**––**

**American International University-Bangladesh**

**DATA WAREHOUSING AND DATA MINING**

**Fall 2017-18**

**Project 1: Supervised Learning**

**Name:** Chowdhury, Jahiruddin

**ID:** 15-28397-1

**Section:** A

**Description:** The data set contains details of glass material amounts. The amount of any particular material or element decides the nature and type of the glass, which identifies in which place a particular glass could be used. The study of classification of types of glass was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence...if it is correctly identified!

In this file the fields are separated by comma.

**Number of Instances:** 214

**Number of Attributes:**  12

**Attribute Information:**

Given is the attribute name, attribute type, the measurement unit and a brief description. The nature of glass is identified to be usable as a window glass or not.

ID, N -- numeric identifier of the instance

RI, N -- refractive index

NA2O, N -- Sodium oxide

MGO, N -- magnesium oxide

AL2O3, N -- aluminum oxide

SIO2, N -- silcon oxide

K2O, N -- potassium oxide

CAO, N -- calcium oxide

BAO, N -- barium oxide

FE2O3, N -- iron oxide

TYPE, N -- An unknown, but must correspond to the types in the paper

Class---- Window or Non-window

**Missing Attribute Values:** None

**Class Distribution:**

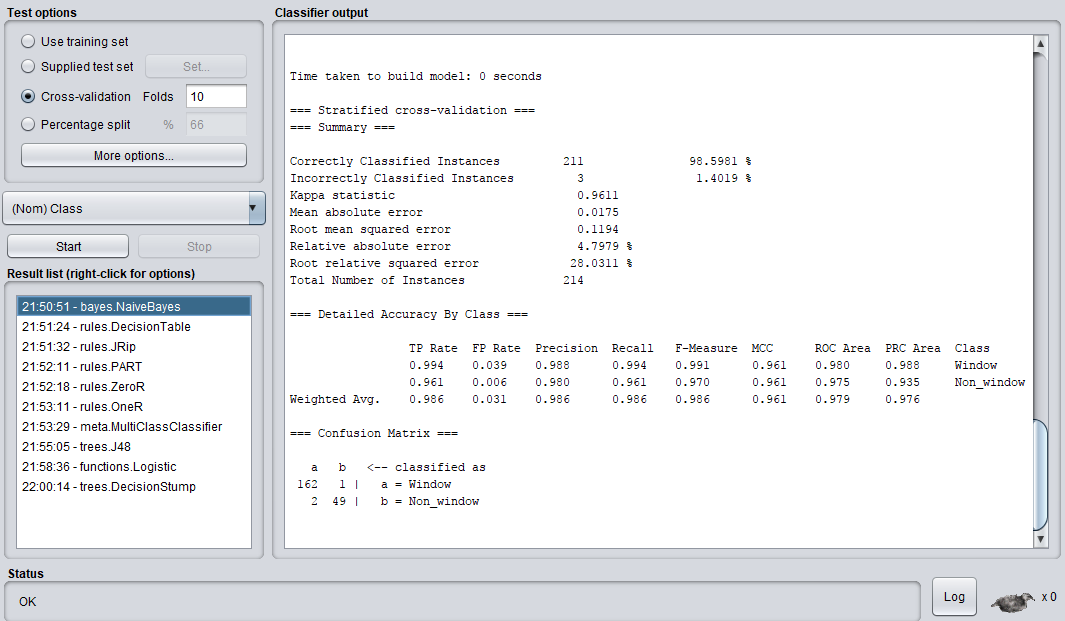
class N N[%]

-----------------------------

Window 163 (76.168%)

