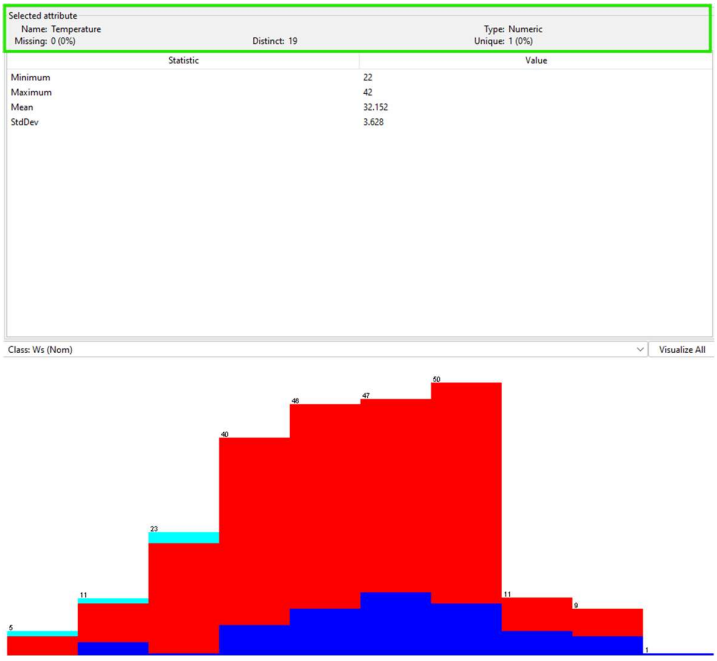
No.	1: day Numeric	2: month Numeric	3: Temperature Numeric	4: RH Numeric	5: FFMC Numeric	6: DMC Numeric	7: DC Numeric	8: ISI Numeric	9: BUI Numeric	10: FWI Numeric	11: Classes Nominal	12: **Ws** Nominal
1	2.0	6.0	29.0	61.0	64.4	4.1	7.6	1.0	3.9	0.4	not fire	'(-inf-...
2	4.0	6.0	25.0	89.0	28.6	1.3	6.9	0.0	1.7	0.0	not fire	'(-inf-...
3	7.0	6.0	33.0	54.0	88.2	9.9	30.5	6.4	10.9	7.2	fire	'(-inf-...
4	9.0	6.0	25.0	88.0	52.9	7.9	38.8	0.4	10.5	0.3	not fire	'(-inf-...
5	10.0	6.0	28.0	79.0	73.2	9.5	46.3	1.3	12.6	0.9	not fire	'(-inf-...
6	16.0	6.0	29.0	89.0	36.1	1.7	7.6	0.0	2.2	0.0	not fire	'(-inf-...
7	29.0	6.0	32.0	47.0	79.9	18.4	84.4	2.2	23.8	3.9	not fire	'(-inf-...
8	10.0	7.0	33.0	69.0	66.6	6.0	9.3	1.1	5.8	0.5	not fire	'(-inf-...
9	12.0	7.0	31.0	75.0	75.1	7.9	27.7	1.5	9.2	0.9	not fire	'(-inf-...
10	14.0	7.0	34.0	61.0	73.9	7.8	22.9	1.4	8.4	0.8	not fire	'(-inf-...
11	27.0	7.0	36.0	48.0	90.3	22.2	108.5	8.7	29.4	15.3	fire	'(-inf-...
12	2.0	8.0	35.0	55.0	78.0	5.8	10.0	1.7	5.5	0.8	not fire	'(-inf-...
13	4.0	8.0	34.0	69.0	85.0	8.2	19.8	4.0	8.2	3.9	fire	'(-inf-...
14	5.0	8.0	34.0	65.0	86.8	11.1	29.7	5.2	11.5	6.1	fire	'(-inf-...
15	11.0	8.0	35.0	63.0	88.9	21.7	77.0	7.1	25.5	12.1	fire	'(-inf-...
16	12.0	8.0	35.0	51.0	81.3	15.6	75.1	2.5	20.7	4.2	not fire	'(-inf-...
17	15.0	8.0	36.0	55.0	82.4	15.6	92.5	3.7	22.0	6.3	fire	'(-inf-...
18	10.0	9.0	33.0	73.0	59.9	2.2	8.9	0.7	2.7	0.3	not fire	'(-inf-...
19	12.0	9.0	29.0	88.0	71.0	2.6	16.6	1.2	3.7	0.5	not fire	'(-inf-...
20	18.0	9.0	32.0	49.0	89.4	9.8	33.1	6.8	11.3	7.7	fire	'(-inf-...
21	21.0	9.0	31.0	55.0	87.8	16.5	57.9	5.4	19.2	8.3	fire	'(-inf-...
22	23.0	9.0	32.0	54.0	73.7	7.9	30.4	1.2	9.6	0.7	not fire	'(-inf-...
23	26.0	9.0	31.0	54.0	82.0	6.0	16.3	2.5	6.2	1.7	not fire	'(-inf-...
24	27.0	9.0	31.0	66.0	85.7	8.3	24.9	4.0	9.0	4.1	fire	'(-inf-...



