**Data Mining Lab**

**Assignment 1**

**Name** – Neelangshu Roy

**Reg No** - 20214060

**Group – C**

Q1. Create a weather table with training data set including attributes :

outlook {sunny,rainy,overcast},

temperature numeric, humidity numeric,

windy {TRUE,FALSE}, play {yes,no}

neeluweather.arff -

@relation NeeluWeather

@attribute outlook {sunny, rainy, overcast}

@attribute temperature numeric

@attribute humidity numeric

@attribute windy {TRUE, FALSE}

@attribute play {yes, no}

% numeric means enter integers as values

% values in square brackets means you can take values only from what is given in set

@data

sunny,85,85,FALSE,no

sunny,80,90,TRUE,no

overcast,83,86,FALSE,yes

rainy,70,96,FALSE,yes

rainy,68,80,FALSE,yes

rainy,65,70,TRUE,no

overcast,64,65,TRUE,yes

sunny,72,95,FALSE,no

sunny,69,70,FALSE,yes

rainy,75,80,FALSE,yes

sunny,75,70,TRUE,yes

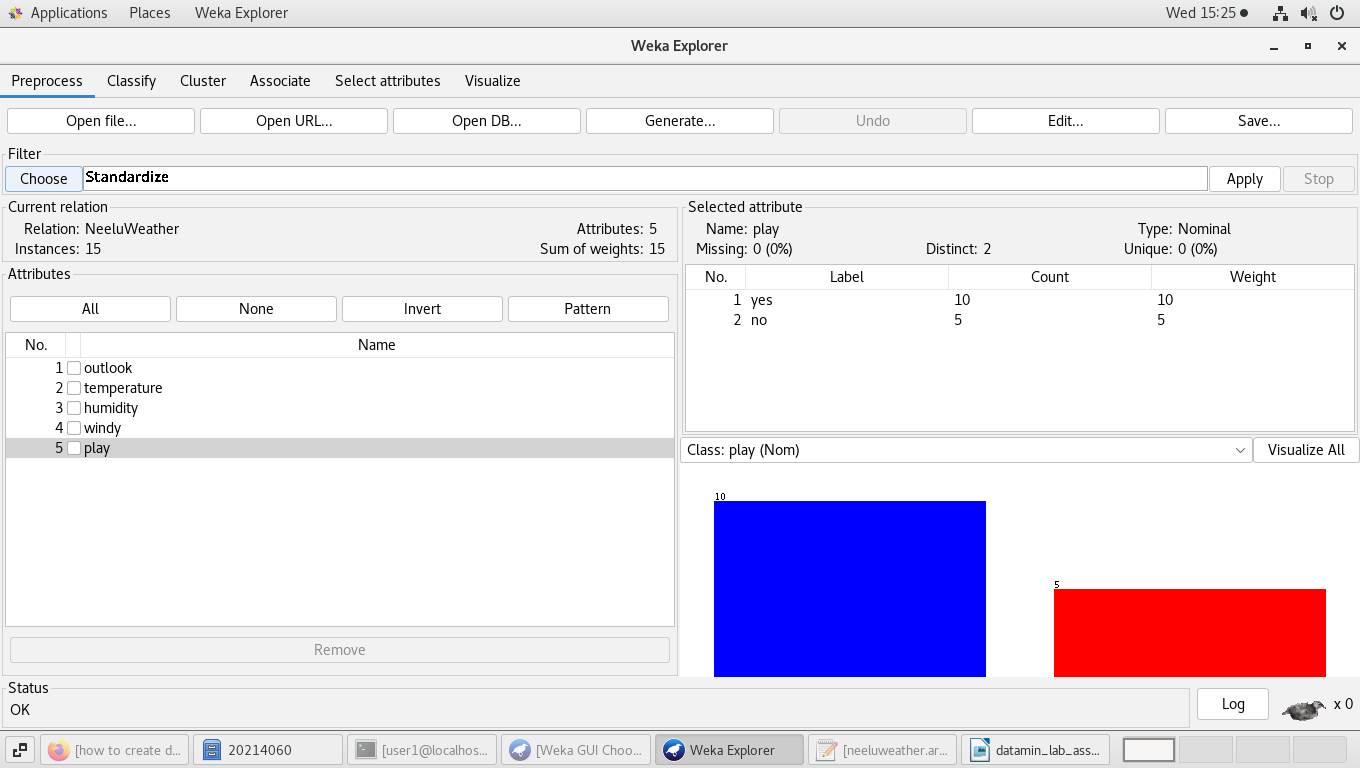
overcast,72,90,TRUE,yes

overcast,81,75,FALSE,yes

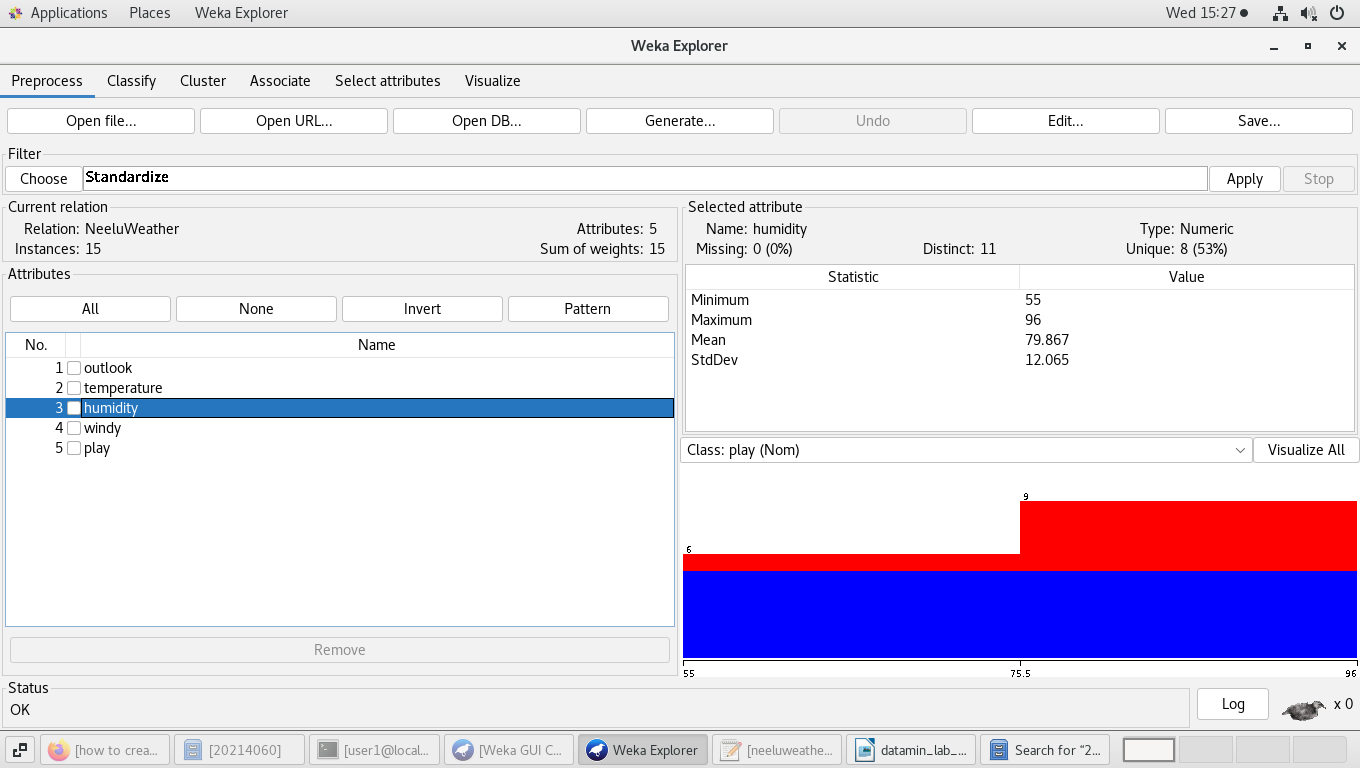
rainy,71,91,TRUE,no

sunny,80,55,TRUE,yes

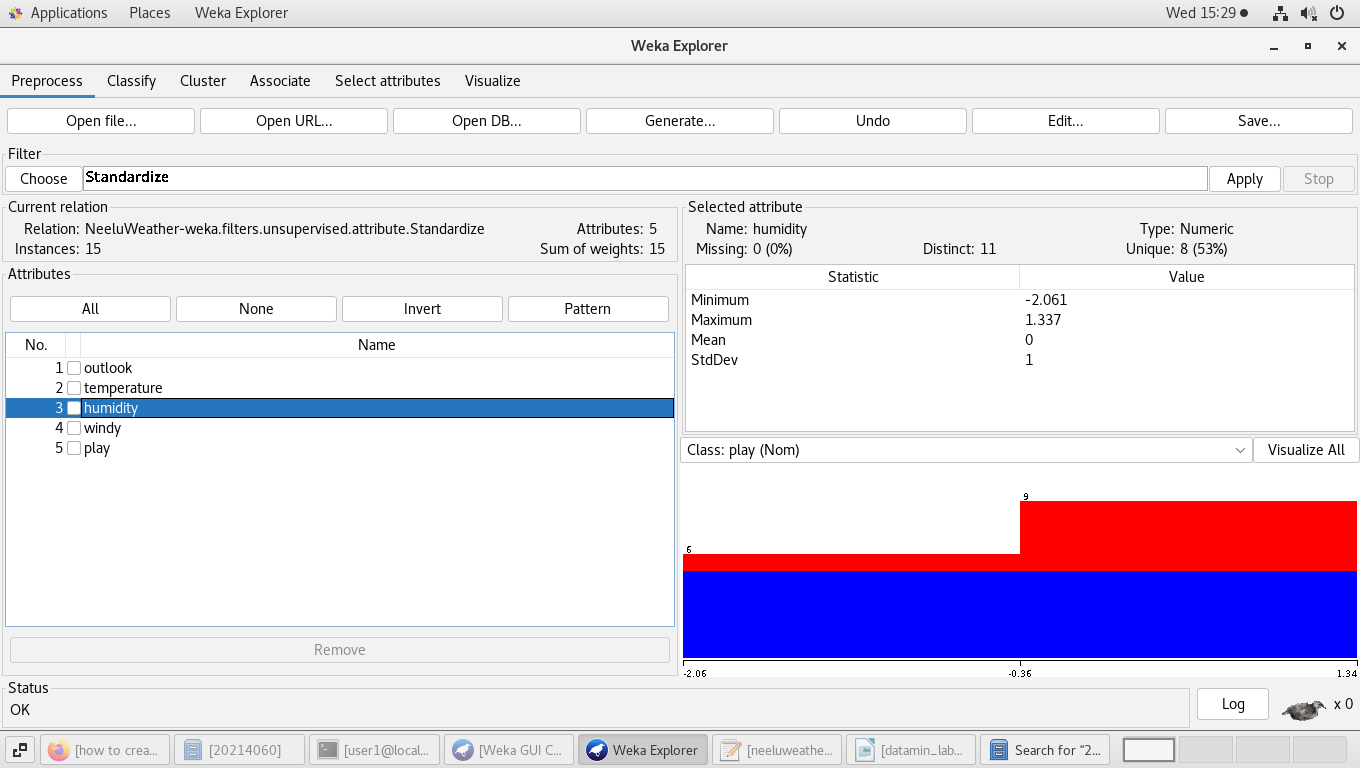
Pre-processing techniques on this self-made dataset:



