KJSCE/IT/LY/SEM-VII/ML/2022-23

**Experiment No.3** 

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

.

KJSCE/IT/LY/SEM-VII/ML/2022-23

**Batch: B1 Roll No.: 1914078 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)**

(Autonomous College Affiliated to University of Mumbai)

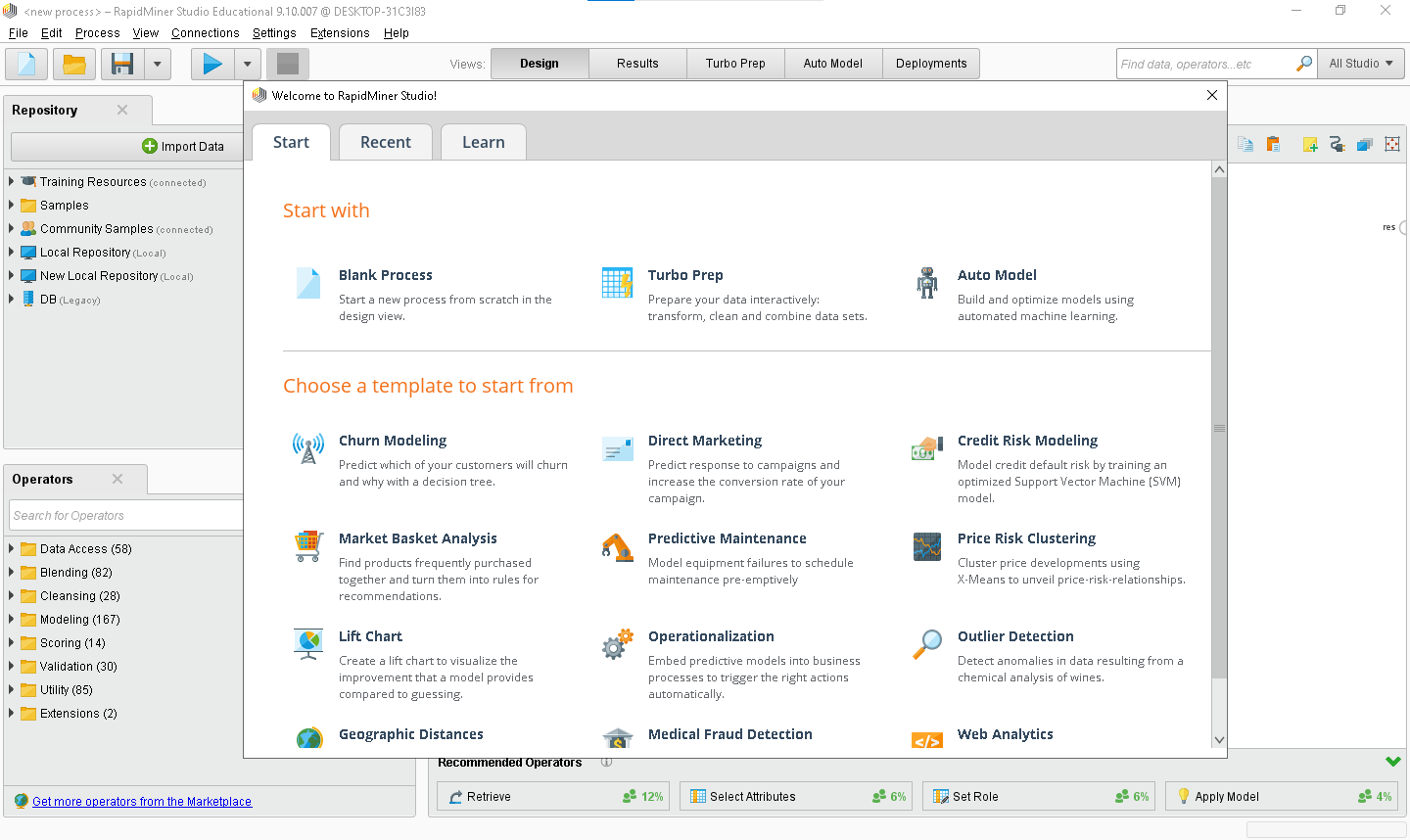
KJSCE/IT/LY/SEM-VII/ML/2022-23

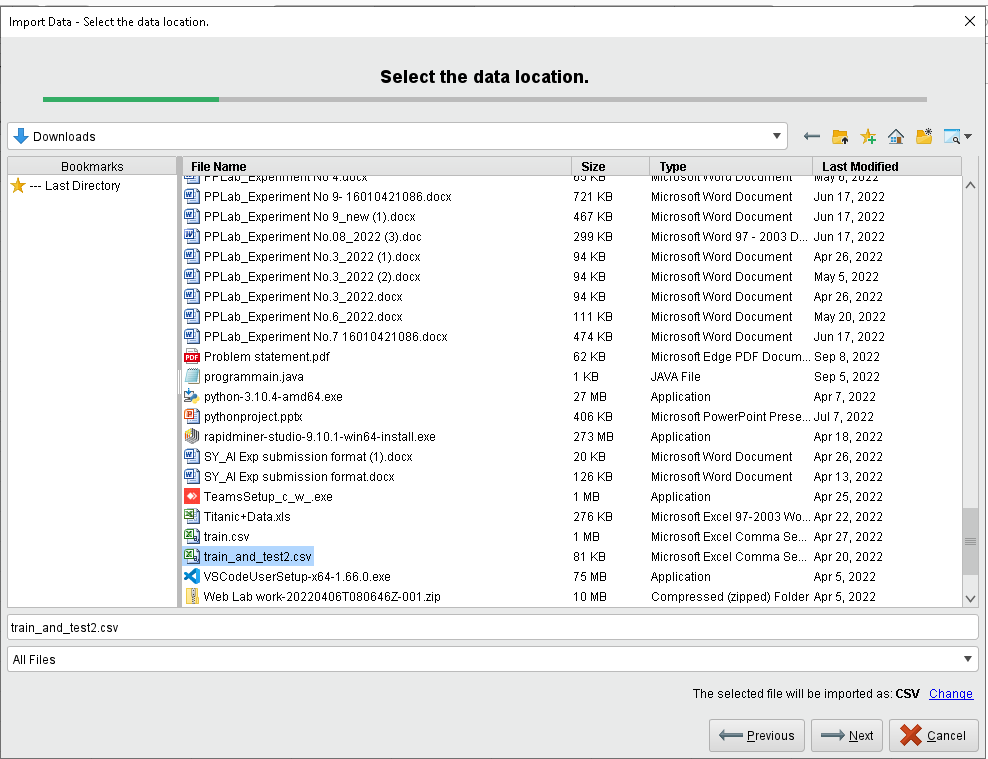
Used for visual design of KDD process

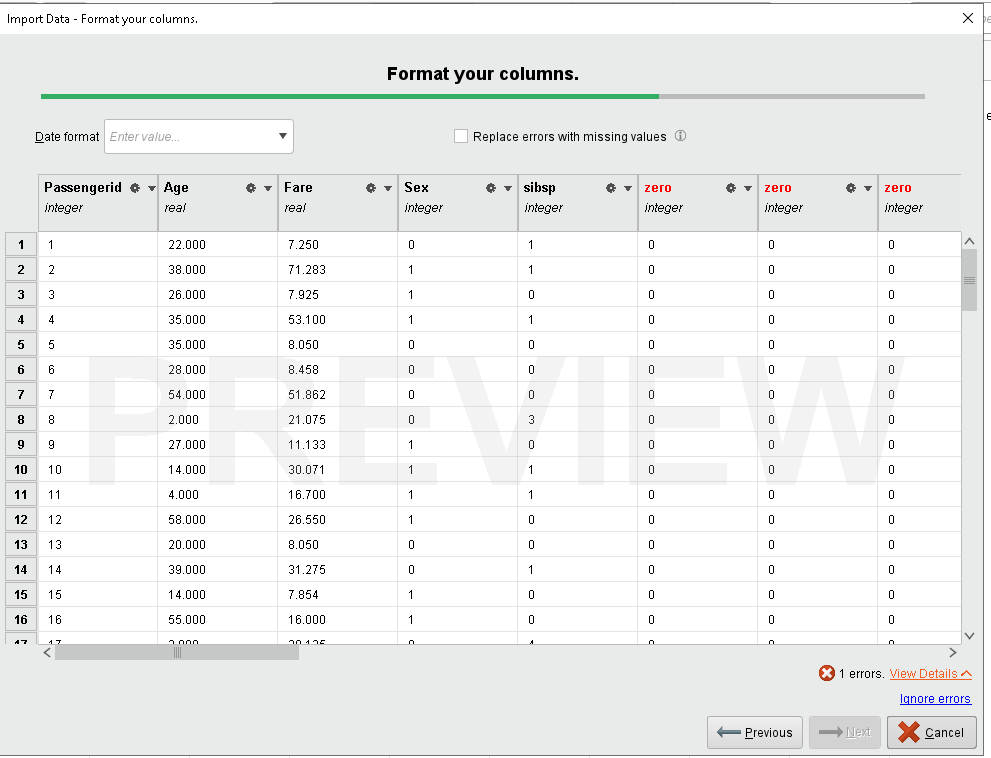
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Procedure / Approach /Algorithm / Activity Diagram:**