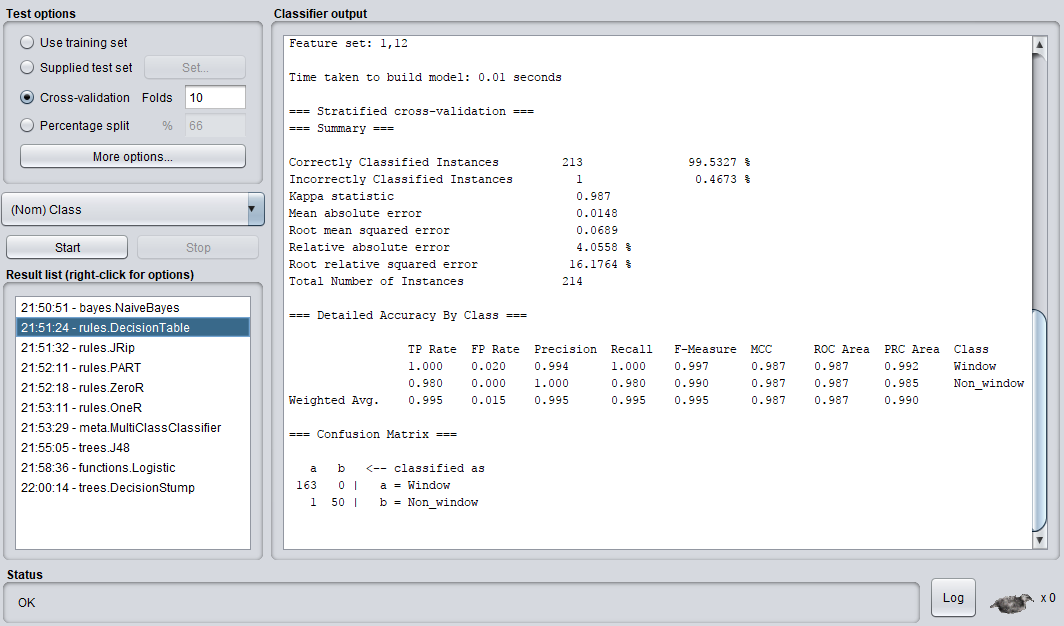
Non-window 51 (23.831%)

**NaiveBayes**



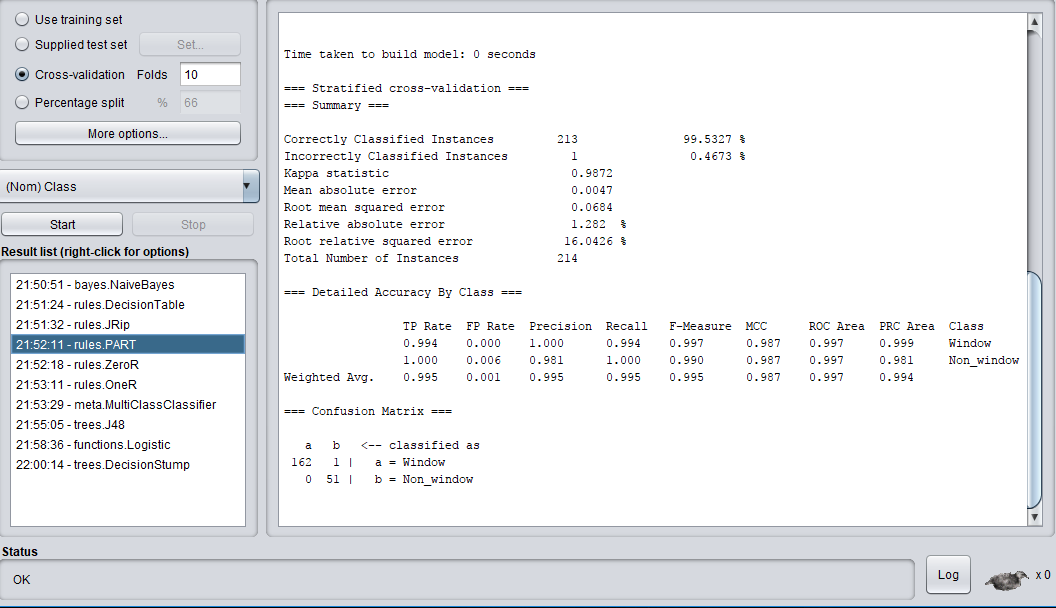
|  |  |  |
| --- | --- | --- |
|  | TPR | FPR |
| Weighted avg. | 0.986 | 0.031 |