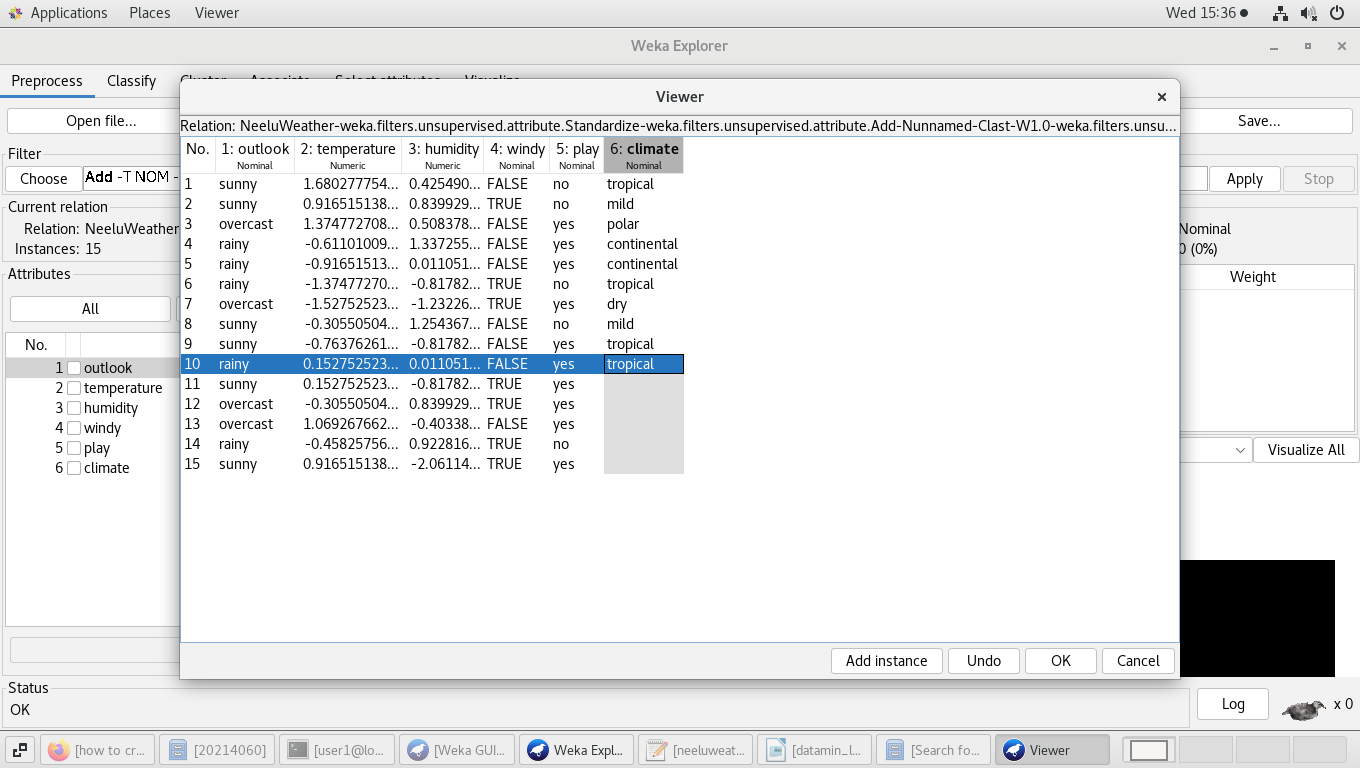
Humidity -



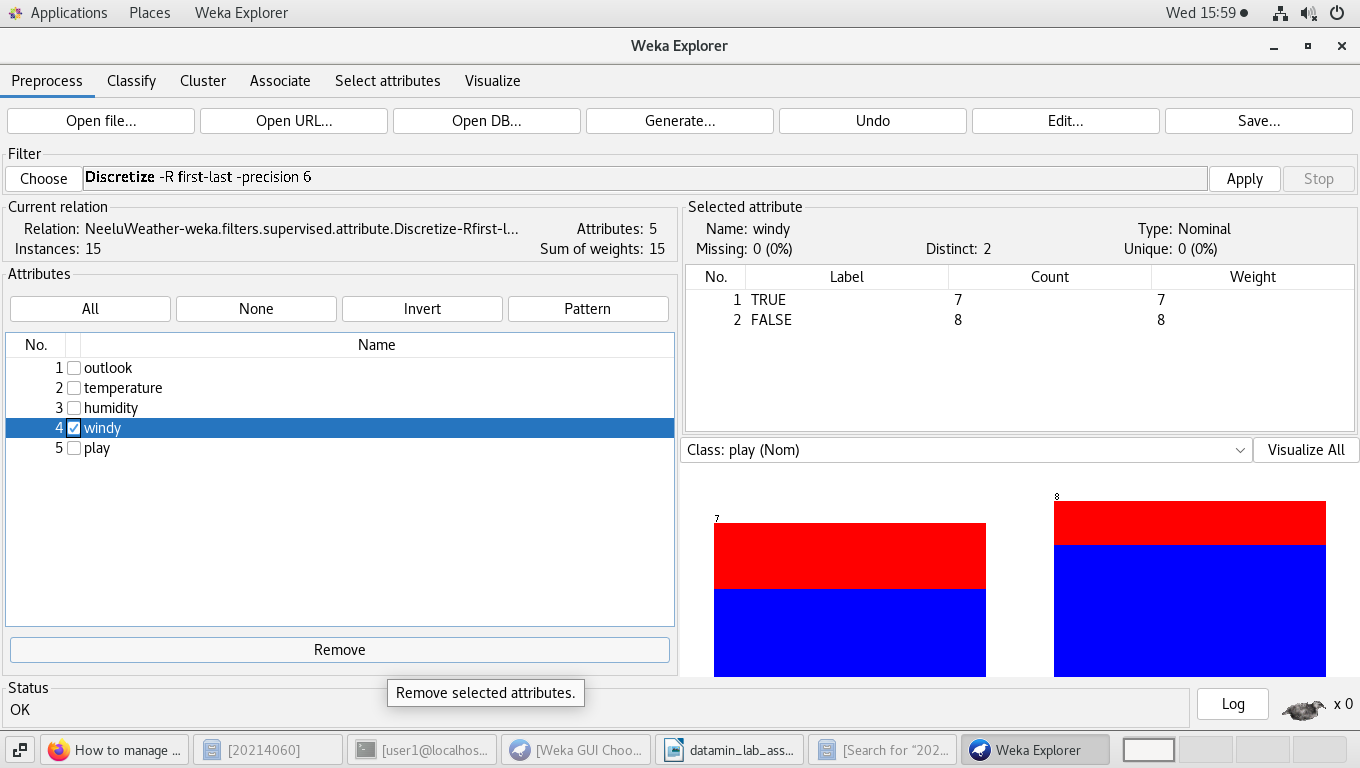
Humidity after standardization i.e. mean is set to zero, so observations