**CIDM 6355 Data Mining Methods Exam 1 Part 2 Submission**

**(50 points in total; due 11:59 pm CDT, October 8th, 2023)**

This exam is open book, open slides, and open notes, but you are not allowed to collaborate nor discuss with anyone else. Sharing your screenshots, RM processes, R script, or answers with other students are considered as cheating for this exam. Should you have any question or unclarity, please contact with the instructor. Please put all your deliverables in a word document and submit it to WT class before the deadline; Make sure that all your screenshots include dates and time [see the examples as below]; otherwise, a penalty of 50% of your grade will be applied. Please type your name below to indicate whether or not you have understood and complied with such requirements in this exam.

Your name (First Last): Trevor Hofmann

**PLEASE MAKE SURE THAT YOU TYPE YOUR NAME AT THE FIRST PAGE; OTHERWISE, YOUR SUBMISSIN WILL NOT BE GRADED AND A ZERO POINT WILL BE ASSIGNED.**

2 Screenshots in RM (6 points for each: 3 pts for your screenshot and 3 pts for your description/discussion).

Screenshot 1 with description (6 pts): A screenshot of your decision tree graph with date and time at Step 2.3 and briefly describe your model. Your description must include root node, split nodes, and leaf nodes.

A screenshot of a computer

Description automatically generated

The root node is Open Price. There are 6 splite nodes including the Root node. There are 7 leaf nodes.

Screenshot 2 with your discussion (6 pts): A screenshot of your RapidMiner Process (the flow chart in your design mode) with date and time and briefly discuss why the operator Nominal to Numerical must be used in your process.

A screenshot of a computer

Description automatically generated

The Neural Net is not able to use text-based data to model with. It relies on numbers to formulate a module. Thus, we need to convert the categorical variables into dummy variables. This will inflate the number of columns in the dataset but will resolve in a working answer.

Screenshot 3: A screenshot of your R codes with date and time to show how you import and prepare the data for modeling and prediction, that is, Steps 6.1-6.3.

A screenshot of a computer

Description automatically generated

Screenshot 4: A screenshot of your R codes with date and time to show Step 6.4.1-6.4.3. Requirements: your screenshot must clearly include all the R codes for your decision tree model and the output of 6.4.3.

A screenshot of a computer

Description automatically generated

Screenshot 5: A screenshot of your R codes with date and time to show Step 6.5.1-6.5.3. Requirements: your screenshot must clearly include all the R codes for your NB model and the output of 6.5.3.

A screenshot of a computer

Description automatically generated

Screenshot 6: A screenshot of your R codes with date and time to show Step 6.6.1-6.6.4. Requirements: your screenshot must clearly include all the R codes for your logistic regression model and the output of 6.6.4.

A screenshot of a computer

Description automatically generated

Screenshot 7: A screenshot of your R codes with date and time to show Step 6.7.1-6.7.4. Requirements: your screenshot must clearly include all the R codes for your NN model and the output of 6.7.4.

A screenshot of a computer

Description automatically generated

Step 7: Comparative Analysis (18 points)

7.3. Please include the following deliverables in your submission:

7.3.1. Please copy and paste your table here. Your table must include the predicted results of 20 records with 8 methods (8 pts in total and each column is worth 1 point).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | RM\_DT | RM\_NB | RM\_LR | RM\_NN | R\_DT | R\_NB | R\_LR | R\_NN |
| 1 | no | no | no | no | no | no | no | no |
| 2 | yes | yes | yes | yes | yes | yes | yes | no |
| 3 | no | yes | yes | yes | yes | yes | yes | yes |
| 4 | yes | yes | yes | yes | yes | yes | yes | yes |
| 5 | yes | no | no | yes | no | no | no | no |
| 6 | no | yes | yes | no | yes | yes | yes | yes |
| 7 | no | no | no | no | no | no | no | no |
| 8 | yes | no | no | yes | yes | no | no | yes |
| 9 | no | no | no | yes | no | no | no | no |
| 10 | yes | no | no | yes | yes | no | no | yes |
| 11 | no | yes | yes | yes | yes | yes | yes | yes |
| 12 | yes | yes | yes | yes | no | yes | yes | no |
| 13 | yes | no | no | no | no | no | no | yes |
| 14 | no | no | no | yes | no | no | no | yes |
| 15 | yes | yes | yes | yes | yes | yes | yes | yes |
| 16 | yes | yes | yes | yes | yes | yes | yes | no |
| 17 | no | no | no | no | no | no | no | no |
| 18 | yes | yes | yes | yes | no | yes | yes | yes |
| 19 | no | yes | yes | yes | yes | yes | yes | yes |
| 20 | no | no | no | yes | no | no | no | yes |

7.3.2. Discuss how many records are predicted to be yes by each method in RM and R. For example, among the 20 records, both RM\_DT and R\_DT predict "yes" for 10 records, while they jointly predict "no" for 5 records (8 pts: 2 pts for each pair of methods)

RM\_DT, RM\_NB, RM\_LR, R\_DT, R\_NB, R\_LR each have 10 yes and 10 no.

RM\_NN had 15 yes and 5 no. R\_NN had 12 Yes and 8 No

7.3.3. At the end, please provide an analysis of the number of records that all eight models predict as "yes" and the number of records that all eight models predict as "no." (2 pts).

ID 3,4,11,15 are all yes. ID 1,7,17 are all No