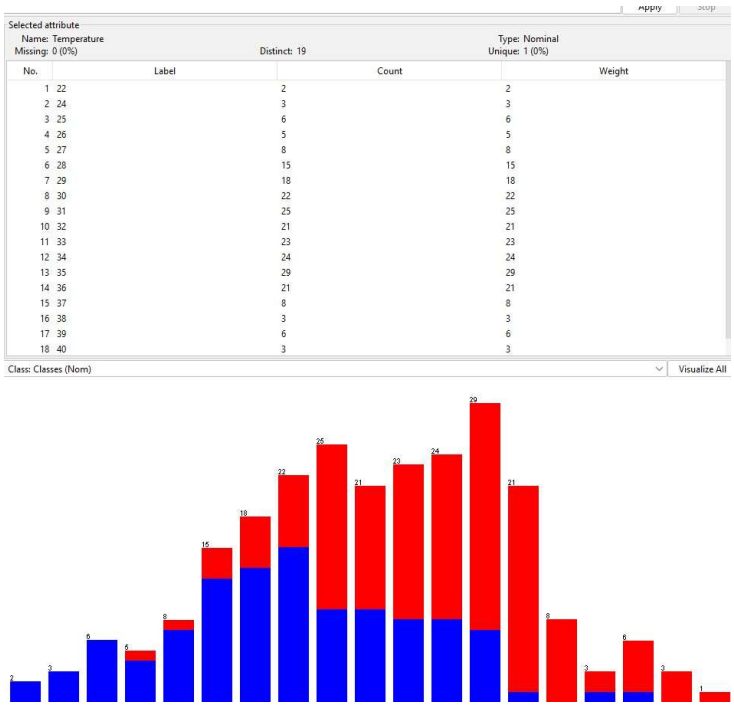
2. Change Numeric to Nominal

- Open dataset
- Select attribute Temperature - numeric
- Weka, filters, unsupervised, attribute,
- NumericToNominal, Click on bar, attribute indices 3,
- Apply

Before applying Numeric to Nominal



After applying Numeric to Nominal



Selected attribute

Name: Temperature | Type: Nominal

Missing: 0 (0%) | Distinct: 19 | Unique: 1 (0%)

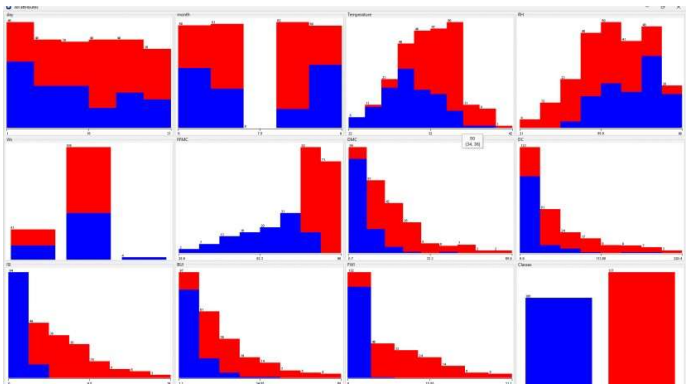
No.	Label	Count	Weight
1	22	2	2
2	24	3	3
3	25	6	6
4	26	5	5
5	27	8	8
6	28	15	15
7	29	18	18
8	30	22	22
9	31	25	25
10	32	21	21
11	33	23	23
12	34	24	24
13	35	29	29
14	36	21	21
15	37	8	8
16	38	3	3
17	39	6	6
18	40	3	3