are scaled accordingly -



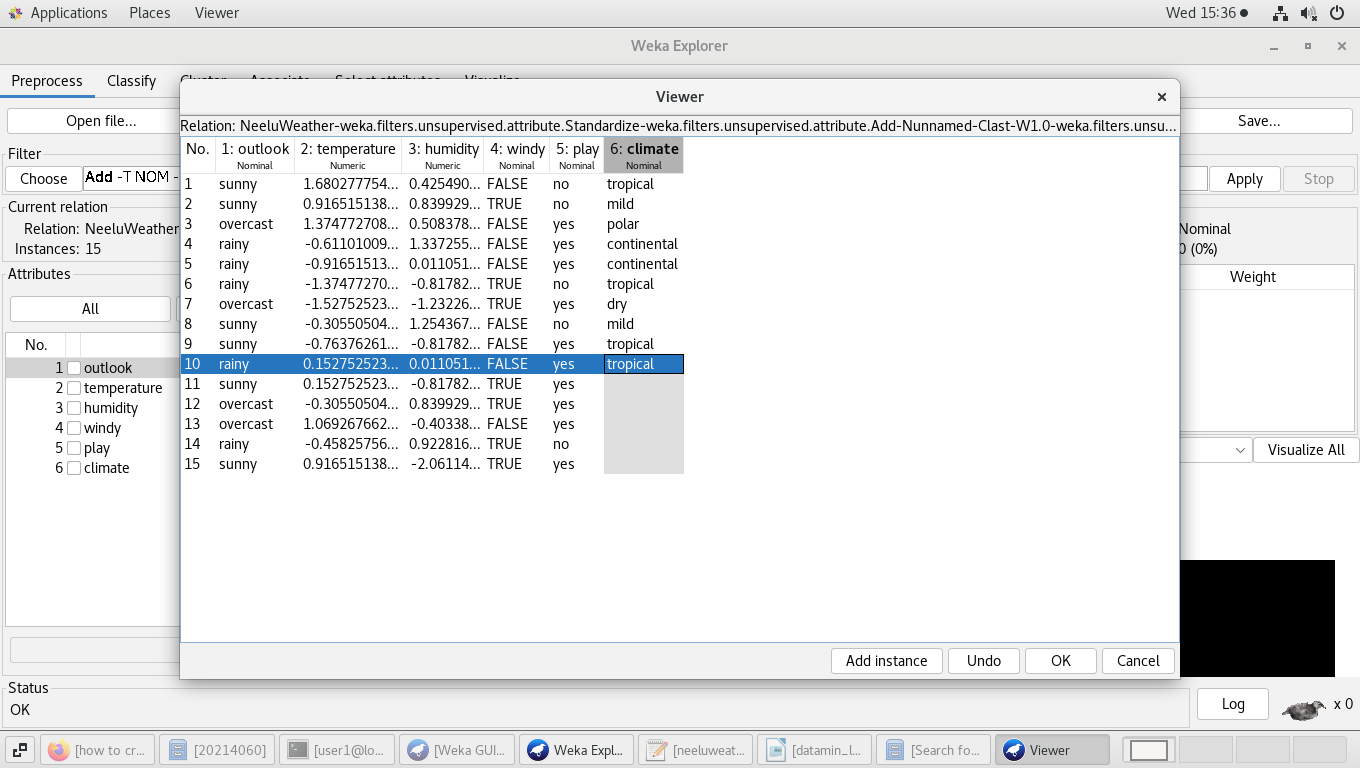
1) Adding attribute climate through the add attribute viewer GUI in the WEKA Explorer :



