|  |
| --- |
| Thông tin sinh viên |
| Họ và tên: Nguyễn Hùng Anh |
| Mã số sinh viên: 21064051 |
| Lớp học phần: DHHTTT17A |
| Mã lớp học phần: 420300344302 |
| Nhóm: 1 |

Contents

[Exercise 1: 2](#_Toc175476675)

[Dữ liệu 1: airline.arff 2](#_Toc175476676)

[Dữ liệu 2: iris.2D.arff 6](#_Toc175476677)

[Dữ liệu 3: ReutersCorn-test 10](#_Toc175476678)

[Dữ liệu 4: ReutersCorn-train 13](#_Toc175476679)

[Dữ liệu 5: ReutersGrain-test 16](#_Toc175476680)

[Exercise 2 19](#_Toc175476681)

# Exercise 1:

## Dữ liệu 1: airline.arff

- What’s the name of the relation?

Airlie\_passengers

- Number of records in each dataset

144

- How many instances does each dataset have?

144

- How many attributes?

2

- What are the values that each attribute can have?

Passerger\_Nuber: Numeric

Date: Date

- Missing: The number (and percentage) of instances in the data for which this attribute is missing (unspecified)

Passerger\_Nuber:0

Date: 0

- Distinct: The number of different values that the data contains for this attribute.

Passerger\_Nuber :6

Date: 144

- Unique: The number (and percentage) of instances in the data having a value for this attribute that no other instances have.

Passerger\_Nuber: 1

Date: 144

- Show the basic statistics on the attributes for the categorical attributes the frequency of each attribute, or for the continuous attributes, show the value of min, max, mean, standard deviation and deviation etc.,

Passerger\_Nuber: A screenshot of a computer

Description automatically generated

Date: A screenshot of a computer

Description automatically generated

- Identify the class attribute (if any). Determine the number of records for each class

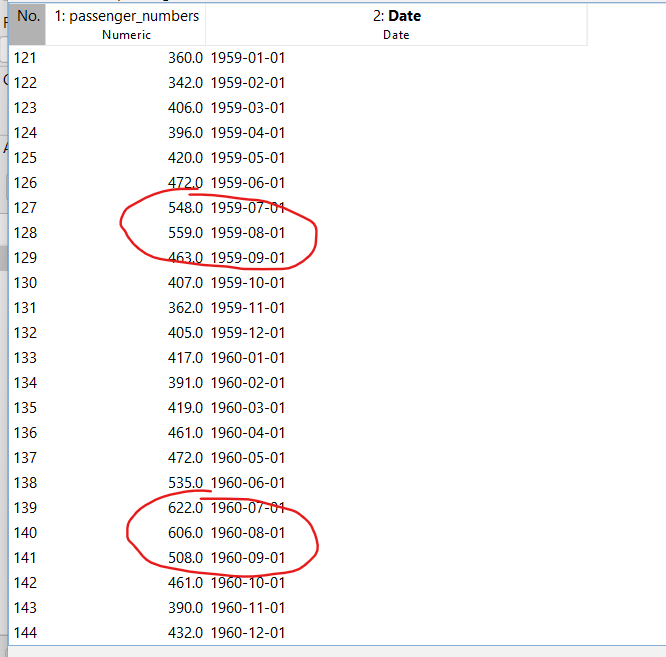
- Plot HistogramA screenshot of a computer screen

Description automatically generated

- Visualize the data in various dimensions. What happens if the class attribute is left on the first column? Compare the visualizations of the 2 datasets (the one with class on the first column, before any manipulation, and the one with the class on the last column) as they appear on the right side of the Preprocessing panel: can you notice any differences?

Do data 2 chọn class không có quan hệ nào với các thuộc tính khác hoặc do data 2 chọn thuộc tính không phải là kiểu class nên không thể hiện được bất cứ liên hệ nào

- Examining data throught a viewer for editing



Dữ liệu sau khi edit bằng cách sort ngày thì dữ liệu của số lượng người lên máy bay tăng mạnh từ tháng 7, 8, 9

## Dữ liệu 2: iris.2D.arff

- What’s the name of the relation?

iris-weka.filters.unsupervised.attribute.Remove-R1-2- Number of records in each dataset

- How many instances does each dataset have?