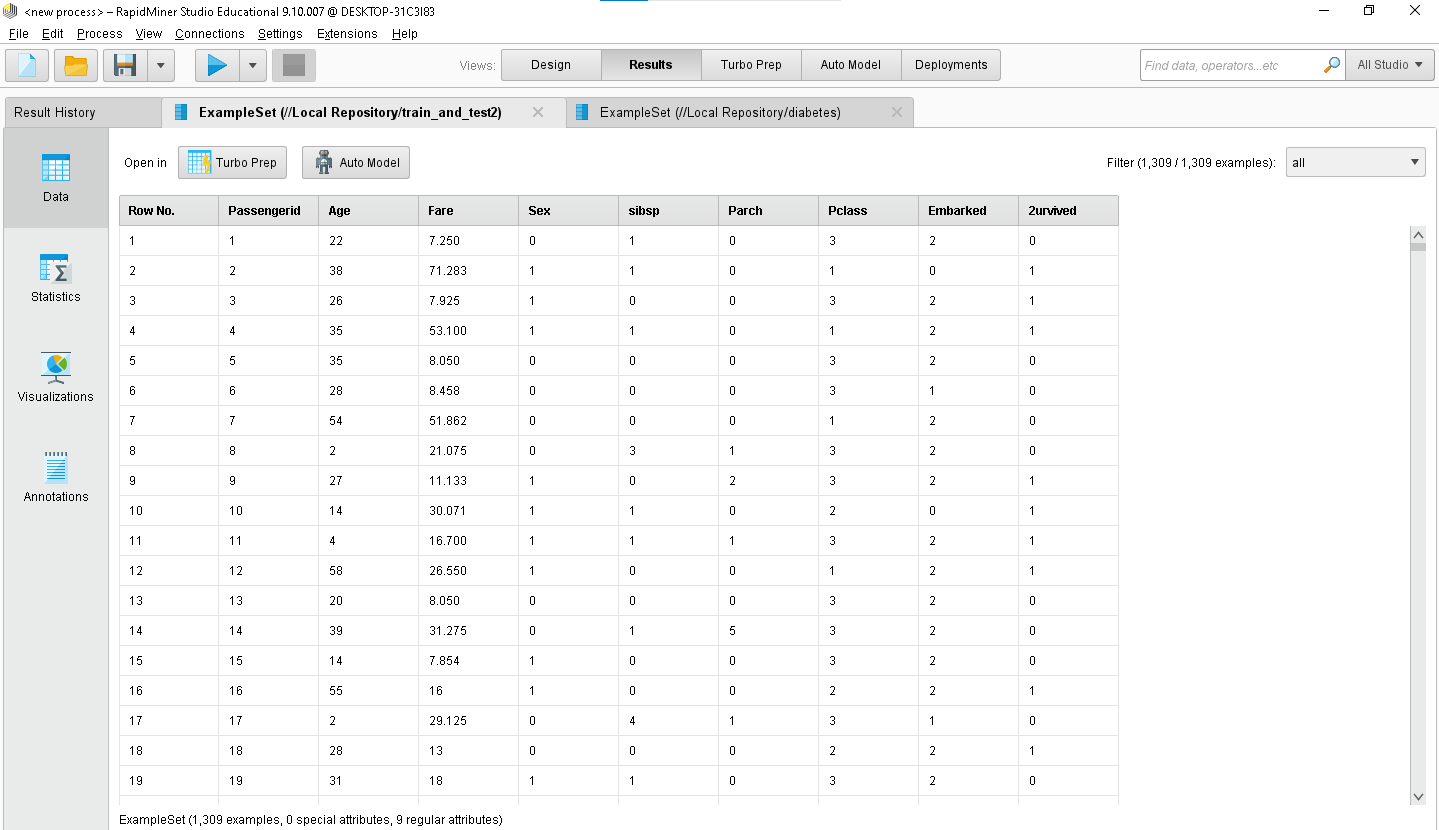
1. Execute any two data mining classification algorithm using Rapidminer tool 2. Analyze the results produced by Rapidminer