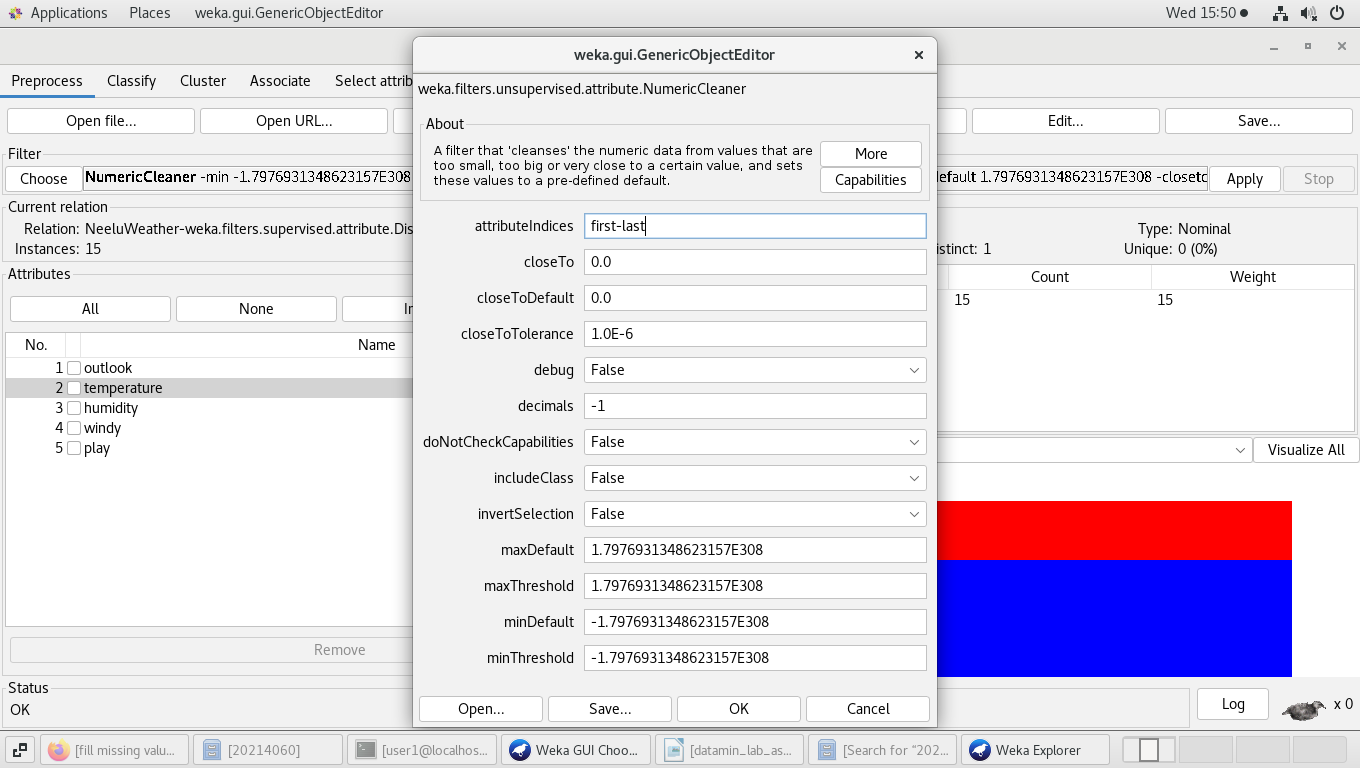
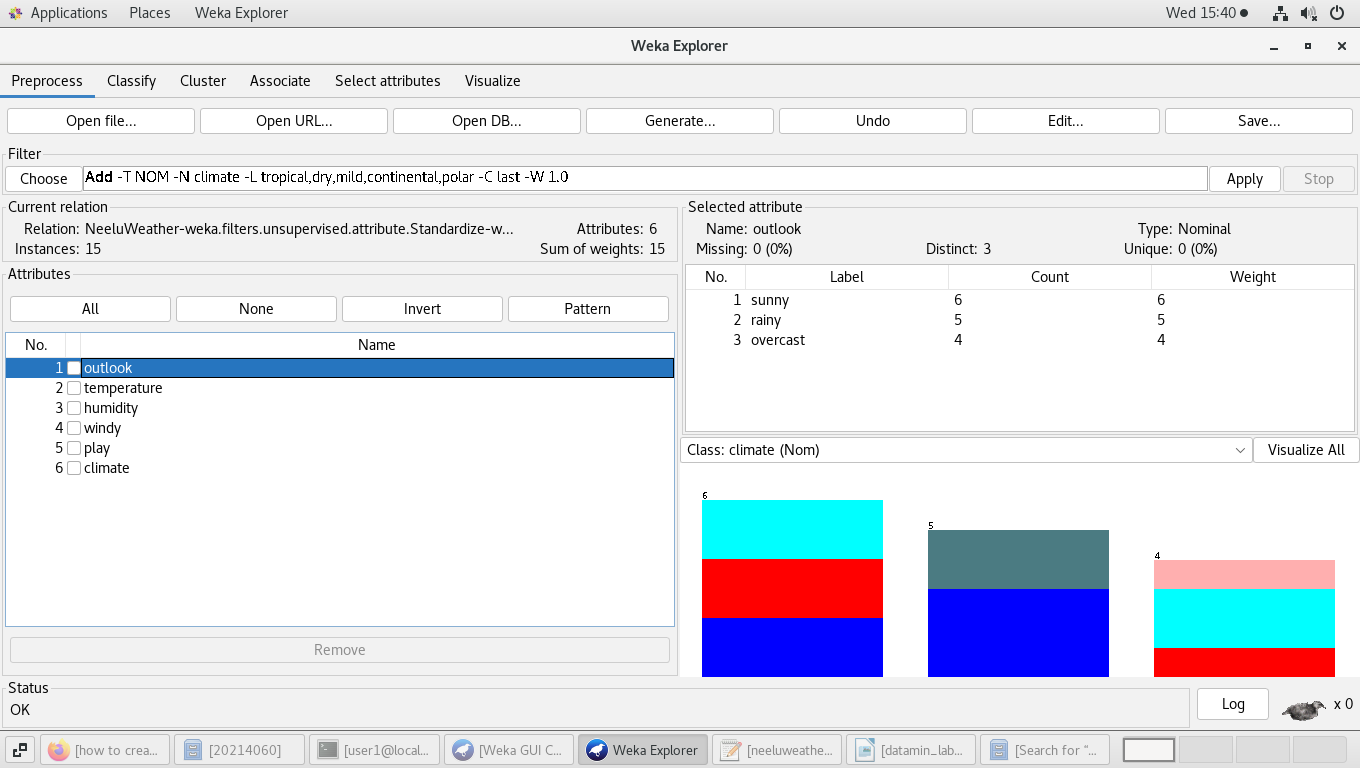
2) Removing attribute (here, wind) as shown below -