Class: Classes (Nom) | Visualize All

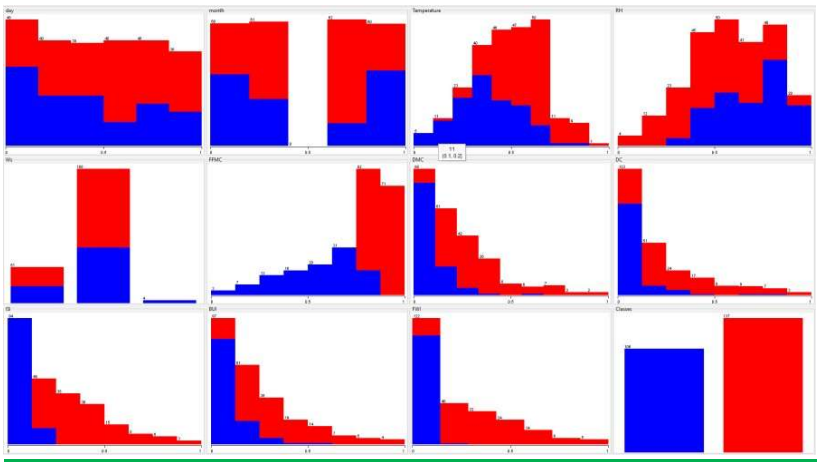
3. Normalize/ Standardize

- Open dataset, check values of all attributes. Each has a different range.
- Weka, filters, unsupervised, attribute, normalize, apply
- (all values between 0 and 1)
- Undo, standardize, apply (mean 0, std dev=1)

