Florida Atlantic University (FAU)

Assignment 4

Project objective: **Module Order Modeling (MOM) -** Ordering of software modules, based on the number of faults predicted by a software quality prediction model.

Written By: Kevin Tudor

Table of Contents

Requirements

- This assignment involves ordering of software modules, based on the number of faults predicted by a software quality prediction model**. Use the predictions obtained with Linear Regression Models we built in Homework I-B**. You will use only two of the three models you obtained in Homework I-B. The models to be used for this assignment are as follows:
    - Linear Regression Model with M5 Method of Attribute Selection:
        - FAULTS = - 0.0516 * NUMUORS + 0.0341 * NUMUANDS - 0.0027 * TOTOTORS - 0.0372 * VG + 0.2119 * NLOGIC + 0.0018 * LOC + 0.005 * ELOC - 0.3091
    - Linear Regression Model with Greedy Method of Attribute Selection:
        - FAULTS = - 0.0482 * NUMUORS + 0.0336 * NUMUANDS - 0.0021 * TOTOTORS - 0.0337 * VG + 0.2088 * NLOGIC + 0.0019 * LOC - 0.3255

- Obtain the predictions for both the fit data set and the test data set using the above two models. Perform Module Order Modeling for both fit and test data sets using both regression models.

- Compare the performances of MOM for both linear regression models. Use Alberg Diagram and Performance Curve for each Model using fit and test data sets.

- Use tables to summarize the results of MOM. Also provide analysis of your summary.

  NOTE:
    - You should NOT start working on this assignment until you have read Reference #5 and #9.
    - You DO NOT need to use Weka or any other software tool to do this homework. Any spreadsheet program like Microsoft Excel will suffice.

Data Predictions

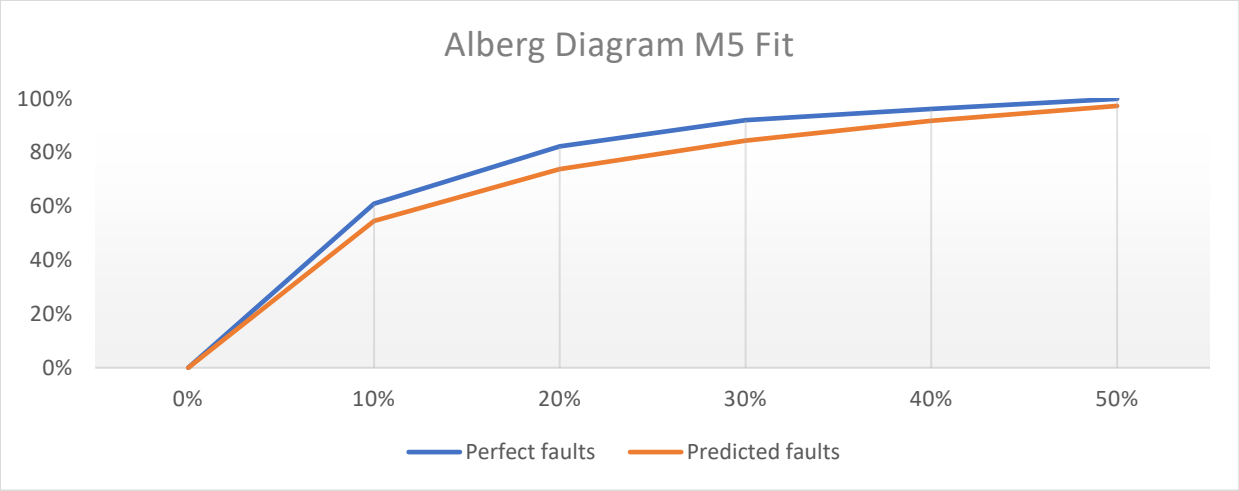
Obtain the predictions for both the fit data set and the test data set using the above two models. Perform Module Order Modeling for both fit and test data sets using both regression models.