150

- How many attributes?

150

- What are the values that each attribute can have?

Petallength: Numeric

petalwidth: Numeric   
Class: Nominal

- Missing: The number (and percentage) of instances in the data for which this attribute is missing (unspecified)

petallength:0

petalwidth:0

Class:0

- Distinct: The number of different values that the data contains for this attribute.

petallength:43

petalwidth:22

Class:3

- Unique: The number (and percentage) of instances in the data having a value for this attribute that no other instances have.

petallength: 10

petalwidth:2

Class:0

- Show the basic statistics on the attributes for the categorical attributes the frequency of each attribute, or for the continuous attributes, show the value of min, max, mean, standard deviation and deviation etc.,

petallength: A screenshot of a computer

Description automatically generated

petalwidth: A screenshot of a computer

Description automatically generated

- Identify the class attribute (if any). Determine the number of records for each class

Class: A screenshot of a computer

Description automatically generated

- Plot Histogram

A screenshot of a graph

Description automatically generated

- Visualize the data in various dimensions. What happens if the class attribute is left on the first column? Compare the visualizations of the 2 datasets (the one with class on the first column, before any manipulation, and the one with the class on the last column) as they appear on the right side of the Preprocessing panel: can you notice any differences?

Do data 2 chọn class không có quan hệ nào với các thuộc tính khác hoặc do data 2 chọn thuộc tính không phải là kiểu class nên không thể hiện được bất cứ liên hệ nào

- Examining data throught a viewer for editing

Trước Edit

A screenshot of a computer

Description automatically generated

Sau Edit giảm dần theo thuộc tình petalwidth

A screenshot of a computer

Description automatically generated

## Dữ liệu 3: ReutersCorn-test

- What’s the name of the relation?

Reuters-21578 Corn ModApte Test-weka.filters.unsupervised.attribute.NumericToBinary-weka.filters.unsupe.

- Number of records in each dataset

604

- How many instances does each dataset have?

604

- How many attributes?

2

- What are the values that each attribute can have?

Text: String

class-att: Nominal

- Missing: The number (and percentage) of instances in the data for which this attribute is missing (unspecified)

Text: 0

class-att: 0

- Distinct: The number of different values that the data contains for this attribute.

Text: 602

class-att: 2

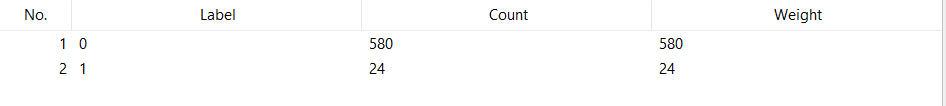
- Unique: The number (and percentage) of instances in the data having a value for this attribute that no other instances have.

Text: 0

class-att: 0

- Show the basic statistics on the attributes for the categorical attributes the frequency of each attribute, or for the continuous attributes, show the value of min, max, mean, standard deviation and deviation etc.,

Text:

class-att: 

- Identify the class attribute (if any). Determine the number of records for each class

- Plot Histogram A screenshot of a computer

Description automatically generated

- Visualize the data in various dimensions. What happens if the class attribute is left on the first column? Compare the visualizations of the 2 datasets (the one with class on the first column, before any manipulation, and the one with the class on the last column) as they appear on the right side of the Preprocessing panel: can you notice any differences?

Do data 2 chọn class không có quan hệ nào với các thuộc tính khác hoặc do data 2 chọn thuộc tính không phải là kiểu class nên không thể hiện được bất cứ liên hệ nào

- Examining data throught a viewer for editing

Thực hiện chỉnh sửa Class att bằng cách sort thì không tìm thấy gì xảy ra

## Dữ liệu 4: ReutersCorn-train

- What’s the name of the relation?