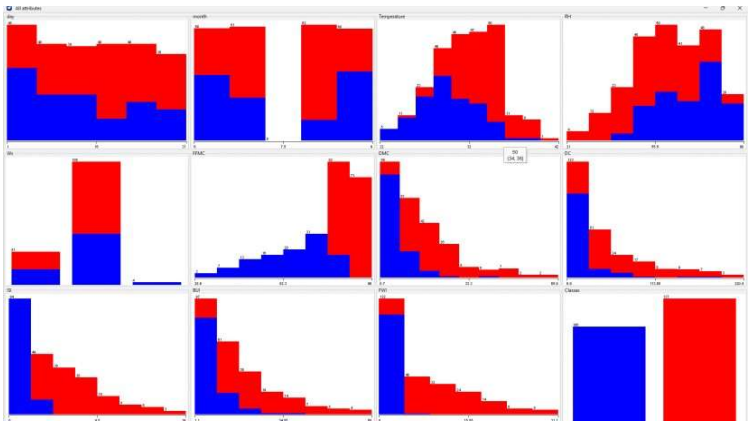
Before applying Normalization



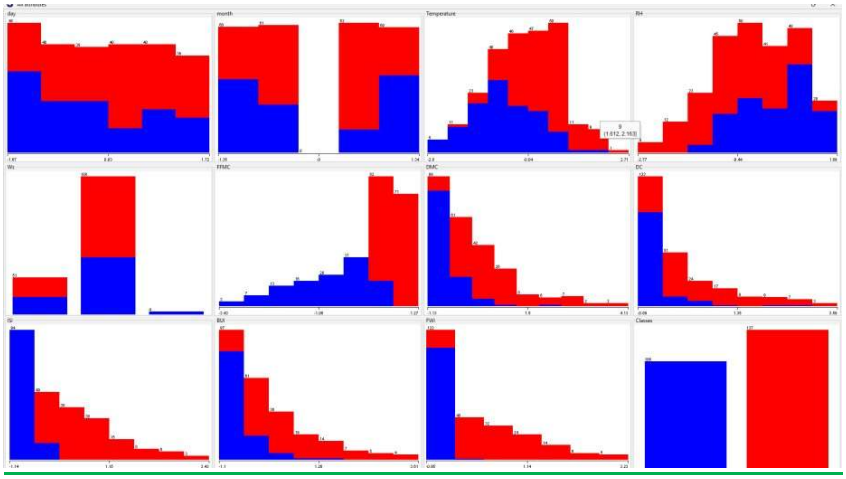
After applying Normalization



Before applying Standardization



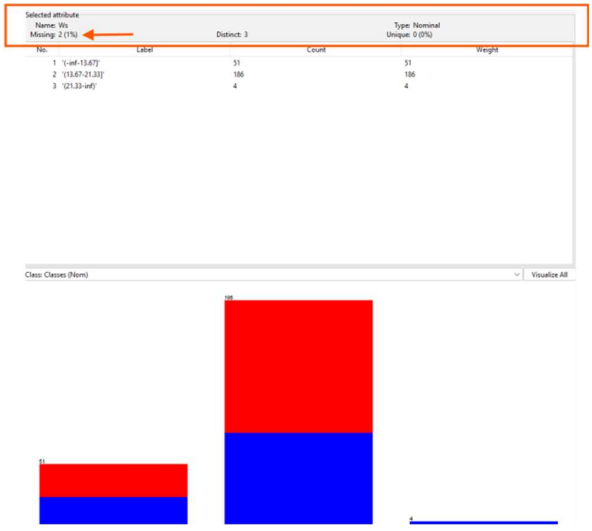
After applying Standardization



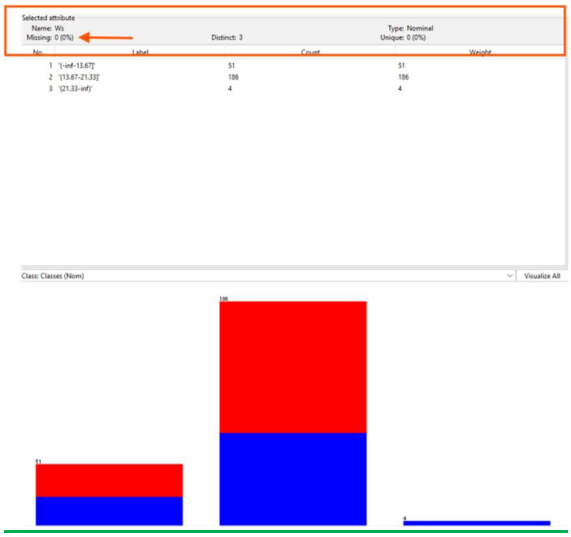
4. Remove Nominal Attribute Missing values

- Open dataset
- Select attribute Ws. It has missing values
- Weka, filters, unsupervised,instance,
- RemoveWithValues, click bar, attribute indices : 5, invert
- Selection: true, matchMissingValues: True, OK, Apply **Before**

