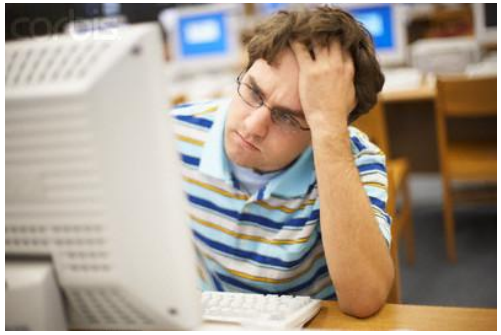
UNC CHARLOTTE

How can we use data to increase retention and quality of student work in freshman computer science courses?



Introduction to Computer Science Course has high rate of Withdraws, D grade, F grade (WDF)

This prevents as many as 30% of freshman from going on to the next course in the major.

Early intervention and making students aware of what they understand of what actions he/she can take may lead to a better chance of passing the course

Data was collected from a single lecture section over course of semester.

PRS “Clicker” quizzes- student answers to multiple choice questions
Attendance in lectures
Participation in a study session
Score on first lecture test
Correctness of each test question for each student

Ran J48 to predict test 1 grade using 7 other attributes:

15:37:12 – trees.J48

```
               ITCS1212Spring2010 (2/15/10 8:47:29 AM)
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

PRS Total <= 17
|   ITCS1212Spring2010 (2/3/10 8:40:05 AM) <= 3
|   |   ITCS1212Spring2010 (2/15/10 8:47:29 AM) <= 1: F (40.63/24.0)
|   |   ITCS1212Spring2010 (2/15/10 8:47:29 AM) > 1
|   |   |   Absences before test 1 <= 0: C (6.3/2.3)
|   |   |   Absences before test 1 > 0
|   |   |   |   ITCS1212Spring2010 (2/8/10 8:51:04 AM) <= 0
|   |   |   |   |   ITCS1212Spring2010 (2/3/10 8:40:05 AM) <= 0: C (6.0/2.0)
|   |   |   |   |   ITCS1212Spring2010 (2/3/10 8:40:05 AM) > 0: F (5.0/1.0)
|   |   |   |   ITCS1212Spring2010 (2/8/10 8:51:04 AM) > 0: C (2.0)
|   ITCS1212Spring2010 (2/3/10 8:40:05 AM) > 3: C (4.06/2.06)
PRS Total > 17
|   ITCS1212Spring2010 (2/3/10 8:40:05 AM) <= 3
|   |   Study Session = y
|   |   |   ITCS1212Spring2010 (2/8/10 8:51:04 AM) <= 0: B (12.0/5.0)
|   |   |   ITCS1212Spring2010 (2/8/10 8:51:04 AM) > 0: A (2.0/1.0)
|   |   Study Session = n
|   |   |   ITCS1212Spring2010 (2/15/10 8:47:29 AM) <= 3
|   |   |   |   ITCS1212Spring2010 (2/8/10 8:51:04 AM) <= 0
|   |   |   |   |   ITCS1212Spring2010 (2/15/10 8:47:29 AM) <= 2
|   |   |   |   |   |   Absences before test 1 <= 3
|   |   |   |   |   |   |   ITCS1212Spring2010 (2/3/10 8:40:05 AM) <= 2
|   |   |   |   |   |   |   |   ITCS1212Spring2010 (2/3/10 8:40:05 AM) <= 0: D (2.0)
|   |   |   |   |   |   |   |   ITCS1212Spring2010 (2/3/10 8:40:05 AM) > 0: B (2.0)
|   |   |   |   |   |   |   ITCS1212Spring2010 (2/3/10 8:40:05 AM) > 2: C (4.0/1.0)
|   |   |   |   |   |   Absences before test 1 > 3: F (2.0/1.0)
|   |   |   |   |   ITCS1212Spring2010 (2/15/10 8:47:29 AM) > 2: C (10.0/5.0)
|   |   |   |   ITCS1212Spring2010 (2/8/10 8:51:04 AM) > 0: D (5.0/1.0)
|   |   |   ITCS1212Spring2010 (2/15/10 8:47:29 AM) > 3
|   |   |   |   ITCS1212Spring2010 (2/3/10 8:40:05 AM) <= 1
|   |   |   |   |   Absences before test 1 <= 2: B (4.0/1.0)
|   |   |   |   |   Absences before test 1 > 2: C (2.0/1.0)
|   |   |   |   ITCS1212Spring2010 (2/3/10 8:40:05 AM) > 1
|   |   |   |   |   ITCS1212Spring2010 (1/27/10 8:25:45 AM) <= 4: D (7.0/3.0)
|   |   |   |   |   ITCS1212Spring2010 (1/27/10 8:25:45 AM) > 4: A (3.0/1.0)
|   ITCS1212Spring2010 (2/3/10 8:40:05 AM) > 3
|   |   Absences before test 1 <= 0: A (8.0/2.0)
|   |   Absences before test 1 > 0: C (4.0/1.0)
```