Reuters-21578 Corn ModApte Train-weka.filters.unsupervised.attribute. NumericToBinary-weka.filters.unsu..

- Number of records in each dataset

1554

- How many instances does each dataset have?

1554

- How many attributes?

2

- What are the values that each attribute can have?

Text: String

class-att: Nominal

- Missing: The number (and percentage) of instances in the data for which this attribute is missing (unspecified)

Text: 0

class-att: 0

- Distinct: The number of different values that the data contains for this attribute.

Text: 1549

class-att: 2

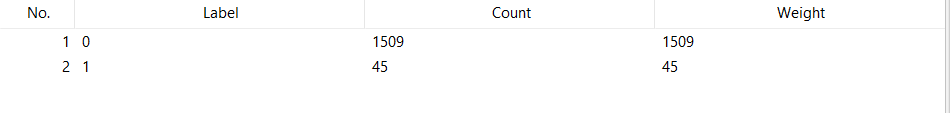
- Unique: The number (and percentage) of instances in the data having a value for this attribute that no other instances have.

Text: 1544

class-att: 0

- Show the basic statistics on the attributes for the categorical attributes the frequency of each attribute, or for the continuous attributes, show the value of min, max, mean, standard deviation and deviation etc.,

Text:

class-att: 

- Identify the class attribute (if any). Determine the number of records for each class

- Plot Histogram A screenshot of a computer

Description automatically generated

- Visualize the data in various dimensions. What happens if the class attribute is left on the first column? Compare the visualizations of the 2 datasets (the one with class on the first column, before any manipulation, and the one with the class on the last column) as they appear on the right side of the Preprocessing panel: can you notice any differences?

Do data 2 chọn class không có quan hệ nào với các thuộc tính khác hoặc do data 2 chọn thuộc tính không phải là kiểu class nên không thể hiện được bất cứ liên hệ nào

- Examining data throught a viewer for editing

Thực hiện chỉnh sửa Class att bằng cách sort thì không tìm thấy gì xảy ra

## Dữ liệu 5: ReutersGrain-test

- What’s the name of the relation?

Reuters-21578 Grain ModApte Test-weka.filters.unsupervised.attribute.NumericToBinary-weka.filters.unsup..

- Number of records in each dataset

604

- How many instances does each dataset have?

604

- How many attributes?

2

- What are the values that each attribute can have?

Text: String

class-att: Nominal

- Missing: The number (and percentage) of instances in the data for which this attribute is missing (unspecified)

Text: 0

class-att: 0

- Distinct: The number of different values that the data contains for this attribute.

Text: 602

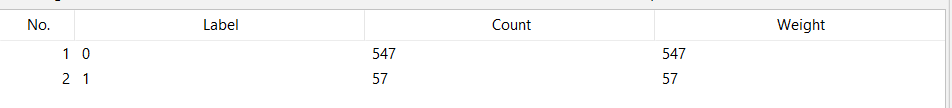
class-att: 2

- Unique: The number (and percentage) of instances in the data having a value for this attribute that no other instances have.

Text: 600

class-att: 0

- Show the basic statistics on the attributes for the categorical attributes the frequency of each attribute, or for the continuous attributes, show the value of min, max, mean, standard deviation and deviation etc.,  
Text:

class-att: 

- Identify the class attribute (if any). Determine the number of records for each class

- Plot Histogram A screenshot of a computer

Description automatically generated

- Visualize the data in various dimensions. What happens if the class attribute is left on the first column? Compare the visualizations of the 2 datasets (the one with class on the first column, before any manipulation, and the one with the class on the last column) as they appear on the right side of the Preprocessing panel: can you notice any differences?

Do data 2 chọn class không có quan hệ nào với các thuộc tính khác hoặc do data 2 chọn thuộc tính không phải là kiểu class nên không thể hiện được bất cứ liên hệ nào

- Examining data throught a viewer for editing

Thực hiện chỉnh sửa Class att bằng cách sort thì không tìm thấy gì xảy ra

# Exercise 2

A screenshot of a computer program

Description automatically generated