Removing missing values



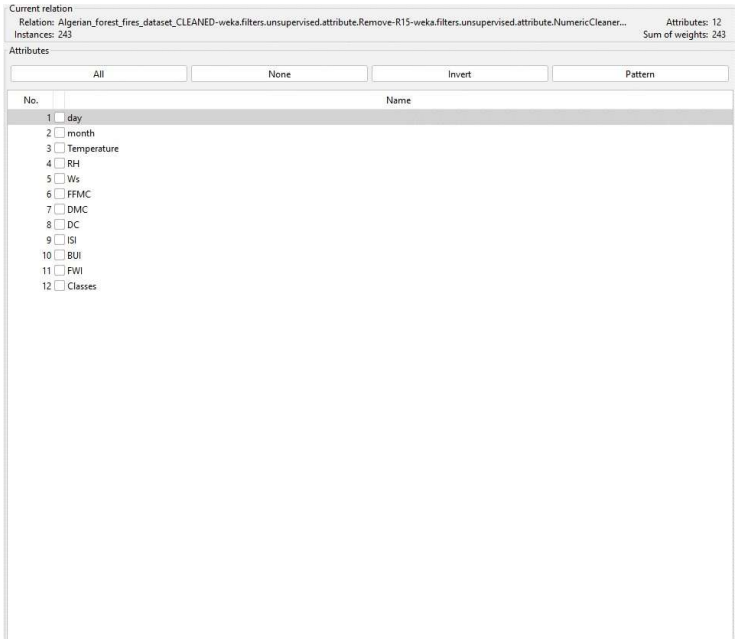
After Removing missing values



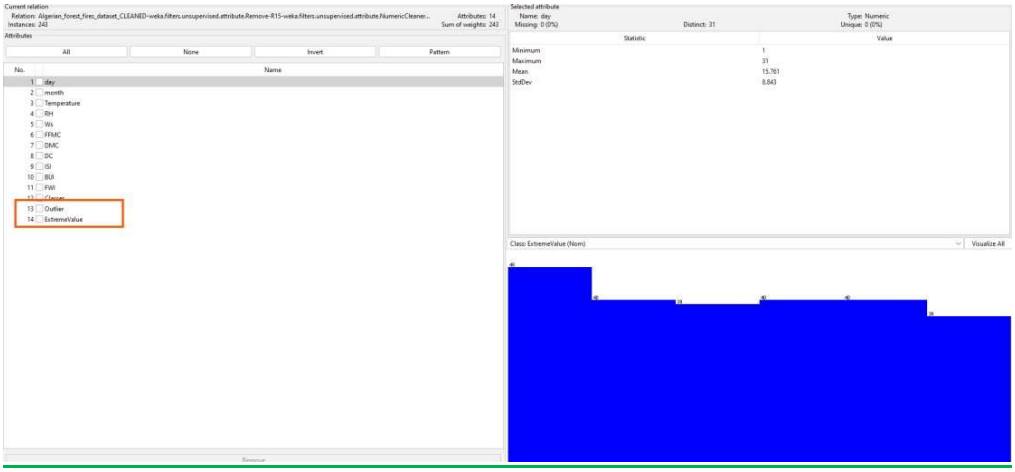
5. Finding and removing Outliers/ Extreme

- Values (Applicable for file having no missing values only and only numeric attributes)
- Open file forest_fire.arrf
- Weka, filters, unsupervised, attribute, InterQuartilerange, Apply
- Two extra columns added. Edit, Select column outlier,
- set class as outlier, OK. visualize
- Weka,Filters , unsupervised, instance,
- removeWithvalues, click on bar
- Attributeindex: 14
- Attribute outlier has two values no(1) and yes(2). We want to remove outliers, so nominal indices=2 or last..
- ok, Apply. save as a new file