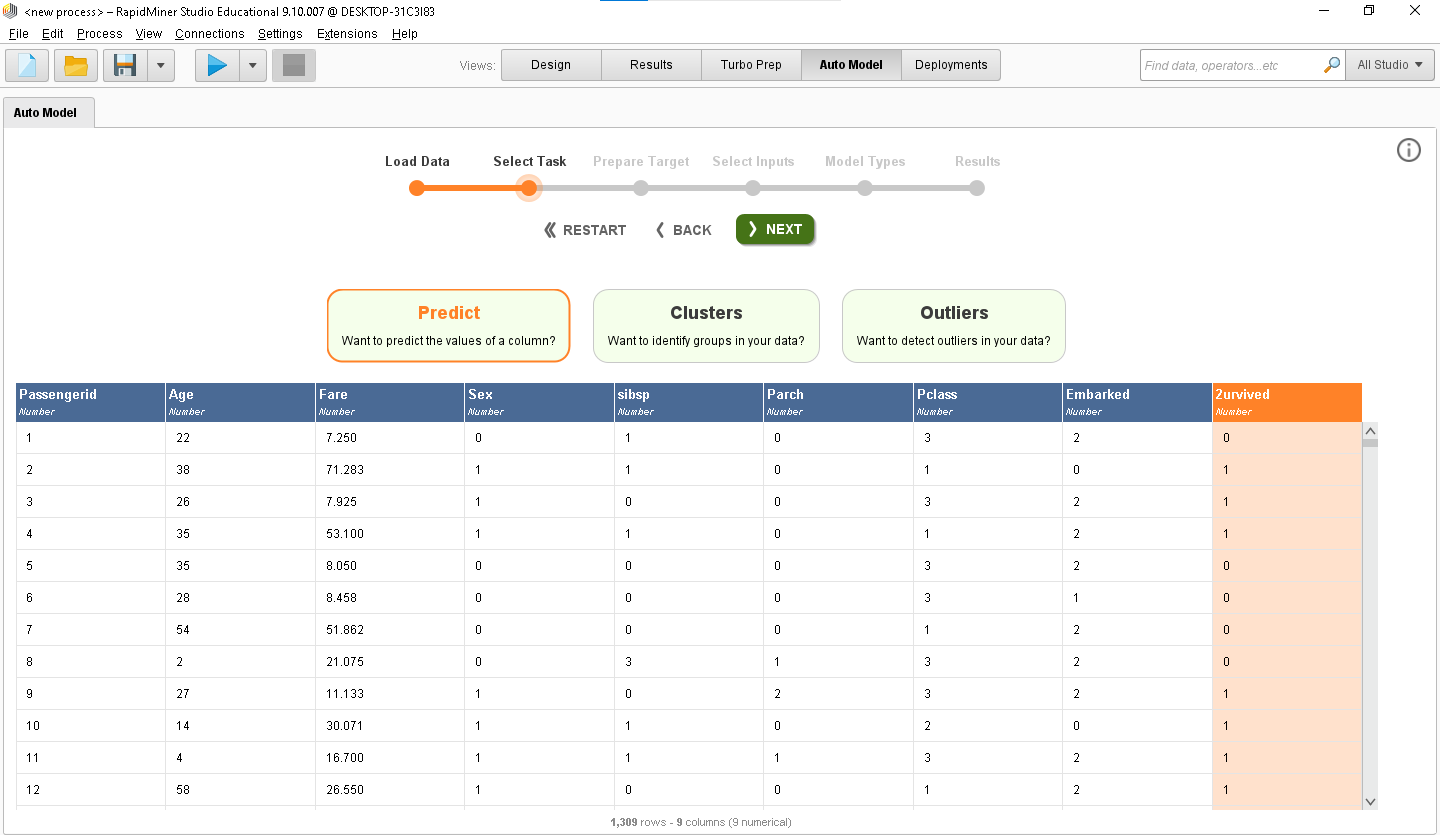
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Results: (Program printout with output / Document printout as per the format)**