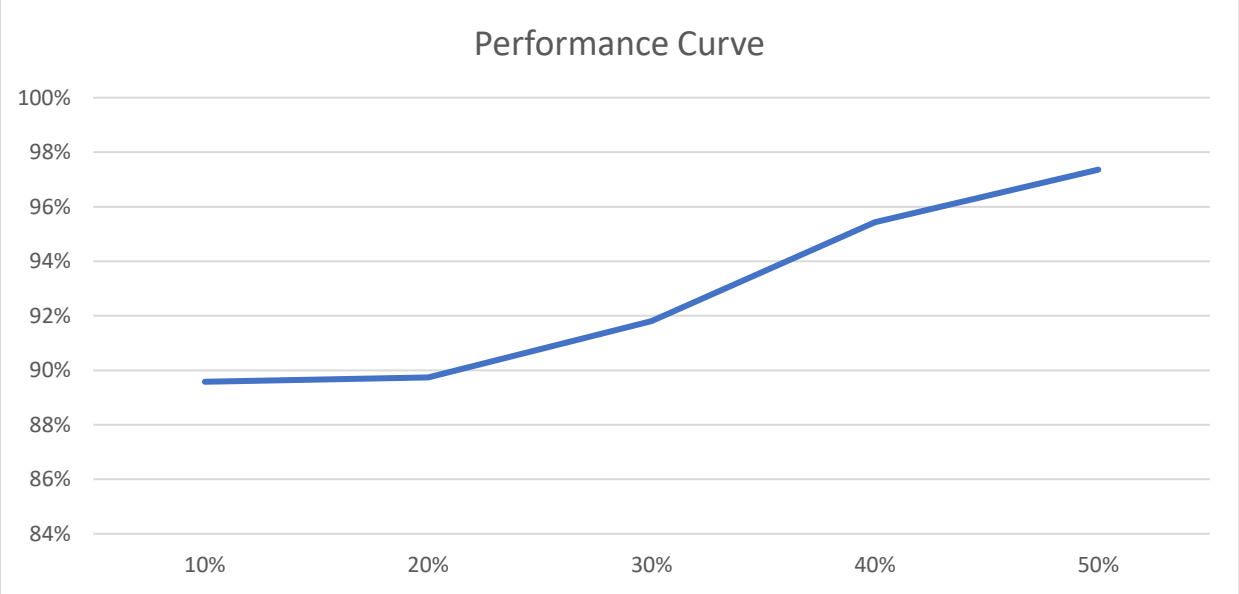
All information for the predicted values is in attached excel file. Not feasible to include 190 + lines of predicted values in a word document. Data predicted is calculated with the Linear Regression Models built in HW I-B then categorized by highest faults to lowest faults and MOM.

Compare Alberg and Performance Curves

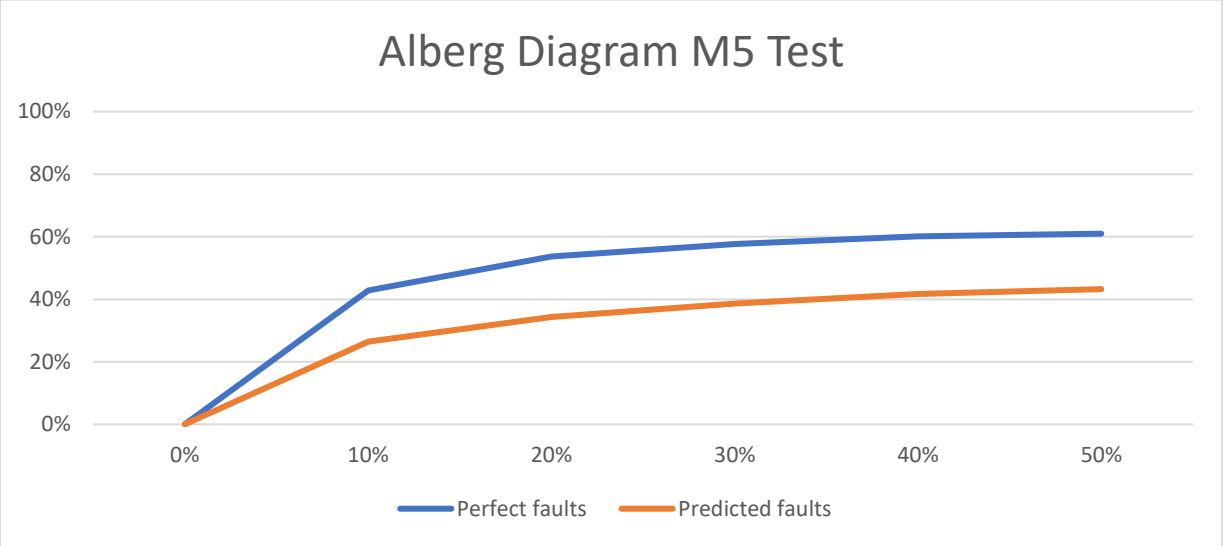
Compare the performances of MOM for both linear regression models. Use Alberg Diagram and Performance Curve for each Model using fit and test data sets.