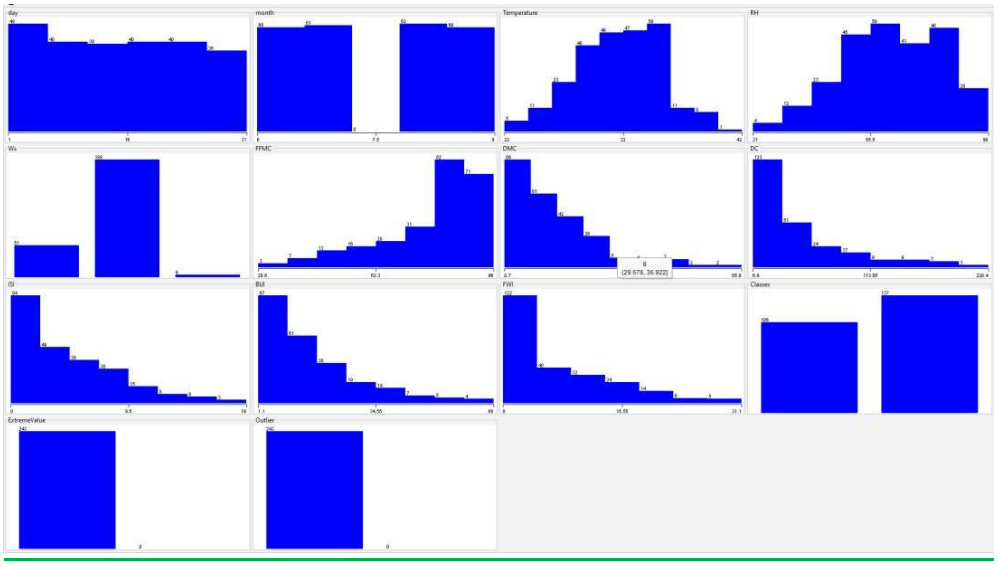
Before Applying InterQuartileRange



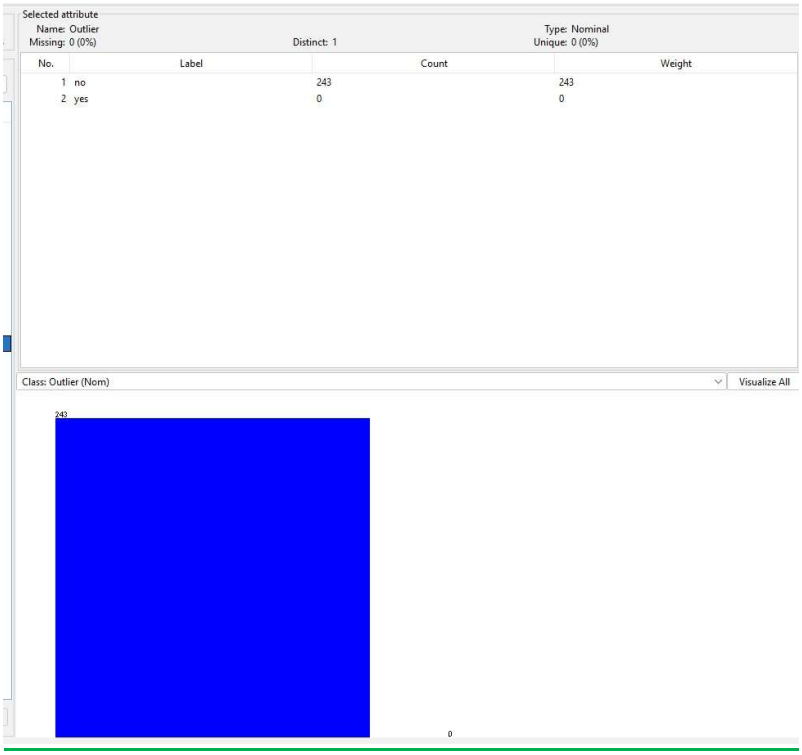
After Applying InterQuartileRange



After setting Outlier as class



After Applying RemoveWithValue in outlier class

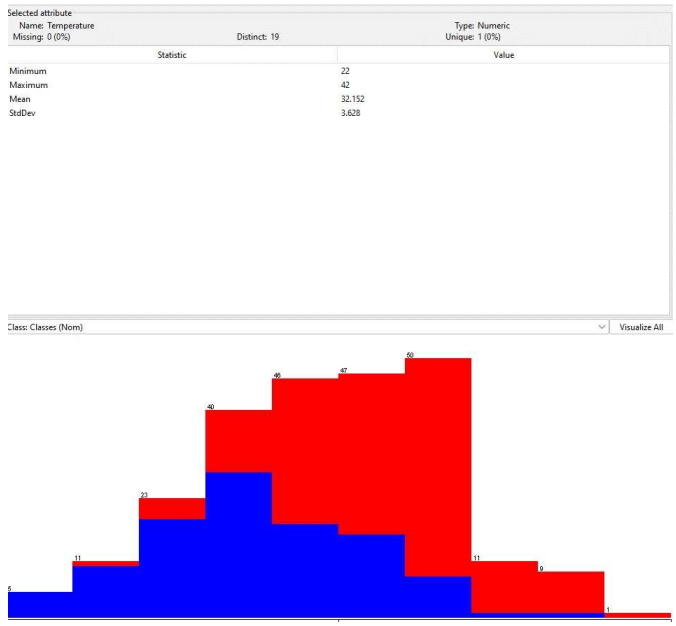


Since my outlier's yes instance was already 0 that's why before and after is same.

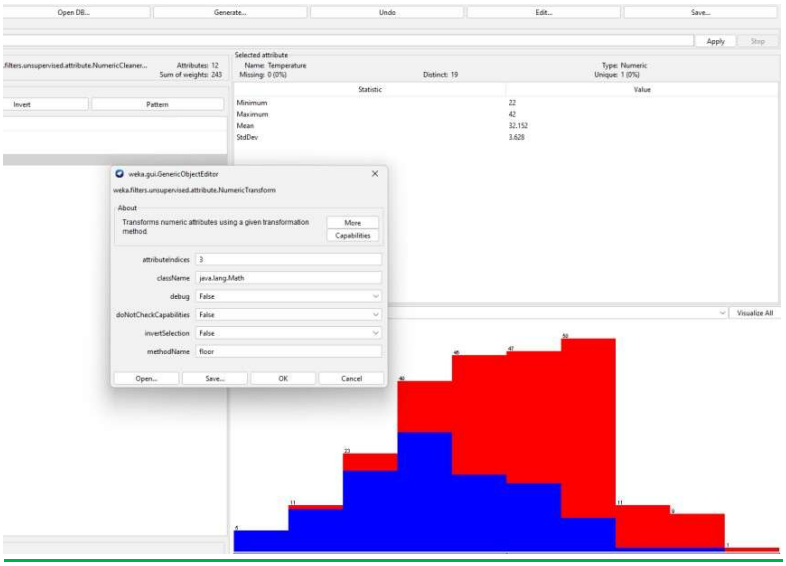
6. Numeric transform

- Forestfire.arff weka filter unsupervised attribute NumericTransform, attributIndices: 3, metod name :
- Floor

