File Needed (use uk eng)

1. Flowcharts and concise description of functions description
2. Design Document/Report – heading, consistent alignment, clear description, illustration used

Overall Software Objectives (requirement)

1. C, C++, C#, Java (or python discussed verbally in the class)
2. Include two modules, C4.5 algorithm and naïve Bayesian
3. Build classification model based on training set (defined in text file)
4. Evaluated by testing (defined in text file)
5. Display classification result of each testing data
6. display metrics for evaluating performance of each trained classifier

Deliverables

1. exe
2. source code
3. deployment and installation manual, user manual
4. .txt file of training set
5. .txt file of testing set
6. Technical report in .pdf
7. Presentation video + demo video

Data Structure

1. Require header
2. Use ‘;’ semicolon to separate data

Input Requirement

1. Allow user enter a file name for training set
2. Allow user to enter file name for testing set after training

Module 1: C4.5 Classification Module

1. Program will develop a tree model using C4.5
2. Illustrate each stages of developing decision tree model
3. Display measure for every iteration in building decision tree
4. Measure for C4.5
   1. Info(D)
   2. InfoA(D)
   3. Gain(A)
   4. SplitInfoA(D)
   5. GainRatio(A)
5. Illustrate resulting tree

Module 2: Naïve Bayesian

1. Notation
   1. Ci = category, xi attribute value
   2. P(Ci), the prior probability of each of the categories
   3. P(xi | Ci), the posterior probability of each of the attribute values conditioned on each of the categories.
2. Illustrate the computational process in classifying data tuples
3. Include functionality for Laplacian correction to avoid any computing probability values of zero

Evaluation Module

1. Accuracy
2. Error Rate
3. Sensitivity (Recall)
4. Specificity
5. Precision
6. F and F1, Harmonic Mean of Precision and Recall