

Name - Yash Lakhtariya

Enrollment number - 21162101012

Branch - CBA Batch - 61

DMW Practical 4

Provide Basic information about the Weka tool (i.e. for what purpose it is used, what are facilities available in that tools, what kind of operations can be performed using it, etc.)

Weka, which stands for *Waikato Environment for Knowledge Analysis*, is a free and open-source software suite developed by the University of Waikato in New Zealand. It's a comprehensive platform designed for data mining and machine learning tasks.

- It offers facilities for data preprocessing, including handling missing values and feature transformation.
- With a vast collection of machine learning algorithms, it supports classification, regression, clustering, association rule mining, and feature selection.
- Users can conduct experiments, evaluate algorithm performance, and visualize results conveniently.
- It also provides scripting capabilities and integration with other languages for customization and extension of functionalities.
- Overall, it serves as a comprehensive platform for both beginners and experts in the field of machine learning.

Name - Yash Lakhtariya

Enrollment number - 21162101012

Branch - CBA Batch - 61

DMW Practical 4

(i) What kind of data format is accepted in Weka tool? Provide brief information about default file format with example.

There are many formats accepted in Weka tool like arff, csv, c4.5, xml, json, etc.

Default Format : ARFF (Attribute-Relation File Format) is the native data format for Weka. It's a plain text file with a specific structure that describes the data and its attributes.

Structure :

- ❖ **Header Section :** Defines the relation name (dataset name) and each attribute's name and data type (numeric, nominal, etc.).
- ❖ **Data Section :** Contains the actual data points, with each line representing an instance and values separated by commas.

Example : Take an example of a Sadhana report sheet of devotee :

```
@relation devotee_activities

@attribute devotee_id NOMINAL
@attribute chanting_rounds INTEGER
@attribute reading_duration INTEGER
@attribute service_performed BOOLEAN

@data
devotee1, 16, 60, true
devotee2, 18, 45, true
devotee3, 32, 45, false
```