Before Applying Numeric Transform on TEMPERATURE Attribute



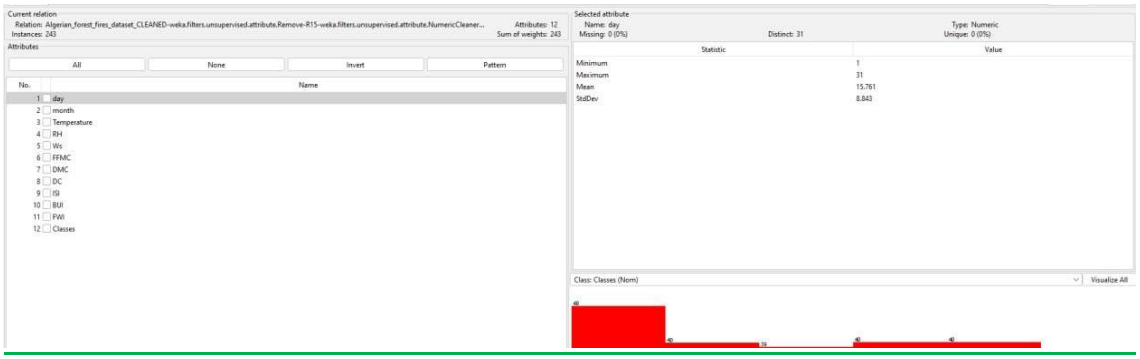
After Applying Numeric Transform on TEMPERATURE Attribute



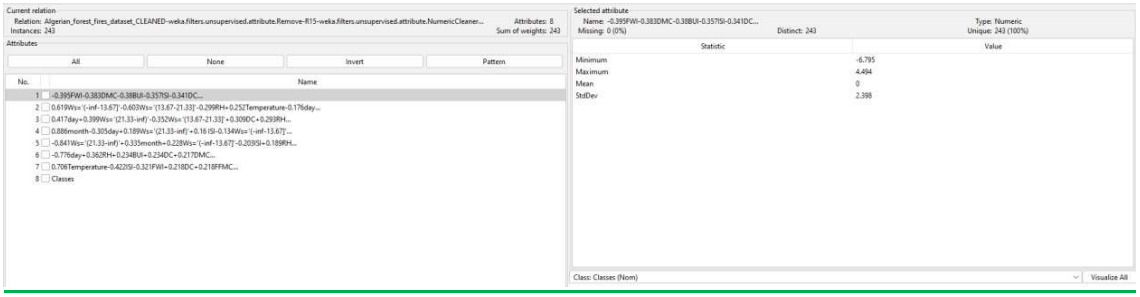
7. PCA

- Open file ForestFire.arff, filter, unsupervised, attribute, PrincipalComponents, click, variance covered:0.95, ok, apply.
- Check for variance/Std deviation on the right. It is the
- maximum variance, Set threshold=50% of the maximum.
- Select them (4,5,6,7) and click remove

Before Applying PCA



After Applying PCA



So as we can see

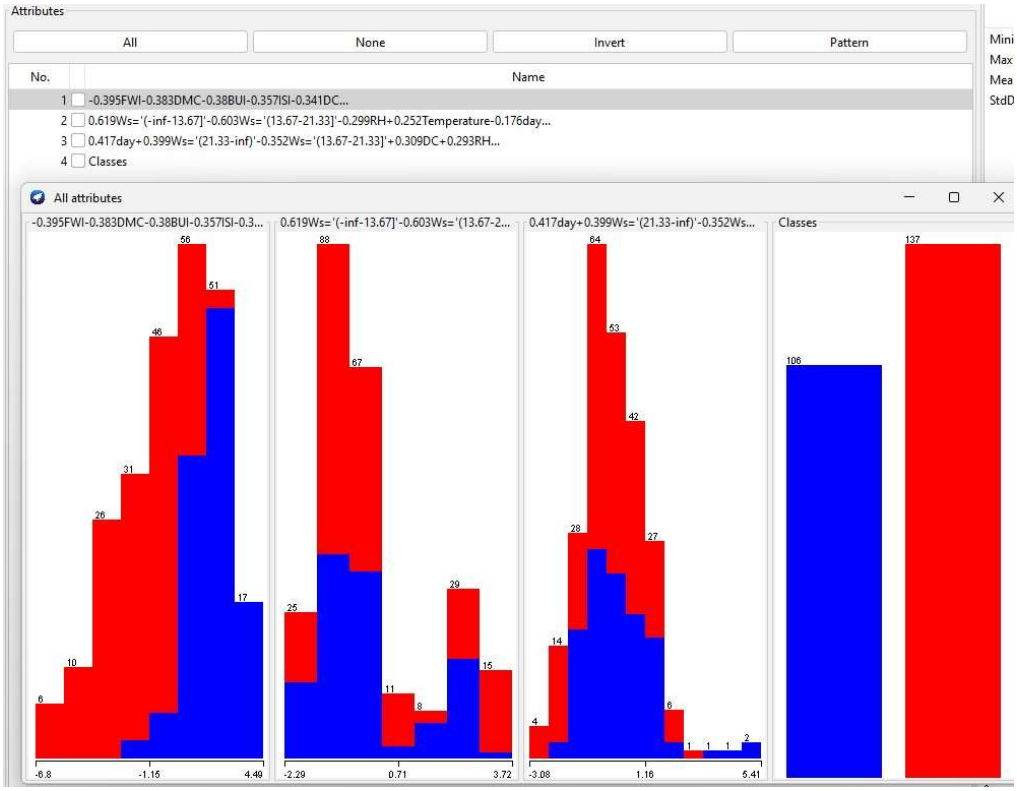
Now the attributes have been placed in decreasing order of standard dev

