CSC 635 – HW1 (Spring 2018)

Dataset – Report

Rohan Saha (Rohan2728)

February 20th, 2018

**Dataset Title**: Qualitative\_Bankruptcy

**Dataset Link**: <http://archive.ics.uci.edu/ml/datasets/Qualitative_Bankruptcy>

**Total number of Instances**: 250

**Number of Attributes**: 6, each corresponding to Qualitative Parameters in Bankruptcy and a classifier label as Result with two classes C1 = Bankruptcy and C2 = Non-Bankruptcy

**Attribute Information**:

(P - Positive, A - Average, N - Negative, B - Bankruptcy, NB - Non-Bankruptcy)

|  |  |  |
| --- | --- | --- |
| # | Attribute Name | Attribute Values |
| 1 | **Industrial\_Risk** | **P, A, N** |
| 2 | **Management\_Risk** | **P, A, N** |
| 3 | **Financial\_Flexibility** | **P, A, N** |
| 4 | **Credibility** | **P, A, N** |
| 5 | **Competitiveness** | **P, A, N** |
| 6 | **Operating\_Risk** | **P, A, N** |
| 7 | **Result** | **B, NB** |

**Result of Distribution**: [143 instances For Non-Bankruptcy] [107 instances For Bankruptcy]

**Training Data**: Used 70 percentage of total number of instances (175 instances) to build the model.

**Decision Tree**: The decision tree generated by the ID3 algorithm is as follows

**('Competitiveness', {'A': ('Credibility', {'A': 'NB', 'N': ('Financial\_Flexibility', {'A': 'NB', 'N': 'B', 'default': 'B'}), 'P': 'NB', 'default': 'NB'}), 'N': 'B', 'P': 'NB', 'default': 'NB'})**

Used the attribute ‘default’ for handling unknown or missing attribute or attribute vales.

**Test Data**: Used the remaining 30 percentage of total number of instances (75 instances) to test the model. Also tested the model with missing and unknown attributes like

c1 = {**'Industrial\_Risk'**: **'P'**, **'Management\_Risk'**: **'P'**, **'Credibility'**: **'A'**, **'Competitiveness'**: **'A'**, **'Operating\_Risk'**: **'P'**}  
c2 = {**'Industrial\_Risk'**: **'P'**, **'Management\_Risk'**: **'P'**, **'Financial\_Flexibility'**: **'A'**, **'Credibility'**: **'A'**, **'Operating\_Risk'**: **'P'**}  
c3 = {**'Competitiveness'**: **'A'**,**'Financial\_Flexibility'**: **'D'**}  
c4 = {**'Competitiveness'**: **'A'**}  
c5 = {**'Credibility'**: **'N'**}

**Result**: The model generated by the ID3 algorithm was able to classify the test data successfully and was also able to handle any missing and unknown attributes.