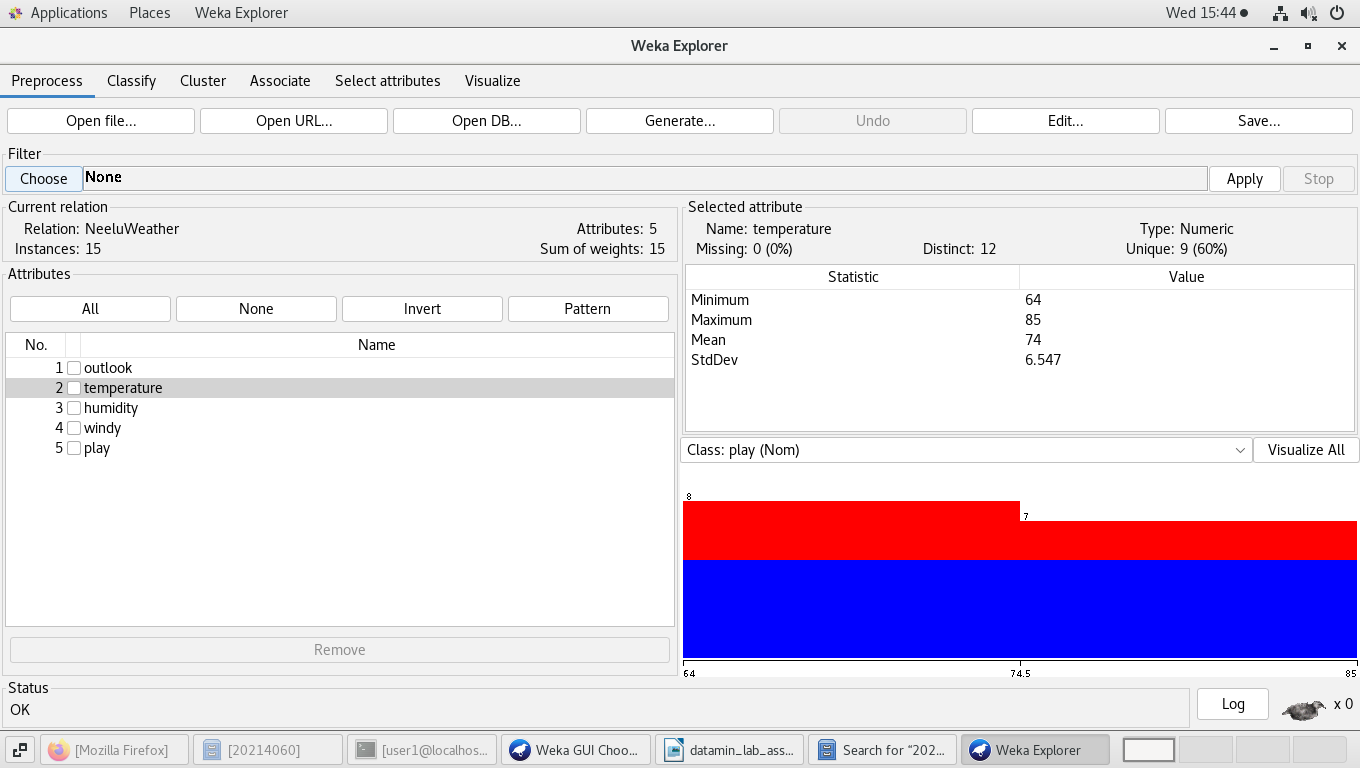
3) Selected attribute – climate



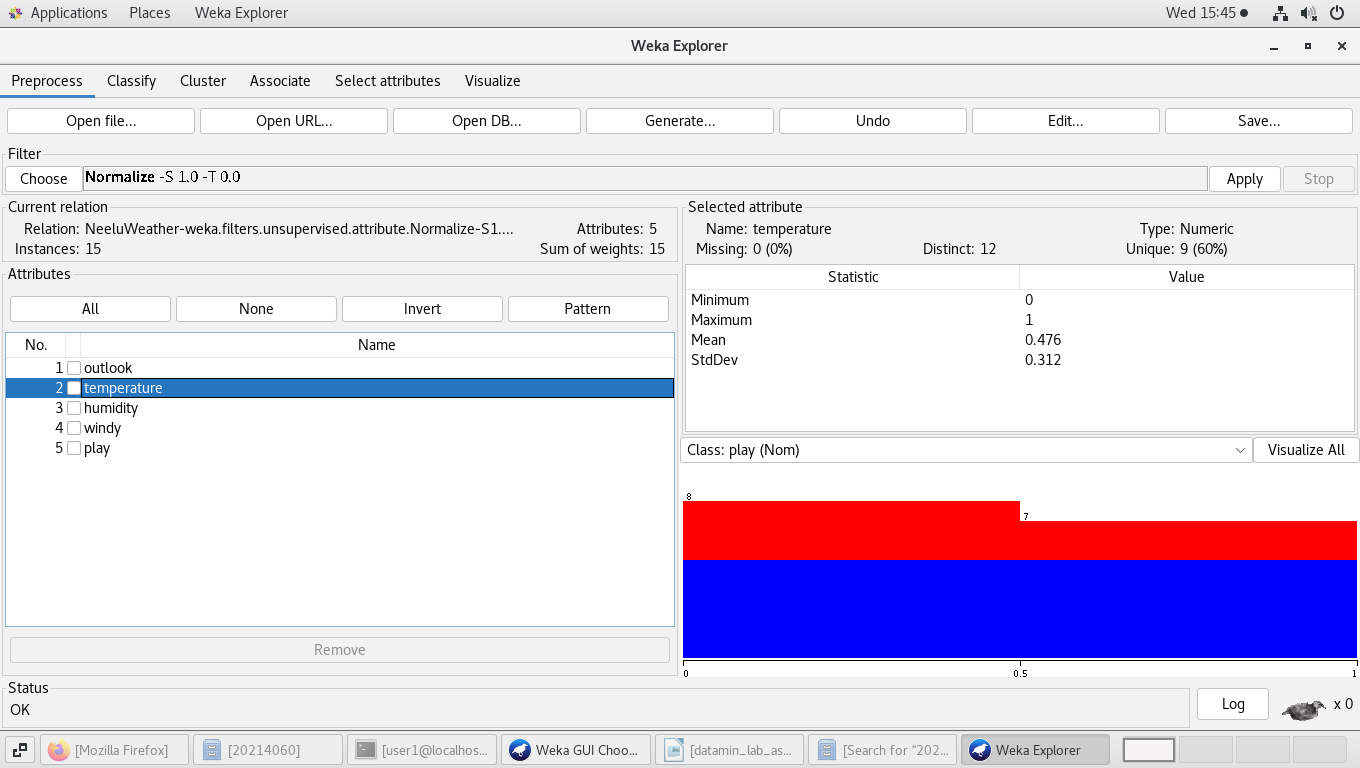
4) Filling missing values is basically using NumericCleaner which basically sets minimum threshold and minimum default values for the missing data items.