My max standard dev is 2.398

Setting ThreshHold = 1.199 (50% of max)

Thus attribute 4,5,6,7 should be removed

Final dataset after removing attributes



Ranker also proves this :

```
Ranked attributes:
 0.7813  1 0.707Classes=fire-0.686-0.395FWI-0.383DMC-0.38BUI-0.357ISI-0.341DC...-0.1430.417day+0.399Ws='(21.33-inf)'-0.352Ws='(13.67-21.33]'+0.309DC+0.293RH...-0.063-0.841Ws='(21.33-inf)'+0.335month+0.228Ws='(-inf-13.67]'-0.203I
 0.6563  2 -0.6890.886month-0.305day+0.189Ws='(21.33-inf)'+0.16 ISI-0.134Ws='(-inf-13.67]'...-0.685-0.841Ws='(21.33-inf)'+0.335month+0.228Ws='(-inf-13.67]'-0.203ISI+0.189RH...-0.2020.706Temperature-0.422ISI-0.321FWI+0.218DC+0.21
 0.5313  3 0.6  0.417day+0.399Ws='(21.33-inf)'-0.352Ws='(13.67-21.33]'+0.309DC+0.293RH...-0.598-0.776day+0.362RH+0.234BUI+0.234DC+0.217DMC...-0.4670.706Temperature-0.422ISI-0.321FWI+0.218DC+0.218FFMC...+0.1480.619Ws='(-inf-13.67
 0.4063  4 -0.7390.619Ws='(-inf-13.67]'-0.603Ws='(13.67-21.33]'-0.299RH+0.252Temperature-0.176day...-0.556-0.776day+0.362RH+0.234BUI+0.234DC+0.217DMC...-0.3570.417day+0.399Ws='(21.33-inf)'-0.352Ws='(13.67-21.33]'+0.309DC+0.293RH
 0.2813  5 0.85 0.706Temperature-0.422ISI-0.321FWI+0.218DC+0.218FFMC...-0.39-0.776day+0.362RH+0.234BUI+0.234DC+0.217DMC...+0.2470.417day+0.399Ws='(21.33-inf)'-0.352Ws='(13.67-21.33]'+0.309DC+0.293RH...+0.1810.619Ws='(-inf-13.67]
 0.1563  6 0.6270.619Ws='(-inf-13.67]'-0.603Ws='(13.67-21.33]'-0.299RH+0.252Temperature-0.176day...-0.6130.417day+0.399Ws='(21.33-inf)'-0.352Ws='(13.67-21.33]'+0.309DC+0.293RH...-0.416-0.776day+0.362RH+0.234BUI+0.234DC+0.217DMC..
 0.0313  7 0.6990.886month-0.305day+0.189Ws='(21.33-inf)'+0.16 ISI-0.134Ws='(-inf-13.67]'...-0.688-0.841Ws='(21.33-inf)'+0.335month+0.228Ws='(-inf-13.67]'-0.203ISI+0.189RH...+0.1540.417day+0.399Ws='(21.33-inf)'-0.352Ws='(13.67-2

Selected attributes: 1,2,3,4,5,6,7 : 7
```