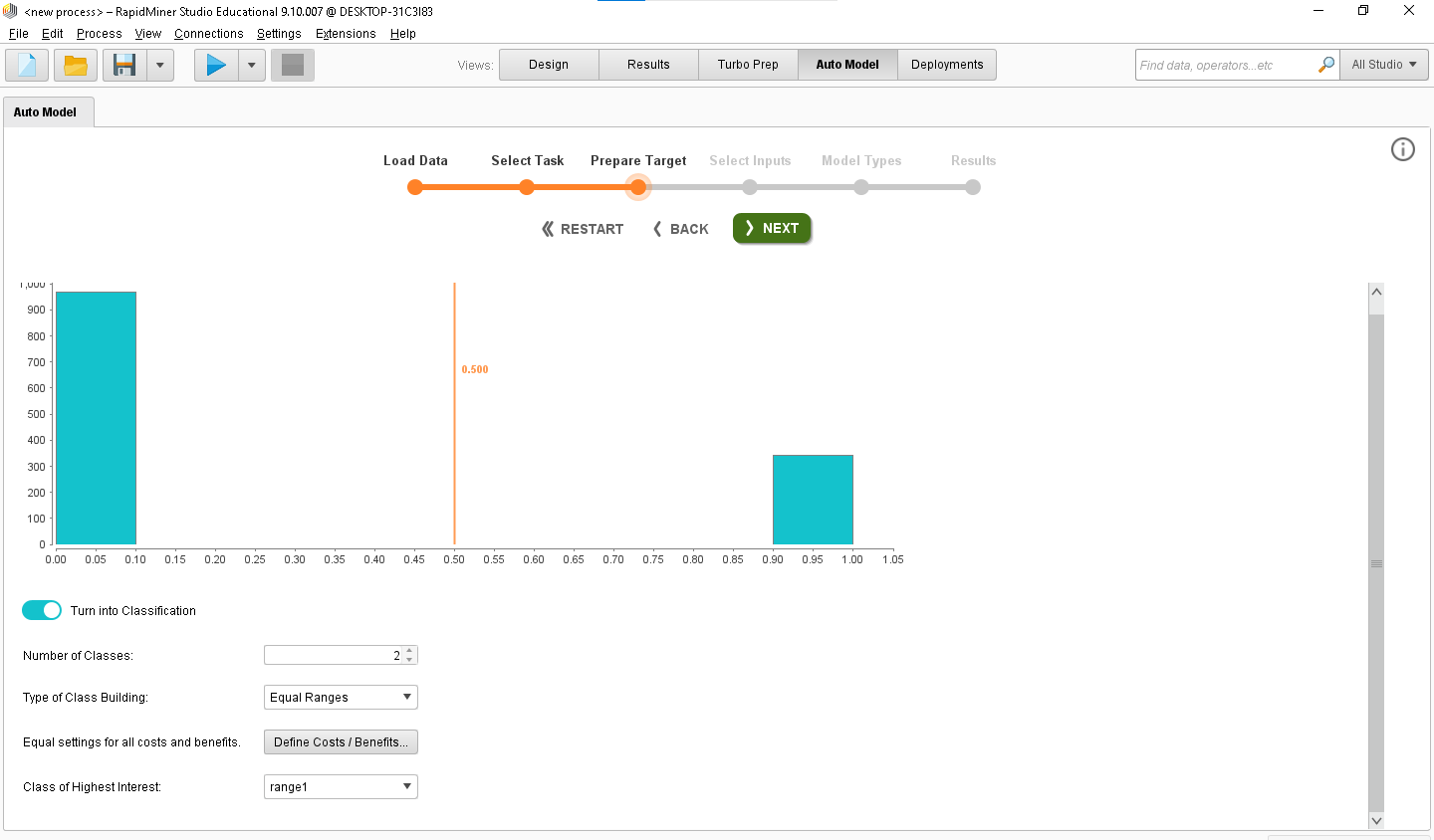


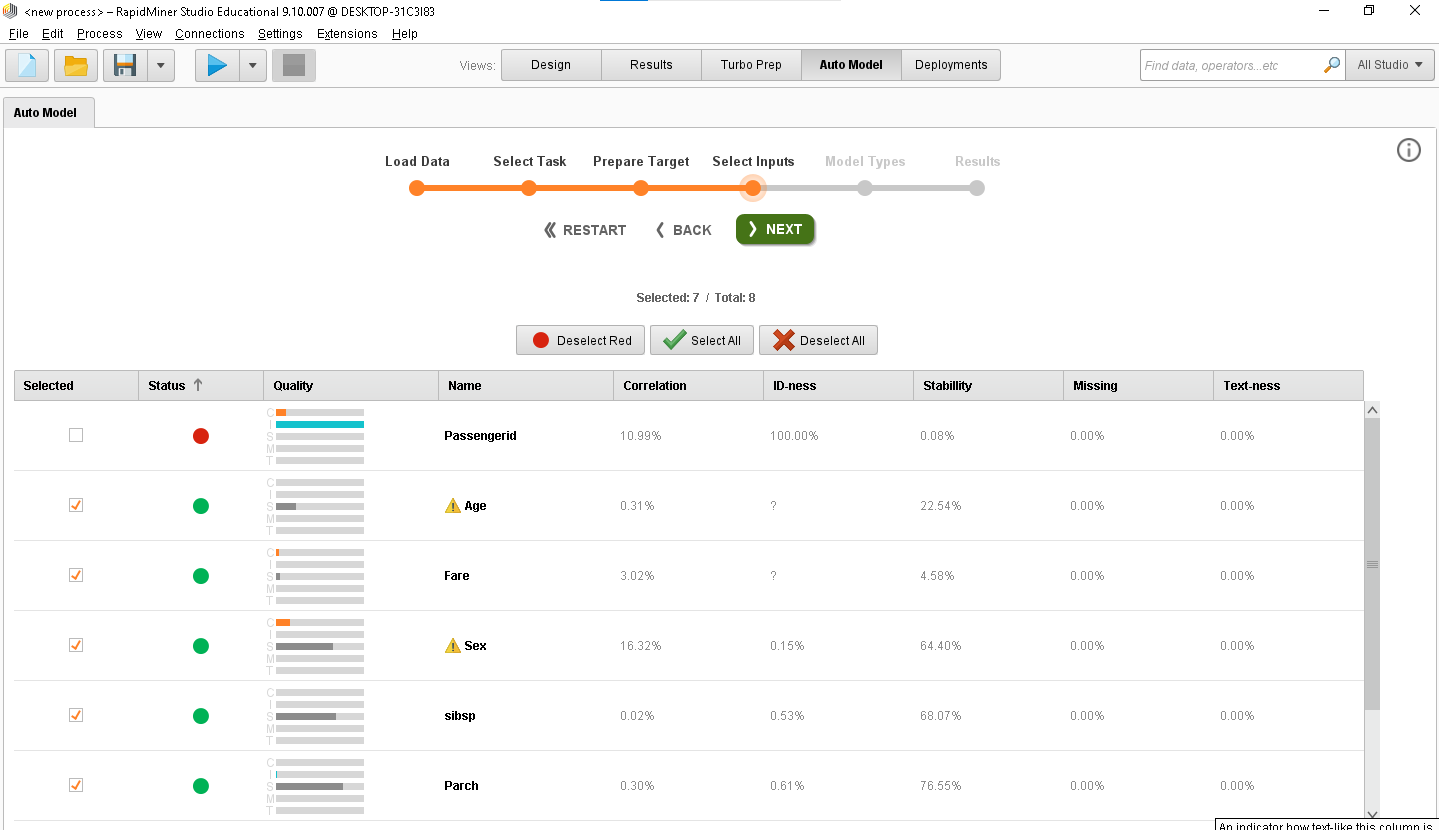


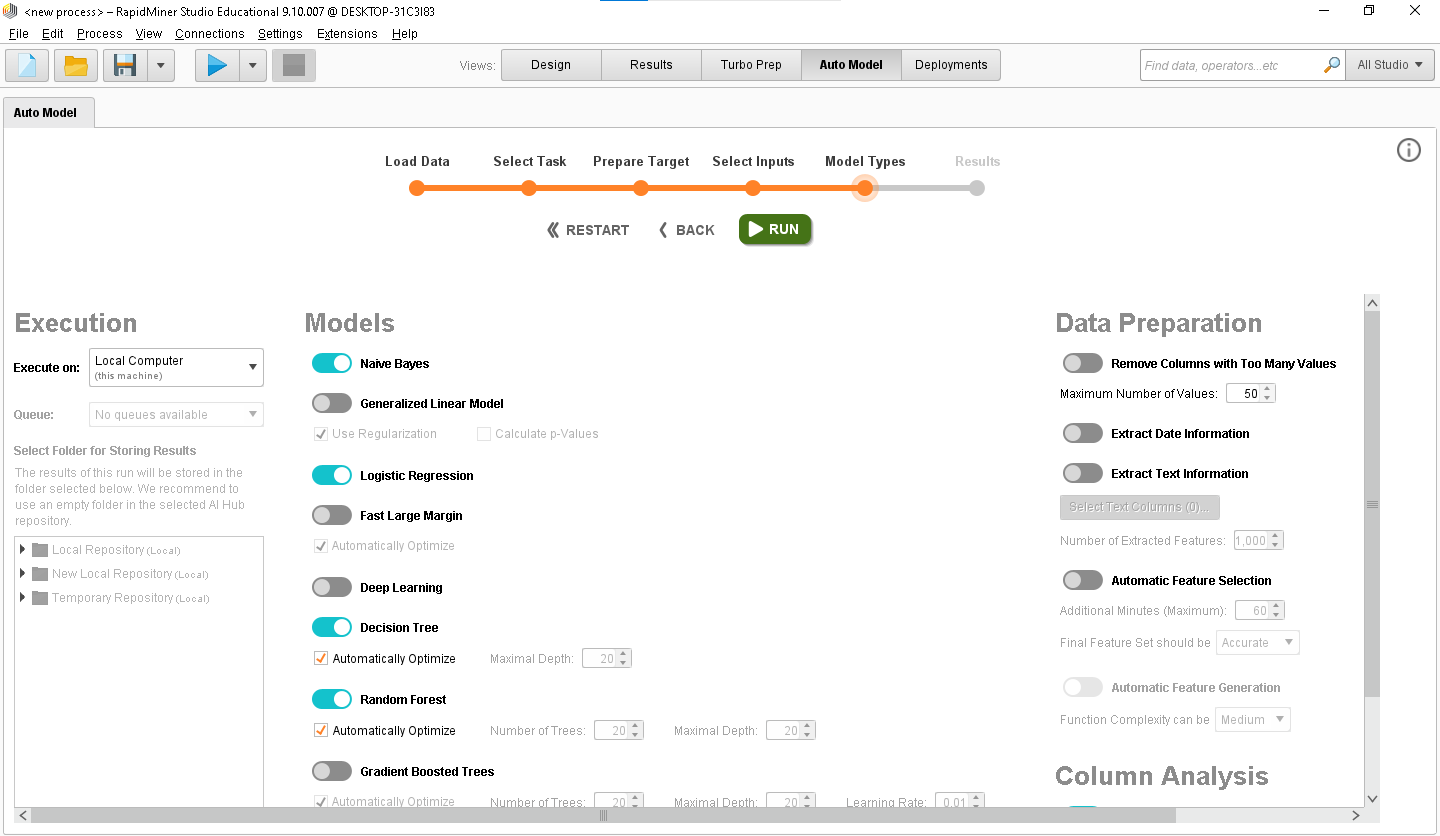


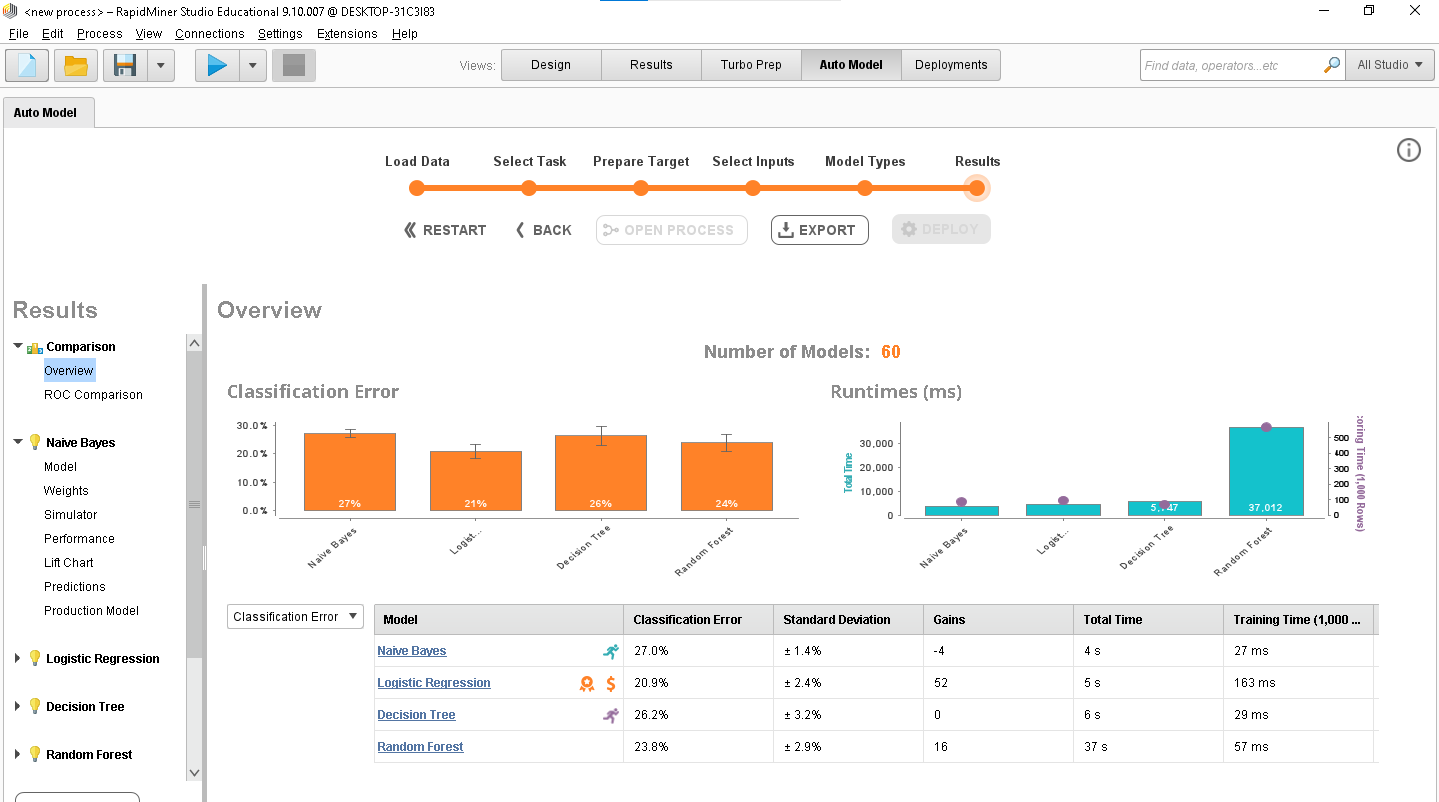


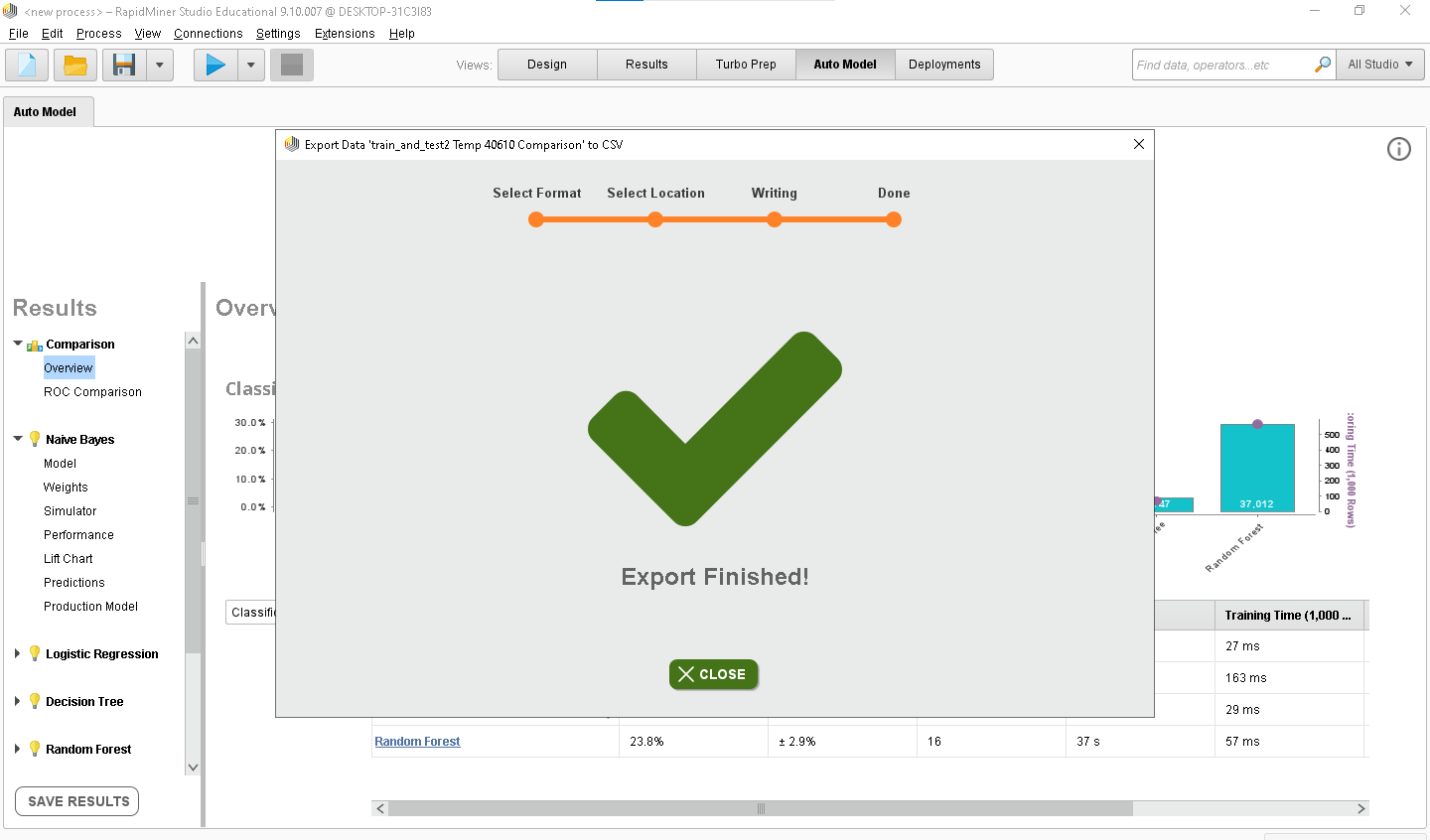












**Questions:**

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

Ans) The five open sources/freeware tools available for data mining are:-

**1. Apache Mahout**

This is a popular distributed linear algebra framework. It is a mathematical expression of Scala DSL, which is designed to let statisticians and data scientists implement their algorithms in a fast manner. The tool supports multiple backends including Apache Spark. Mahout allows apps to analyze large datasets in a faster manner.