Name - Yash Lakhtariya

Enrollment number - 21162101012

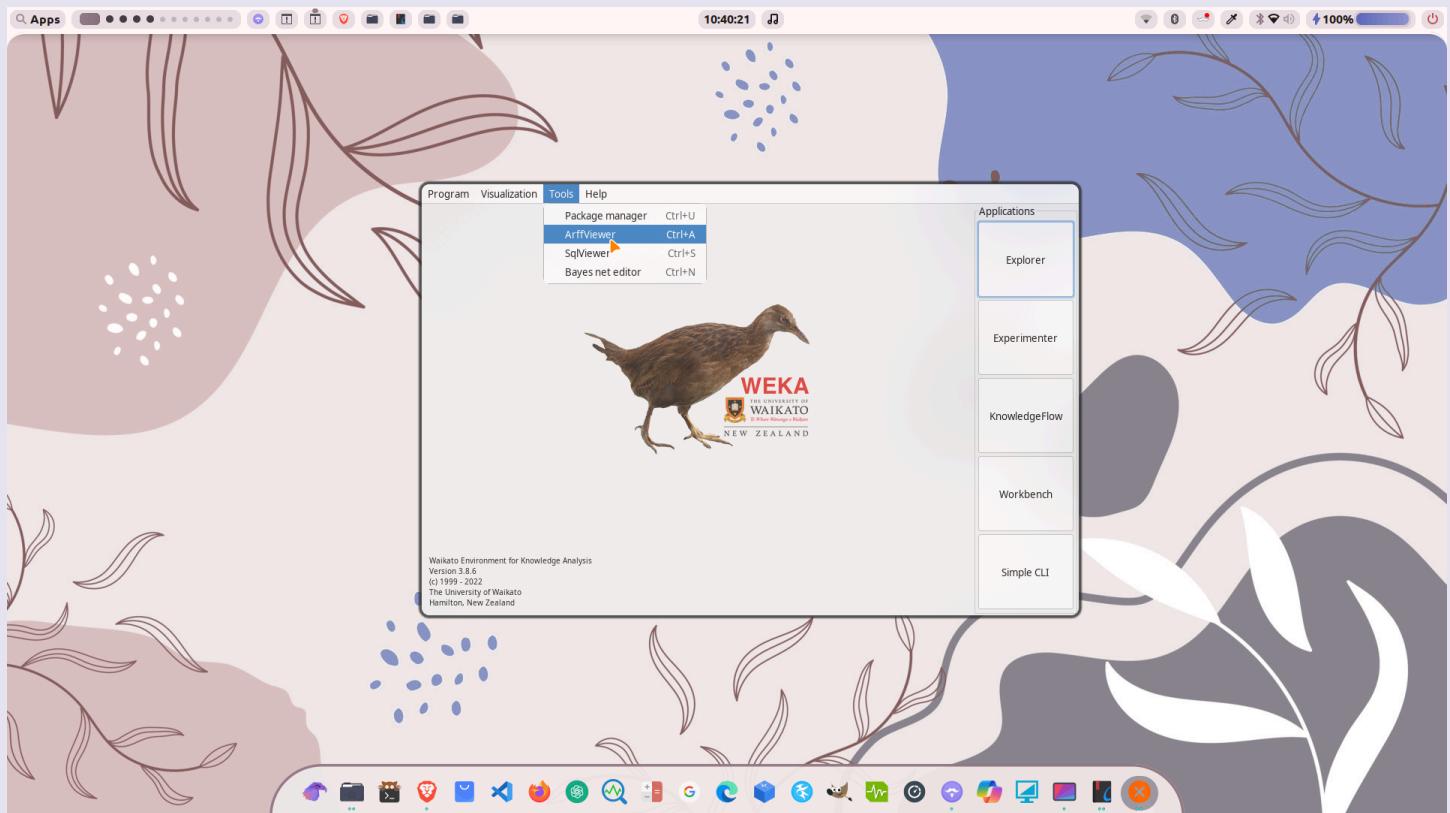
Branch - CBA Batch - 61

DMW Practical 4

(ii) Demonstrate how to convert the csv file into the file format which is accepted by Weka.

Conversion of CSV to default format of Weka (i.e. ARFF) :