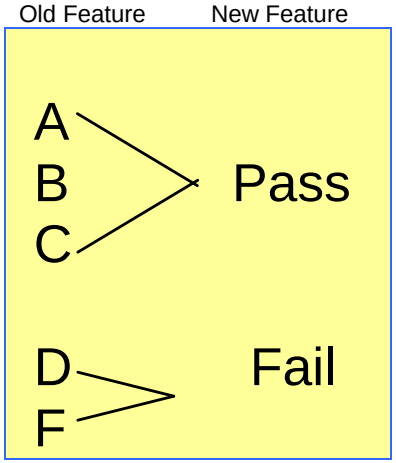
Accuracy: 29%
Classification Error: 71%
Kappa: .09
Root Mean Squared Error: .42

A	F	C	D	U	B
5	2	1	4	0	2
1	17	9	1	0	3
4	10	11	6	0	6
3	8	5	3	0	3
0	7	0	1	0	0
4	3	8	2	0	2

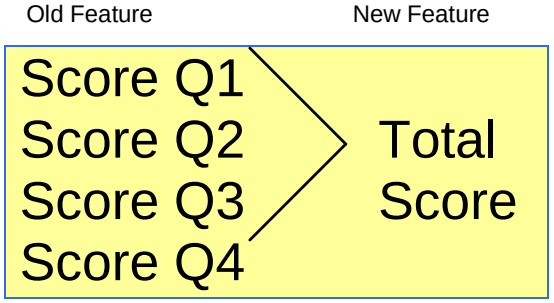
Confusion Matrix

Can re-forming the question and feature consolidation/re-selection give more meaningful results?

Consolidate Test Scores



Consolidate Clicker Scores

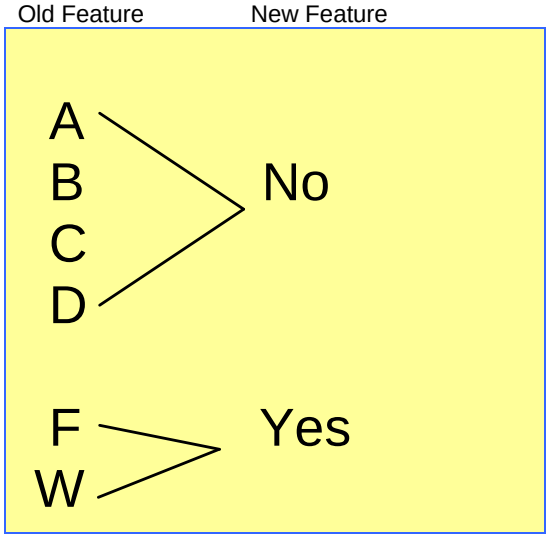


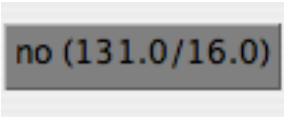
Ran J48 to predict WF from two new features and attendance.