**2. ELKI**

ELKI stands for Environment for Developing KDD-Applications Supported by Index-Structures. It is an open-source data mining software written using Java. The platform is designed for algorithm research. ELKI platform aims to research in algorithms, with a focus on unsupervised methods in cluster analysis. ELKI provides data index structures such as R\*-tree. The platform offers an extensive collection of highly parameterizable algorithms.

**3. KNIME**

The Eclipse-based framework is written in Java. It is a multi-language software development environment. KNIME is a free data analytics, reporting, and integration platform. It allows you to choose from 2000 nodes and create visual workflows with an intuitive and user-friendly interface.

**4. Orange**

The open-source, component-based data mining software for machine learning and data visualization is best suited for a range of data visualization, exploration, preprocessing and modelling techniques. Orange has interactive data visualization which can perform simple data analysis.

**5. Rattle**

This open-source GUI for data mining presents statistical and visual summaries of data. Rattle allows for transforming data to allow ready modelling. It also builds unsupervised and supervised ML models from data. The best feature in Rattle is that it provides considerable data mining functionality by exposing the power of R statistical software.



**Outcomes: CO2: Apply concepts of different types of Learning and Neural Networks**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Conclusion:** We carried out the execution of 4 data mining algorithms using RapidMiner. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Autonomous College Affiliated to University of Mumbai)

**Grade: AA / AB / BB / BC / CC / CD /DD** Signature of faculty in-charge with date

KJSCE/IT/LY/SEM-VII/ML/2022-23

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **References:**

Books/ Journals/ Websites:

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



(Autonomous College Affiliated to University of Mumbai)