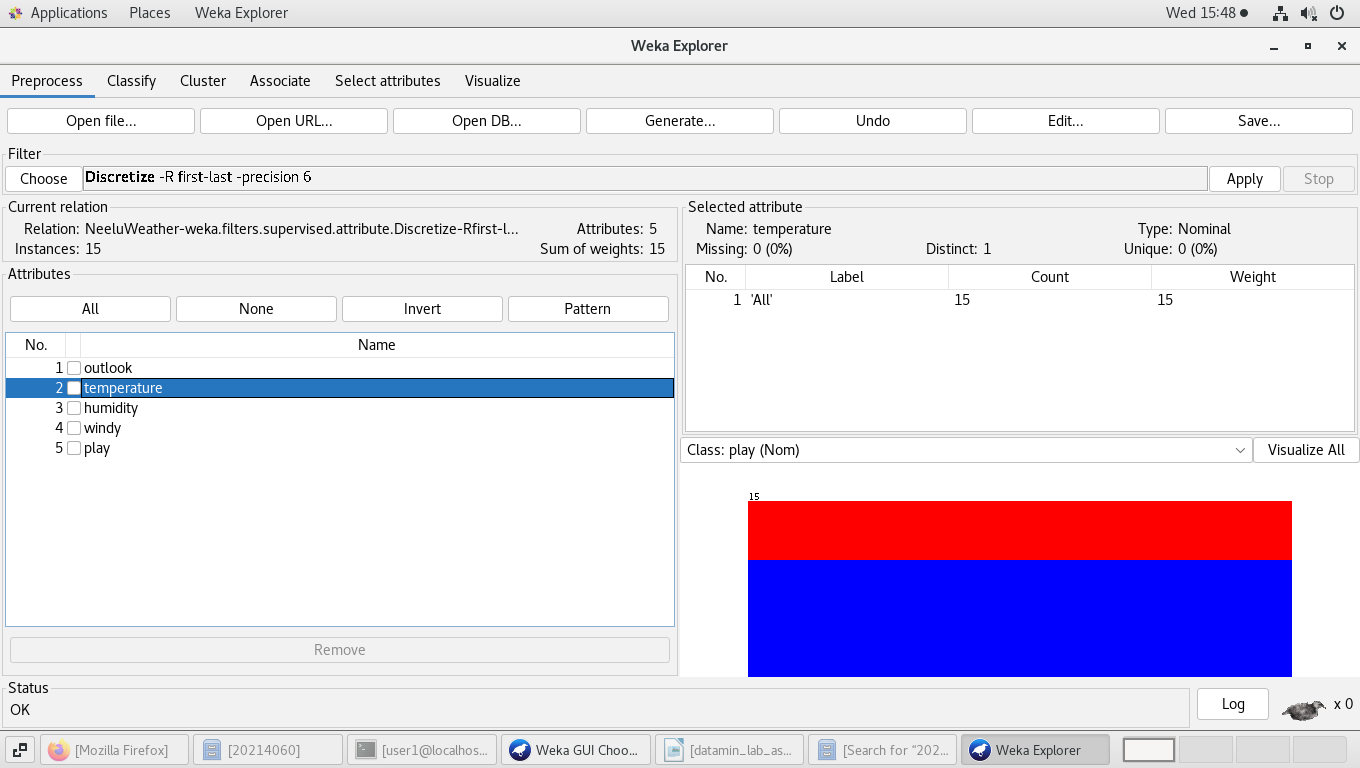
5) Before normalization -



After normalization i.e. values are scaled between 0 and 1 (max = 1, min = 0) -



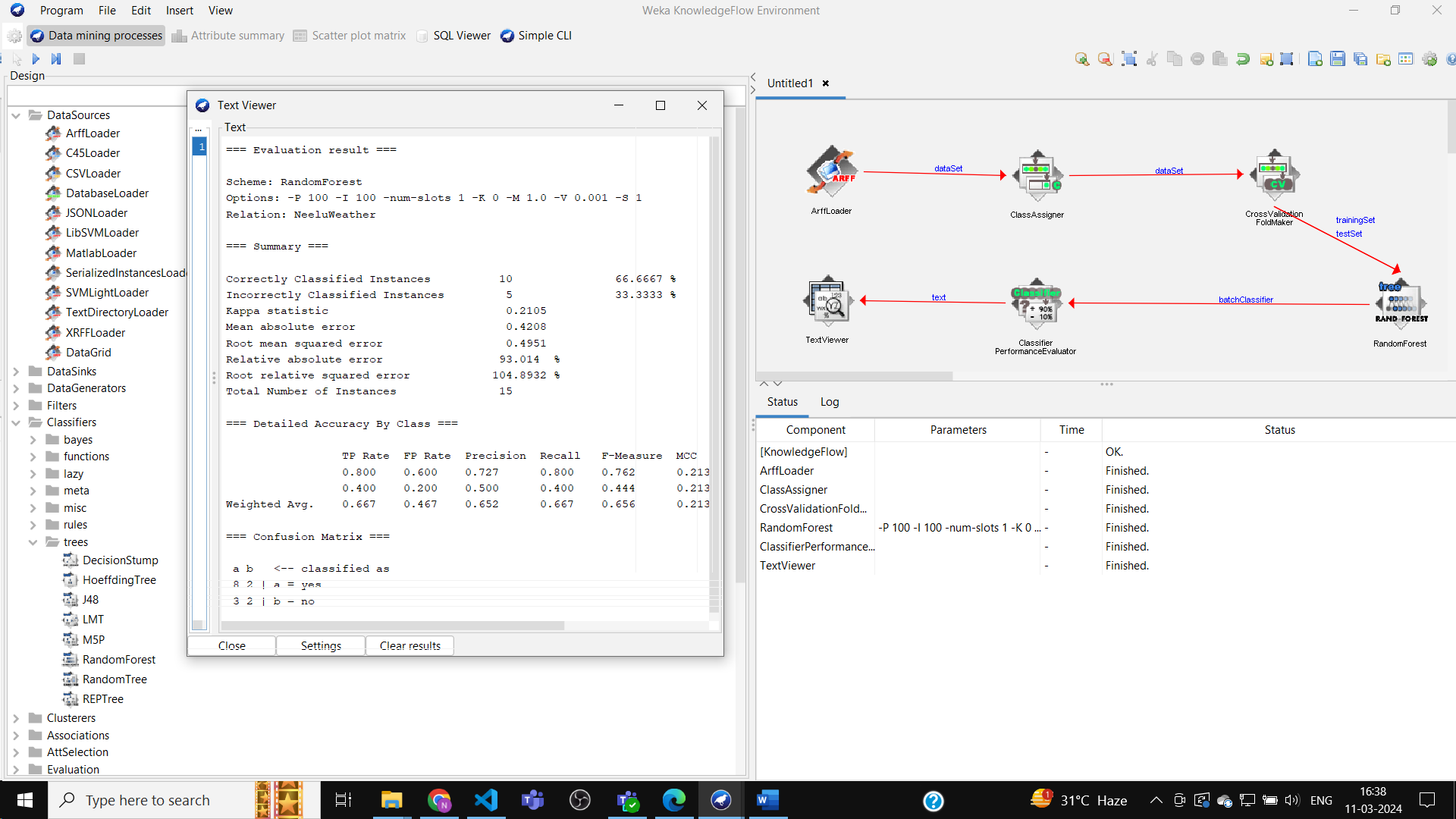
6) After discretization –



Doing the same using WEKA KnowledgeFlow –

Pre-processing and classifying the above data using tree(random forest) classifier –

The nominal class, one being determined here, is play i.e. whether to play outside or not depending on the weather outside.



After using normalization and scaling features to 1.0 and using J48 classifier :