➤ Open ARFF viewer in Weka



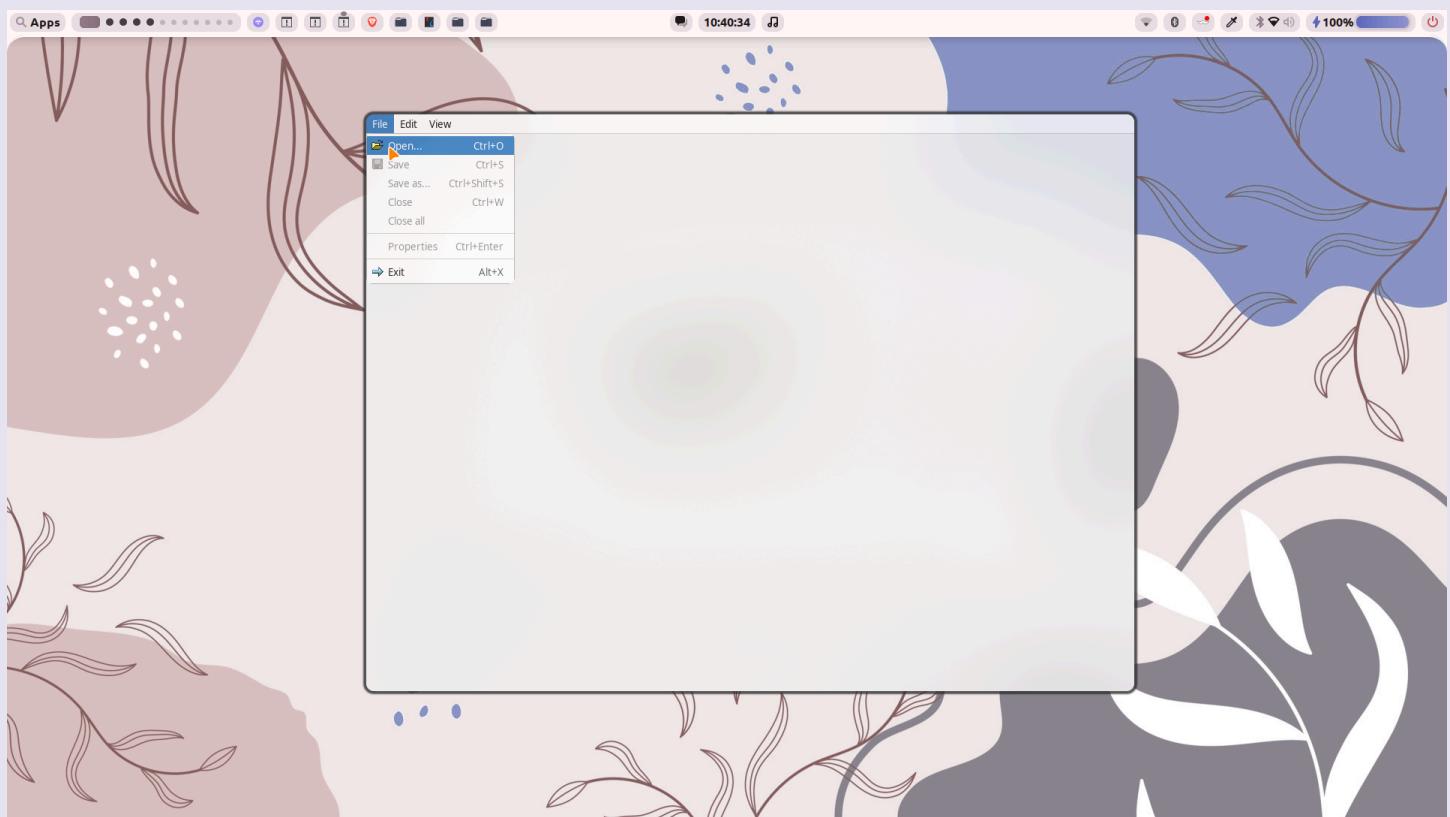
Name - Yash Lakhtariya

Enrollment number - 21162101012

Branch - CBA Batch - 61

DMW Practical 4

➤ Open CSV file in the window using File>Open

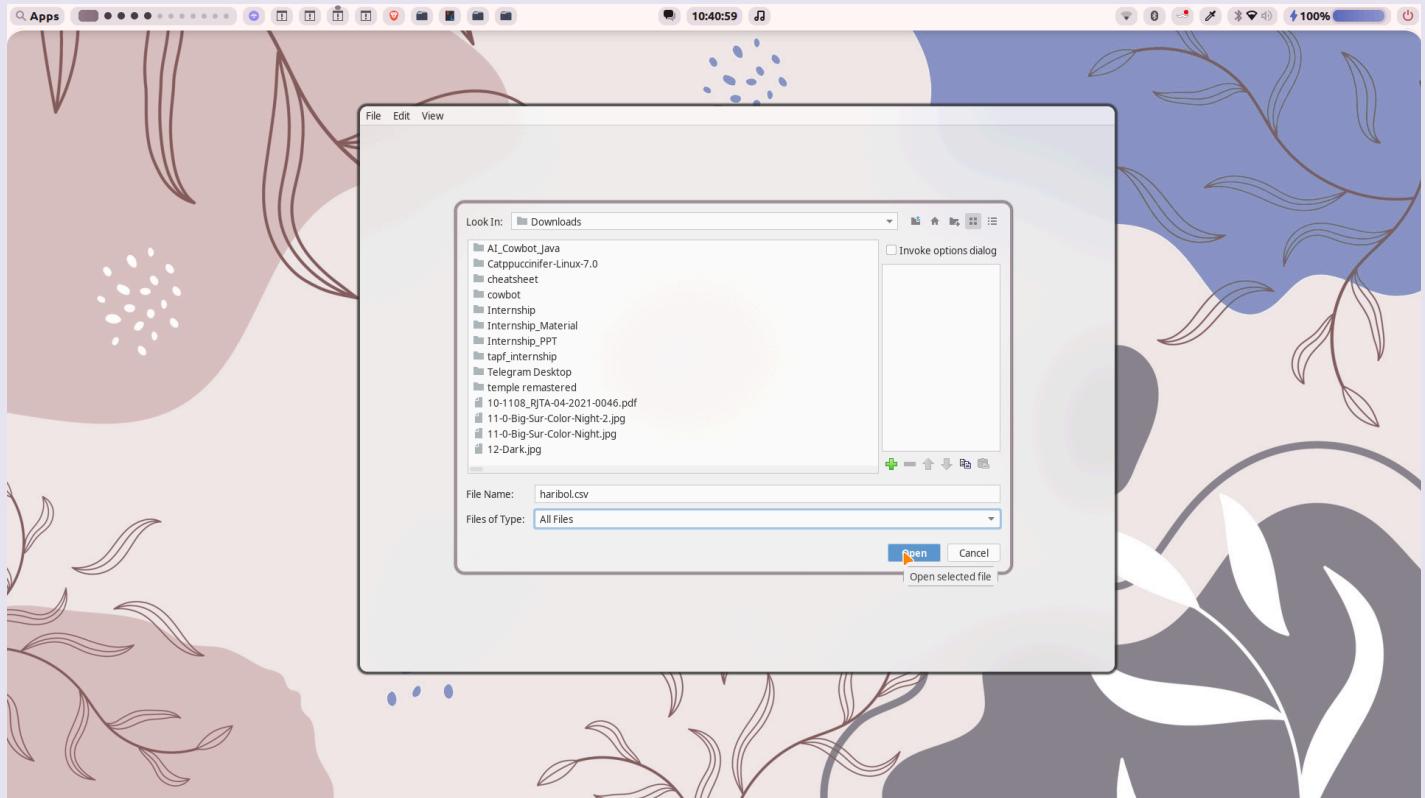


Name - Yash Lakhtariya

Enrollment number - 21162101012

Branch - CBA Batch - 61

DMW Practical 4



No.	1:id	2:name	3:age	4:contact	5:designation	6:salary
1	1.0	Madhav	21.0	8.4949...	CEO	64000.0
2	2.0	Keshav	22.0	5.9694...	Manager	32000.0
3	3.0	Govind	23.0	9.6894...	Engineer	16000.0
4	4.0	Giridhar	19.0	9.8975...	Developer	64320.0
5	5.0	Madh...	25.0	9.4869...	HR	64160.0
6	6.0	Mada...	24.0	9.4694...	Co-Founder	32160.0
7	7.0	Shya...	27.0	6.1946...	Financer	84664.0
8	8.0	Ramc...	28.0	8.7706...	Manager	64646.0
9	9.0	Vishnu	21.0	6.8984...	Manager	98980.0
10	10.0	Vaman	25.0	4.8965...	Sales	89941.0
11	11.0	Narsi...	30.0	6.8408...	Engineer	89676.0
12	12.0	Chaita...	25.0	8.4898...	Co-Founder	18494.0