More meaningful results?

not quite!
not yet…

Consolidate Course Grade





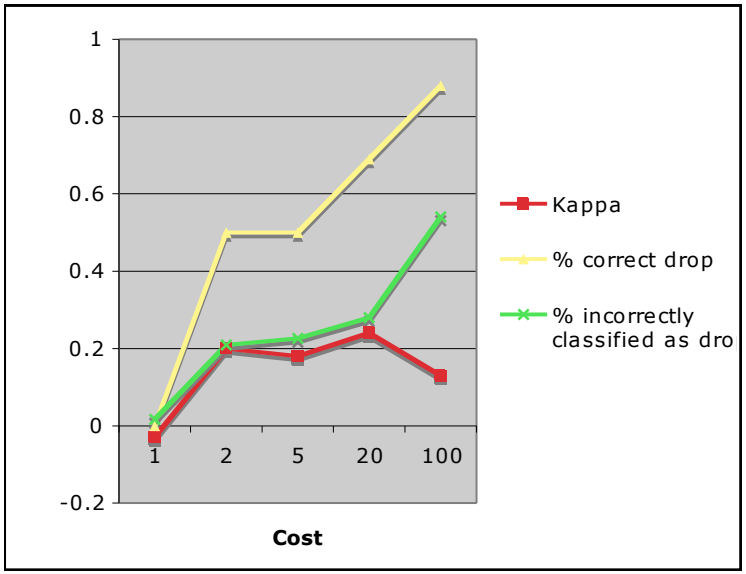
decision tree

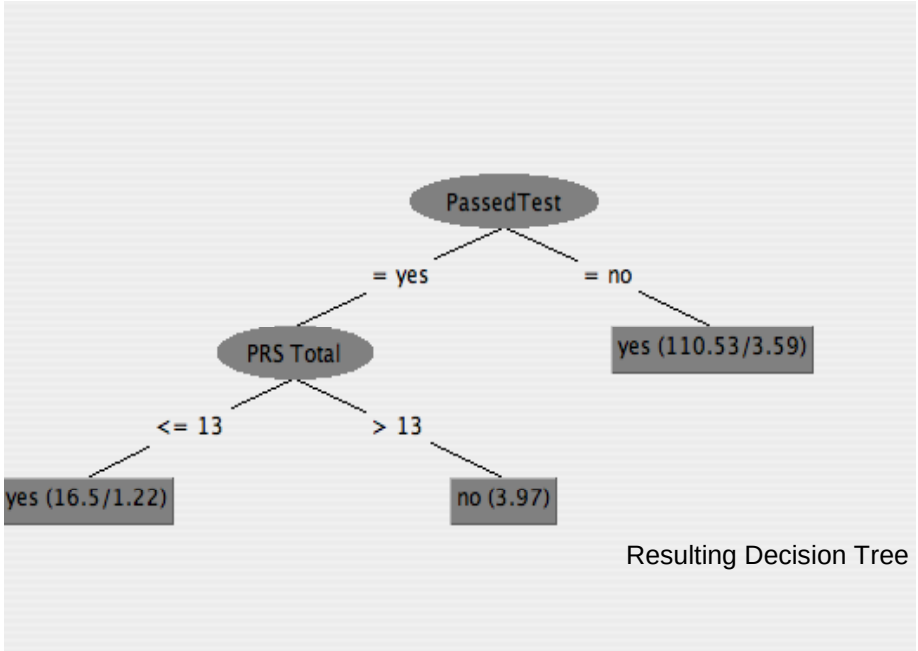
Because the number of WF in this data was small compared to the total number of rows, the algorithm made the decision on one feature alone.

Since the driving issue is to identify students at risk of being a “WF”, it is important to predict the true drops as correctly as possible.

A cost sensitive classifier was applied to the J48 algorithm.
A high cost was placed on incorrectly predicting students as passing.

The purpose is to intervene early to prevent failure. Identifying these students is thought to be more important than the cost of unnecessarily intervening with students incorrectly predicted as failing.





Resulting Decision Tree