

DATA MINING LAB ASSIGNMENT – 1

GERMAN CREDIT DATASET:-

OBSERVATIONS:-

The german_credit dataset classifies people to either good or bad credit risks.

Filter

Choose None

Current relation

Relation: german_credit

Instances: 1000

Attributes: 21

Sum of weights: 1000

Attributes

All None Invert Pattern

No.	Name
1	checking_status
2	duration
3	credit_history
4	purpose
5	credit_amount
6	savings_status
7	employment
8	installment_commitment
9	personal_status
10	other_parties
11	residence_since
12	property_magnitude
13	age
14	other_payment_plans
15	housing
16	existing_credits
17	job
18	num_dependents
19	own_telephone
20	foreign_worker
21	class

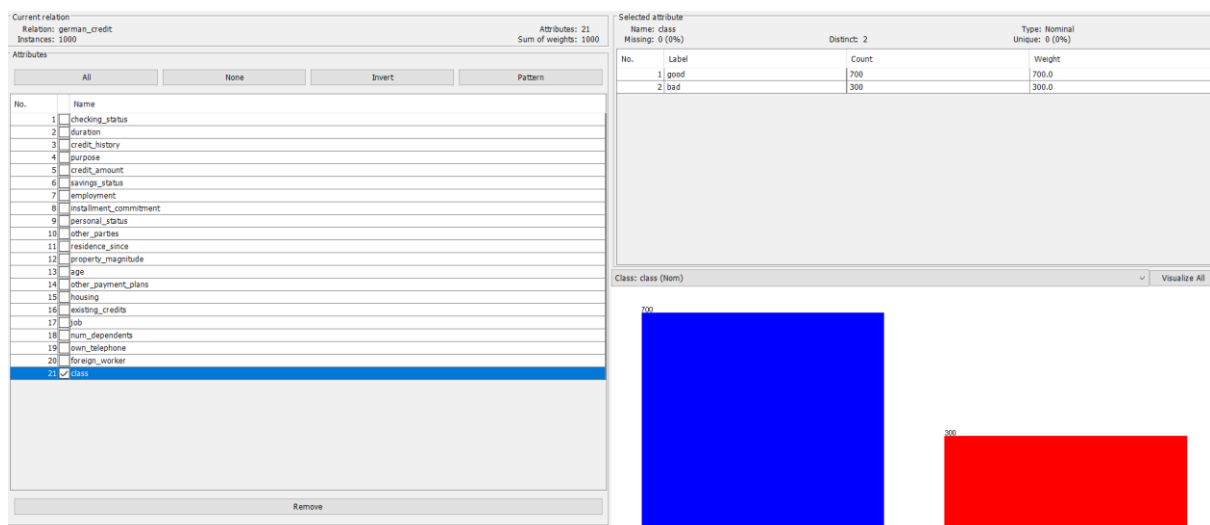
Number of records, attribute names and types:-

The german credit dataset consist of 1000 records with 21 attributes.

Attribute Name	Attribute Type
checking_status	Nominal
duration	Numeric
credit_history	Nominal
purpose	Nominal
credit_amount	Numeric
savings_status	Nominal
employment	Nominal

Installment_commitment	Numeric
Personal_status	Nominal
Other_parties	Nominal
residence_since	Numeric
property_magnitude	Nominal
age	Numeric
other_payment_plans	Nominal
housing	Nominal
existing_credits	Numeric
job	Nominal
num_dependents	Numeric
own_telephone	Nominal
foreign_worker	Nominal
class	Nominal

Class attribute in the dataset:



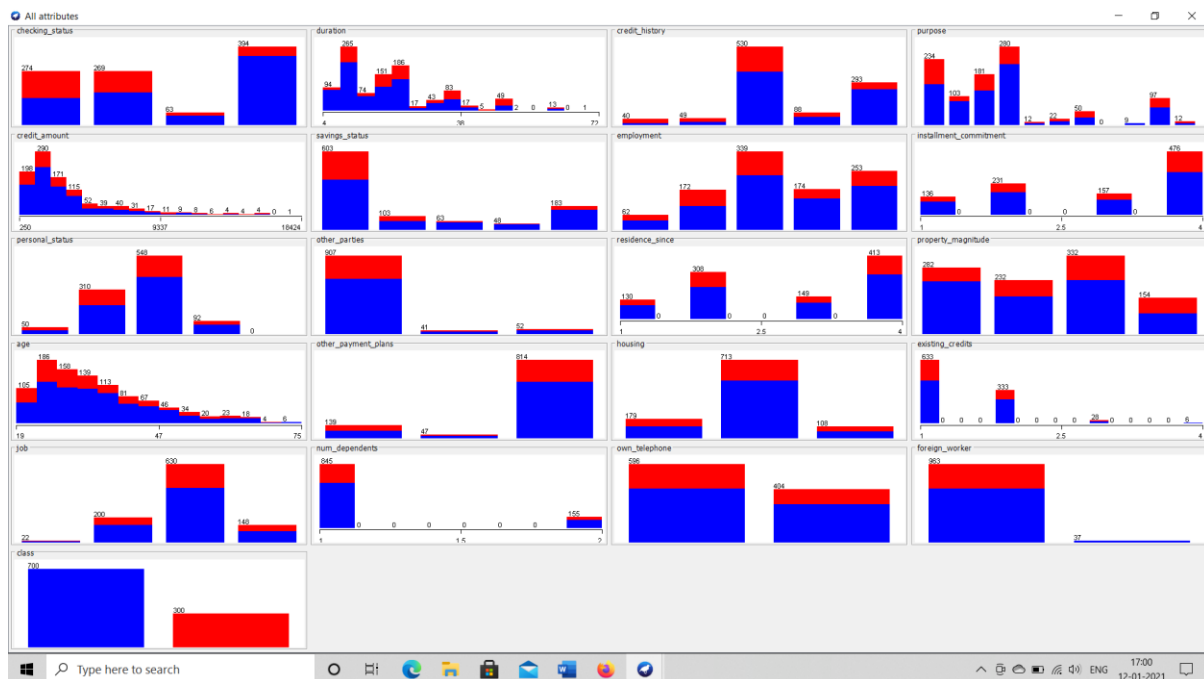
Number of records for class:

Good – 700

Bad – 300

Selected attribute			
Name: class		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
		Distinct: 2	
No.	Label	Count	Weight
1	good	700	700.0
2	bad	300	300.0

Plot histogram:-



Blue colour represents good and red colour represents bad class.

IRIS DATASET:-

OBSERVATIONS:-

It is a dataset through which we classify flowers into 3 categories.

Number of records, attribute names and types:

Current relation

Relation: iris

Instances: 150

Attributes: 5

Sum of weights: 150

Attributes

All

None

Invert

Pattern

No.

1

2

3

4

5

☐

☐

☐

☐

☐

sepal.length

sepal.width

petal.length

petal.width

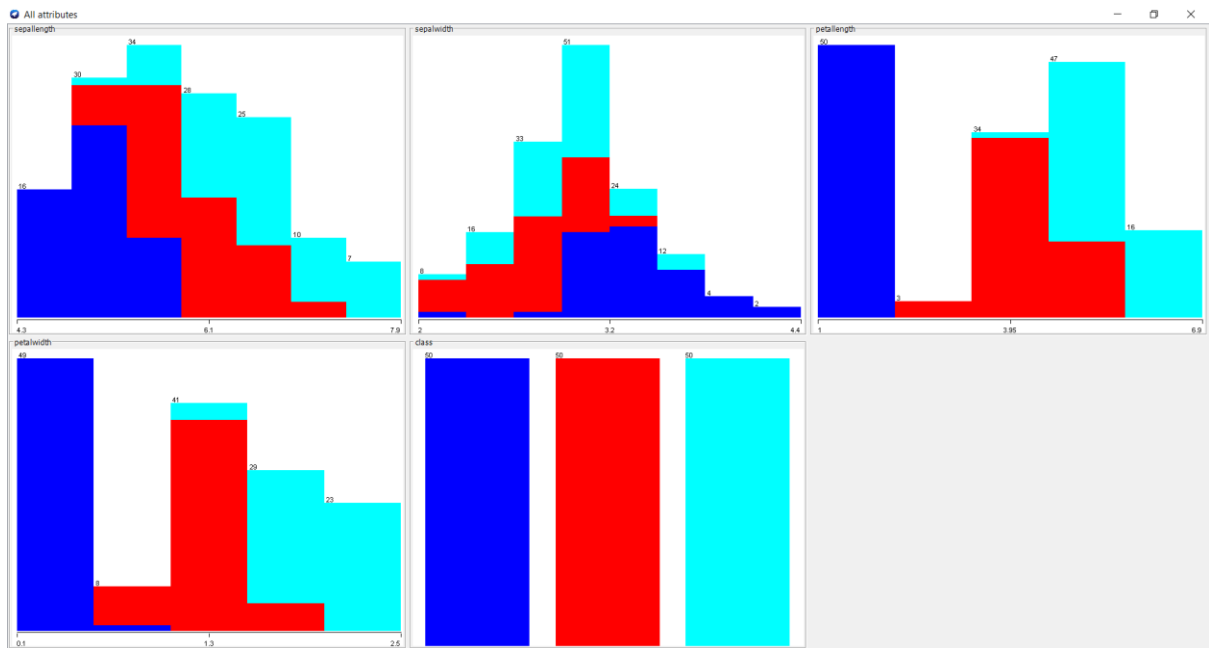
class

The iris dataset consists of 150 records with 5 attributes.

Attribute Name	Attribute Type
sepal.length	Numeric

sepalwidth	Numeric
petallength	Numeric
petalwidth	Numeric
class	Nominal

Plot Histogram:



There are 3 classes in the dataset. Where blue represents Iris-setosa, red represents Iris-versicolor and light blue represents Iris-virginica.