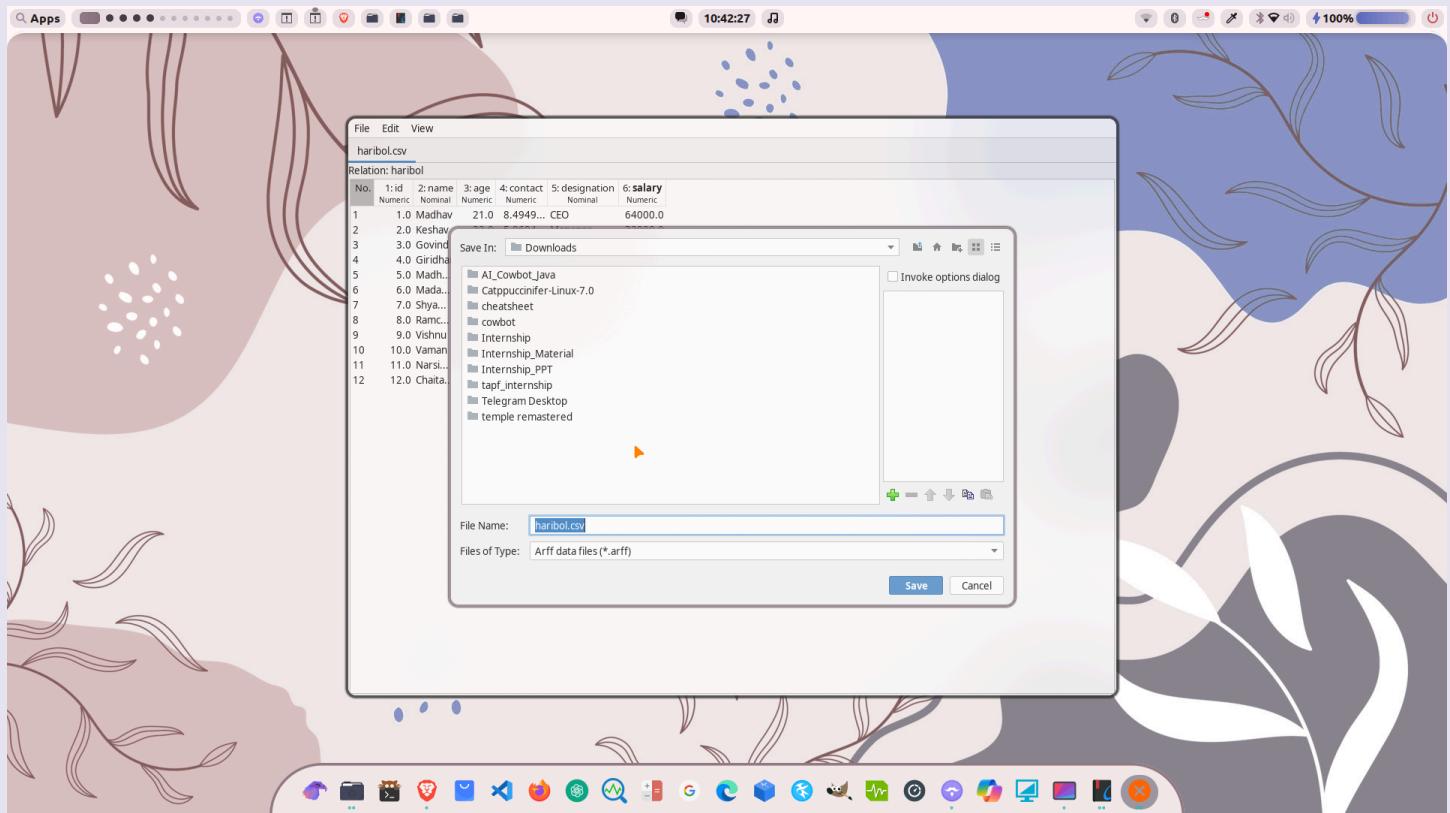
Name - Yash Lakhtariya

Enrollment number - 21162101012

Branch - CBA Batch - 61

DMW Practical 4

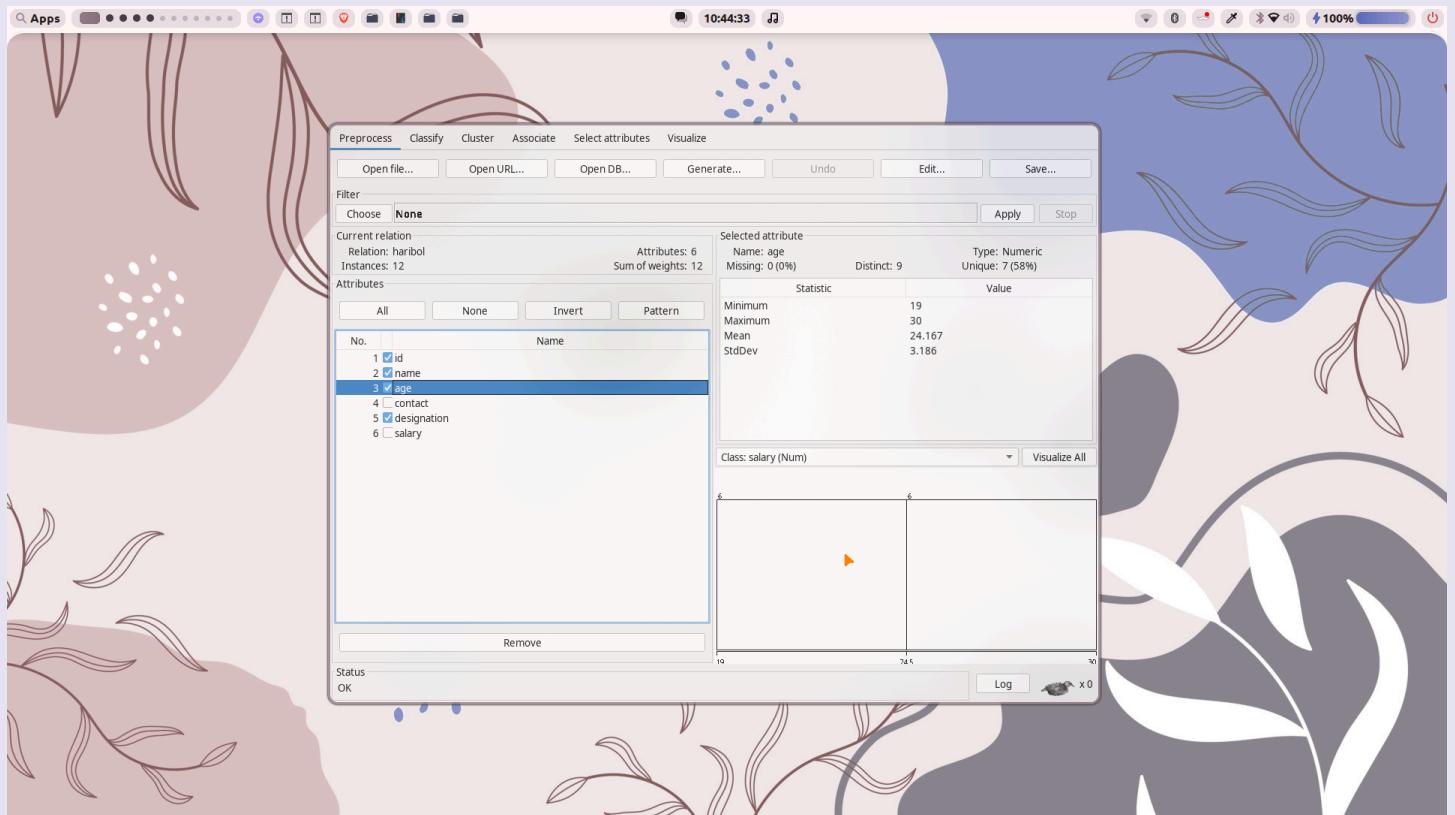
➤ Save the file with ARFF extension



Name - Yash Lakhtariya
Enrollment number - 21162101012
Branch - CBA Batch - 61
DMW Practical 4

Weka exploration :

→ Opening ARFF file in Explorer for data preprocessing, where attributes can be filtered, removed selected, inspected for min, max, avg, etc. analysis



Name - Yash Lakhtariya

Enrollment number - 21162101012

Branch - CBA Batch - 61

DMW Practical 4

→ Visualization can be done as per selecting the base attribute for analysis, say here name is taken as base attribute and every other data is based and filtered as per name

