

**Experiment No.3**

**Title:** Execution of classification algorithm using Rapidminer.

# Batch:B4 Roll No.:16010420121 Experiment No.:3

**Aim:** Execution of data mining algorithm using RapidMiner.

**Resources needed:** Any RDBMS, Java

# Theory:

Rapidminer is a collection of open source of many data mining and machine learning algorithms, including,

* pre-processing on data
* Classification
* clustering
* Association rule extraction



A dataset is a collection of examples, each one of class Instance. Each Instance consists of a number of attributes, any of which can be nominal (= one of a predefined list of values), numeric (= a real or integer number) or a string (= an arbitrary long list of characters, enclosed in "double quotes"). The external representation of an Instances class is an ARFF file, which consists of a header describing the attribute types and the data as comma- separated list.

**Rapidminer Main Features:**

Main features are as follows:

* 49 data pre-processing tools
  + 76 classification/regression algorithms
  + 8 clustering algorithms
  + 15 attribute/subset evaluators + 10 search algorithms for feature selection.
  + 3 algorithms for finding association rules
  + 3 graphical user interfaces

# The Explorer (exploratory data analysis)

Used for pre-processing, attribute selection, learning, visualization

# The Experimenter (experimental environment)

Used for testing and evaluating machine learning algorithms

# The Knowledge Flow (new process model inspired interface)

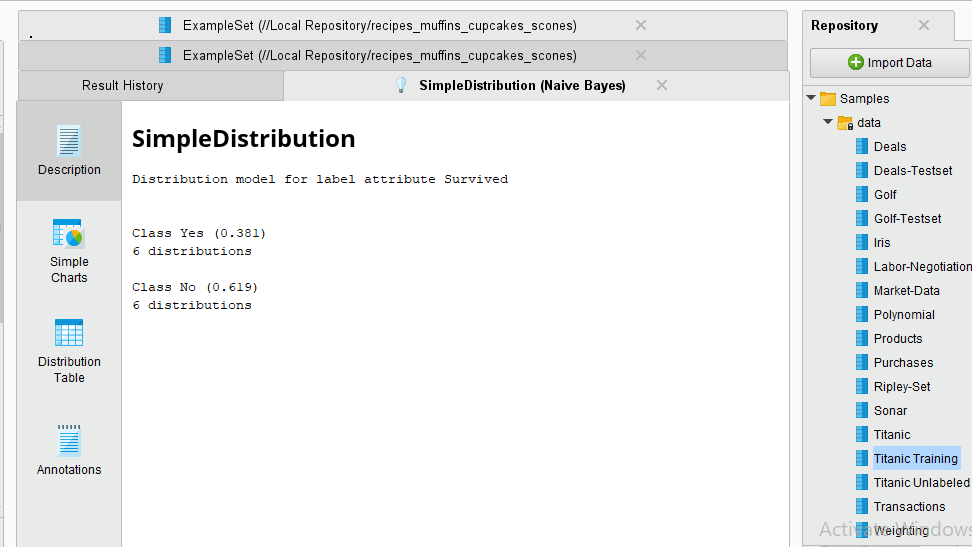
Used for visual design of KDD process

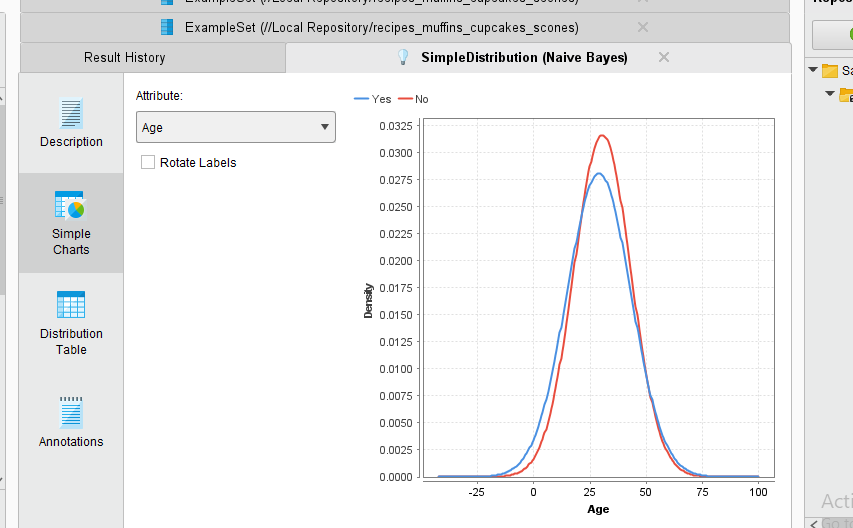
# Procedure / Approach /Algorithm / Activity Diagram:

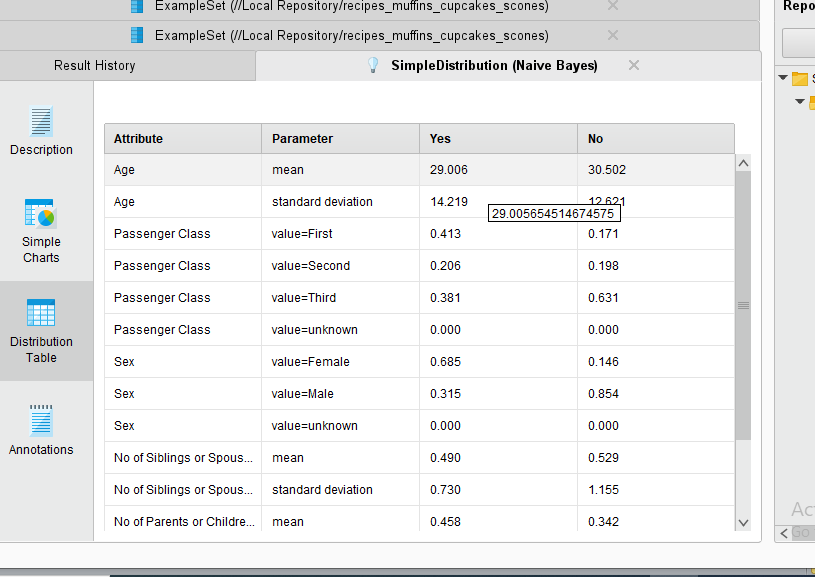
1. Execute any two data mining classification algorithm using Rapidminer tool
2. Analyze the results produced by Rapidminer

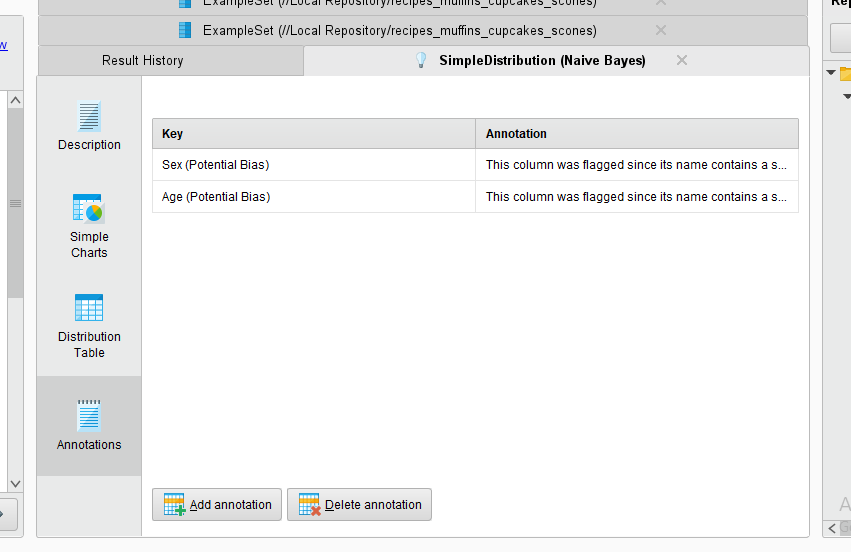
# Results: (Program printout with output / Document printout as per the format)