**DecisionTable:**



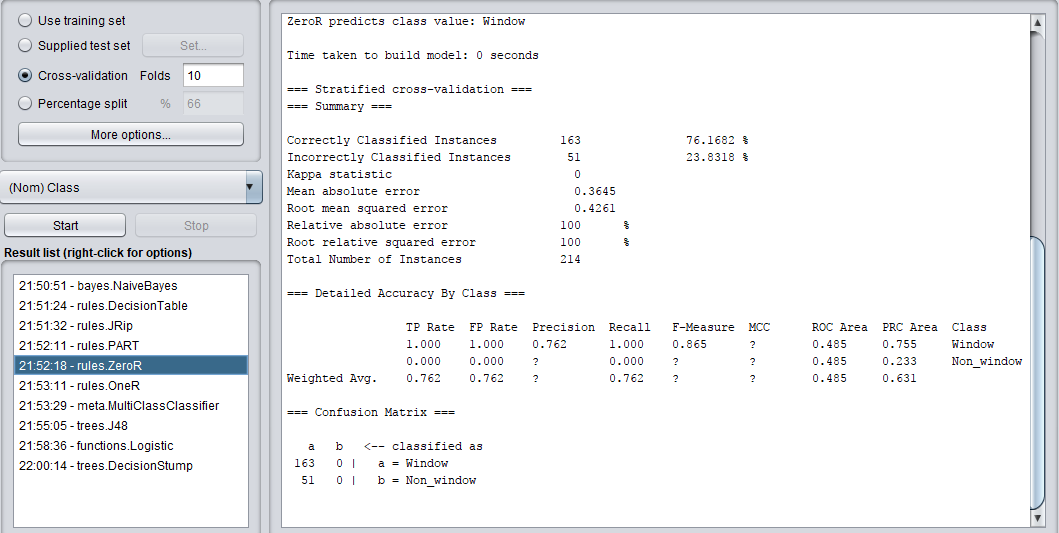
|  |  |  |
| --- | --- | --- |
|  | TPR | FPR |
| Weighted avg. | 0.995 | 0.15 |

**PART**



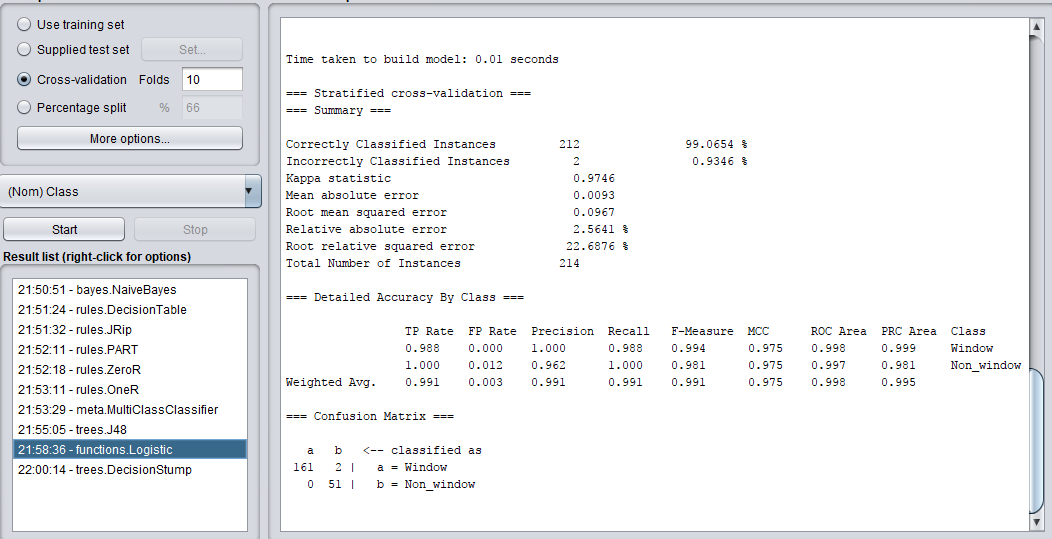
|  |  |  |
| --- | --- | --- |
|  | TPR | FPR |
| Weighted avg. | 0.995 | 0.001 |

**ZeroR**



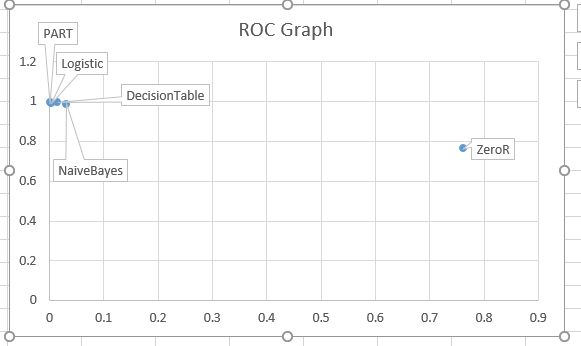
|  |  |  |
| --- | --- | --- |
|  | TPR | FPR |
| Weighted avg. | 0.762 | 0.762 |

**Logistic**



|  |  |  |
| --- | --- | --- |
|  | TPR | FPR |
| Weighted avg. | 0.991 | 0.003 |
|  |  |  |
|  |  |  |

**Receiver Operating Characteristics Graph:**



Comment: According to ROC graph PART algorithm is the best classifier because it has the highest TP rate.