

**ISM 6136 – Datamining/Predictive Analytics**

**Class Assignment 4**

**5 points**

**TASK: Performing classification using Naïve Bayes Classifier in XLMiner**

**Part 1**

1. Dataset provided has test results on the ﬂying ﬁtness tests for 40 pilots. There are six other categorical variables (var2 - var 6) indicative of the

performance of the pilots on various physical and psychological tests. Predict the outcome class of the pilots in the ‘new data’ sheet of the dataset using the Naïve Bayes Classifier.

1. Follow the datamining steps below:
2. Understand the problem and purpose of data mining task
3. Obtain the dataset for analysis
4. Explore, clean and preprocess data
   1. Cleanup any column that is not a predictor
   2. Perform ‘Missing Data Handling’ *(Delete any record that is corrupt)*
5. Reduction of data dimension (if needed)
6. Partition data
7. Choose the data mining techniques/algorithms – Classify, Naïve Bayes and build **at least 2 models**
8. Interpret the results and depending on the model selection criteria choose the **best model.** *Check the full criteria mentioned on the lecture slide.*
9. Deploy **best model** on the new data and paste a screen shot of your prediction results
10. Submit the **Excel workbook and this word document with explanation for steps c) through h).**

**Part 2**

1. Dataset provided has been taken from Zoo database that was created by Richard Forsyth in 1990 to illustrate its PC-Beagle program (rule finding program). It contains a list of animals in rows and their associated attributes described in 17 distinct qualitative variables (columns): hair, feathers, eggs, milk, airborne, aquatic, predator, toothed, backbone, breathes, venomous, fins, legs, tail, domestic, catsize. Based on these qualitative variables the animals are classified into outcome class type 1 through 7. Predict the class type of the animals in the ‘new data’ sheet using Naïve Bayes Classifier.
2. Follow the datamining steps below:
3. Understand the problem and purpose of data mining task
4. Obtain the dataset for analysis
5. Explore, clean and preprocess data
   1. Cleanup any column that is not a predictor
   2. Perform ‘Missing Data Handling’ *(Delete any record that is corrupt)*
6. Reduction of data dimension (if needed)
7. Partition data
8. Choose the data mining techniques/algorithms – Classify, Naïve Bayes and build **at least 2 models**
9. Interpret the results and depending on the model selection criteria choose the **best model.** *Estimate the Precision and Sensitivity by looking at the diagonal of the matrix and pick the model with higher values. Check the full criteria mentioned on the lecture slide.*
10. Deploy **best model** on the new data and present your prediction results
11. Submit the **Excel workbook and this word document with explanation for steps c) through h).**