# Naïve Bayes algo on titanic dataset

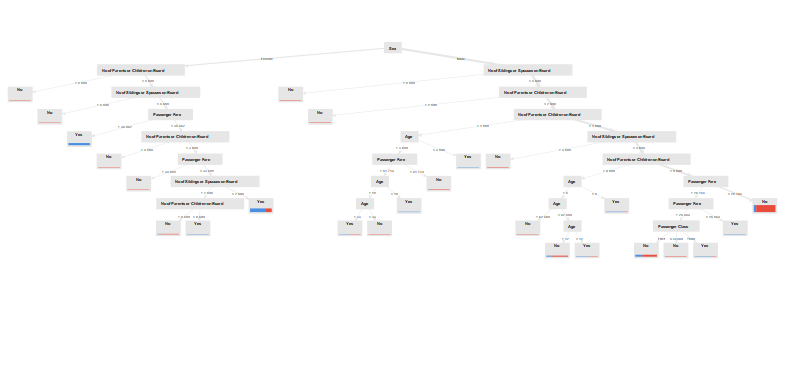
****

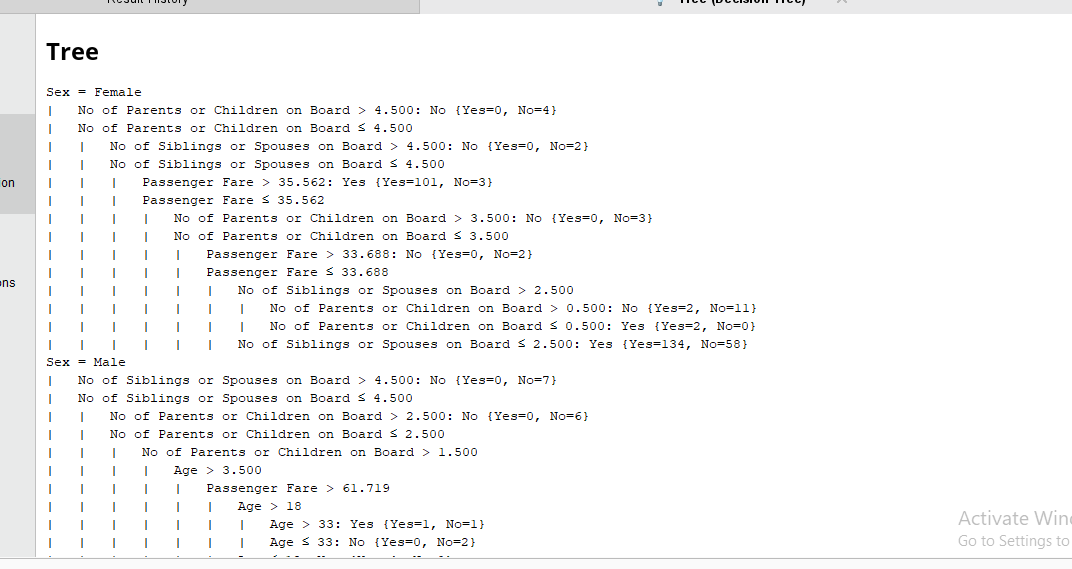
****

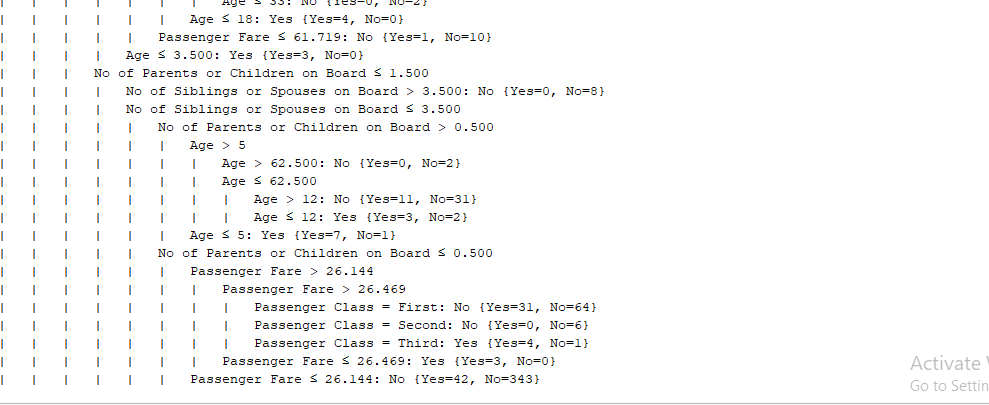
****

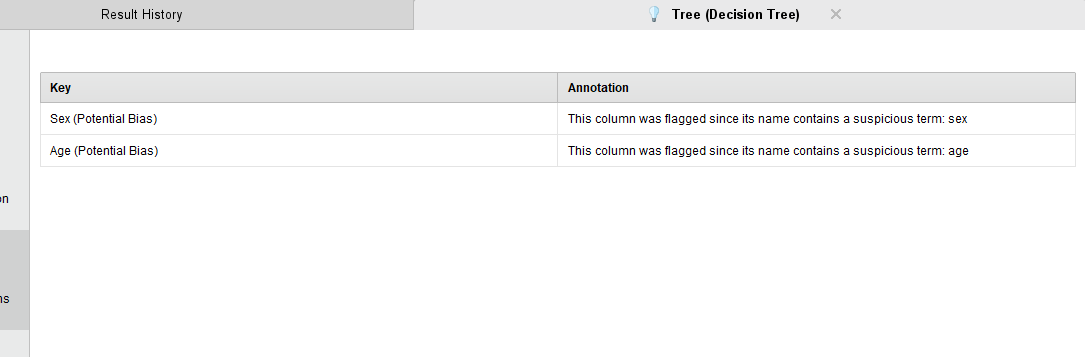
****

**Decision tree on titanic dataset:**

****

****

****

****

**Questions:**

* 1. List any five open sources / freeware tools available for data mining.



# Outcomes:

**Conclusion: (Conclusion to be based on the objectives and outcomes achieved)**

# Grade: AA / AB / BB / BC / CC / CD /DD

Signature of faculty in-charge with date

# References:

Books/ Journals/ Websites:

1. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3nd Edition

