

## Experiment-2

- Objective:- Implementation of data preprocessing using weka explorer.
- Tool used:- weka software
- Dataset used:- IRIS dataset (.csv)
- Theory:- Data preprocessing and data cleaning are crucial steps in any data analysis or machine learning project. They involve transferring raw data into a clean, consistent and usable format that can be effectively analysed.

① Detect outliers → An outlier is a data point that differs significantly from other observations in a dataset.  
In weka, outliers can be detected using the Interquartile Range (filter / weka unsupervised /)  
Click on apply to apply the filter. Two columns would appear as outlier and extreme also, these would be indicated differently using a different colour.

② Detect & remove Null values → A Null value is represents the absence of a value or a missing value in a database or programming context in weka, go to

EXPERIMENT 2

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

None

Current relation

Relation: None

Instances: None

Attributes

All

None

Invert

Pattern

Attributes: None

Sum of weights: None

Selected attribute

Name: None

Missing: None

Weight: None

Distinct: None

Type: None

Unique: None

Apply

Stop


Remove

Visualize All

Status

Welcome to the Weka Explorer

Log

 x 0

Remove with values in the filter tab (filter / unsupervised / attribute / remove with values). Configure it to remove instances with missing values in the desired attributes.

③ Convert / transform data types → In weka, select the Numeric transform (filter / unsupervised / attribute / numeric transform) and select the instances you want to apply the transformation from.

④ fill substitute of Null values:- In weka, we can use Replace missing values (filter / unsupervised / attribute / replace missing values) click on apply button and the missing values will be replaced.

⑤ Result

• Result :- We have successfully implemented all the various operations successfully.

• Viva question?

Q1 What is meant by data preprocessing?

→ data preprocessing is the process of cleaning, transforming and organizing raw data into a suitable format for analysis for machine learning.



# EXPERIMENT 2

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Plot Matrix

outlook

temperature

humidity

windy

play

play

outlook

temperature

humidity

windy

play

outlook

temperature

humidity

windy

play

PlotSize: [89]

PointSize: [3]

Iter:

Colour play (Nom)

☐ Fast scrolling (uses more memory)

Update

Select Attributes

Subsample %: 100

Status

OK

Log

x 0

Q2

State the methods / processes of data preprocessing that can be implemented by weka?

→ Methods / process of data preprocessing that can be implemented using WEKA.

- ① Handling missing values
- ② data transformation
- ③ discretization
- ④ feature selection
- ⑤ data sampling
- ⑥ attribute construction
- ⑦ removing unwanted attribute