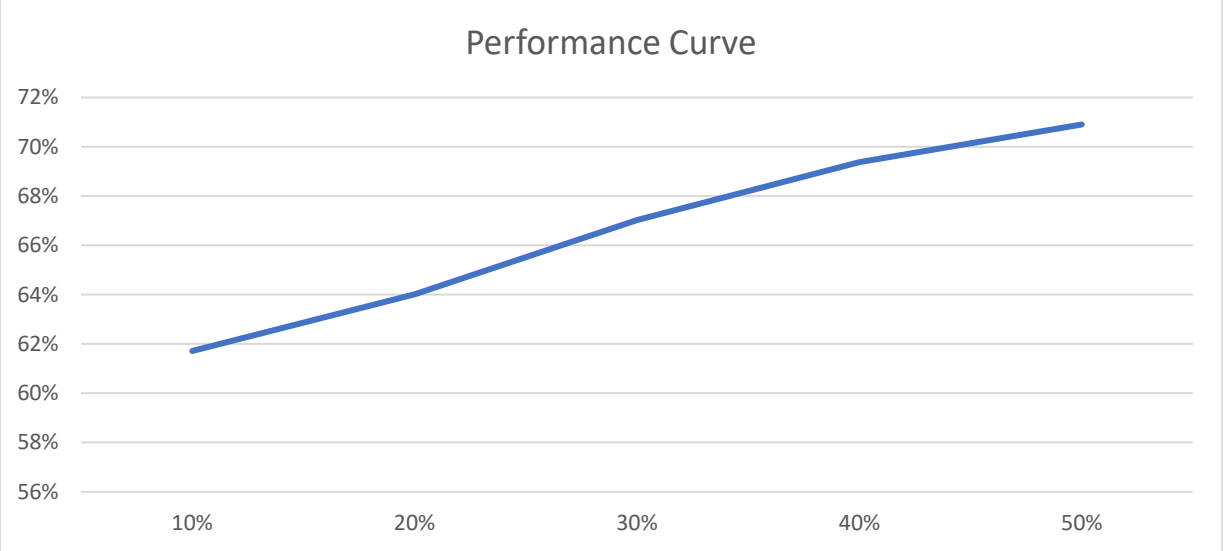
M5 Fit



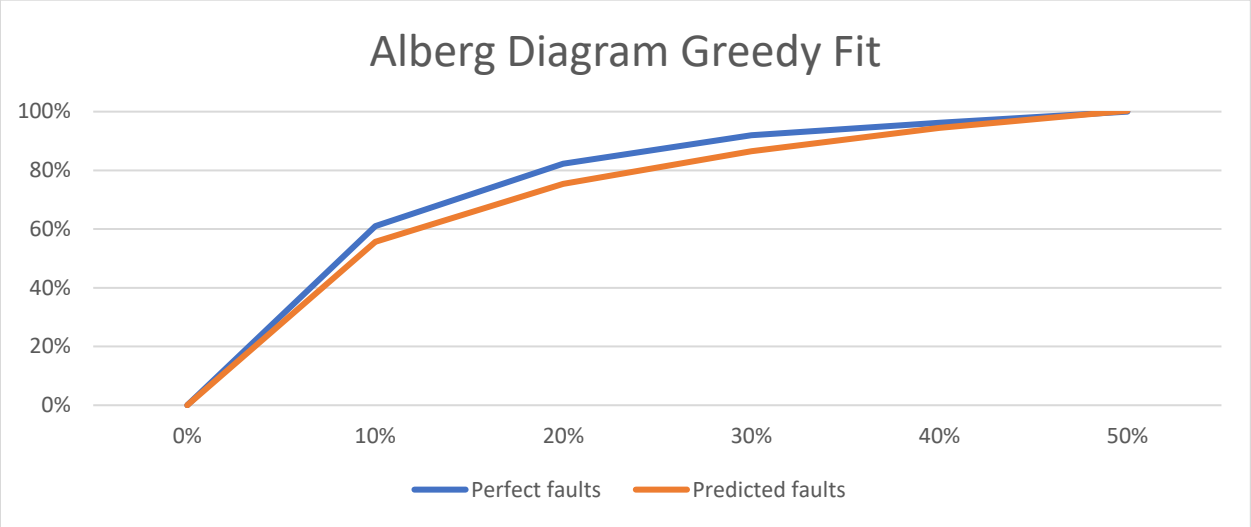


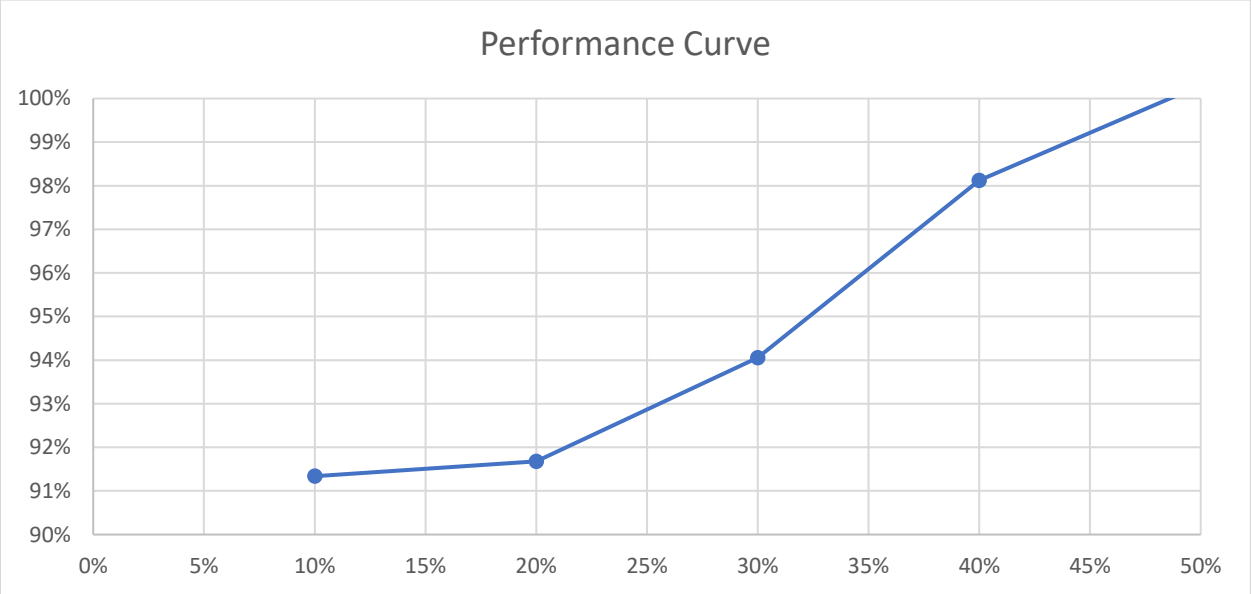
M5 Test


Alberg Diagram M5 Test
100%
80%
60%
40%
20%
0%
0%
10%
20%
30%
40%
50%
Perfect faults
Predicted faults


Performance Curve
72%
70%
68%
66%
64%
62%
60%
58%
56%
10%
20%
30%
40%
50%

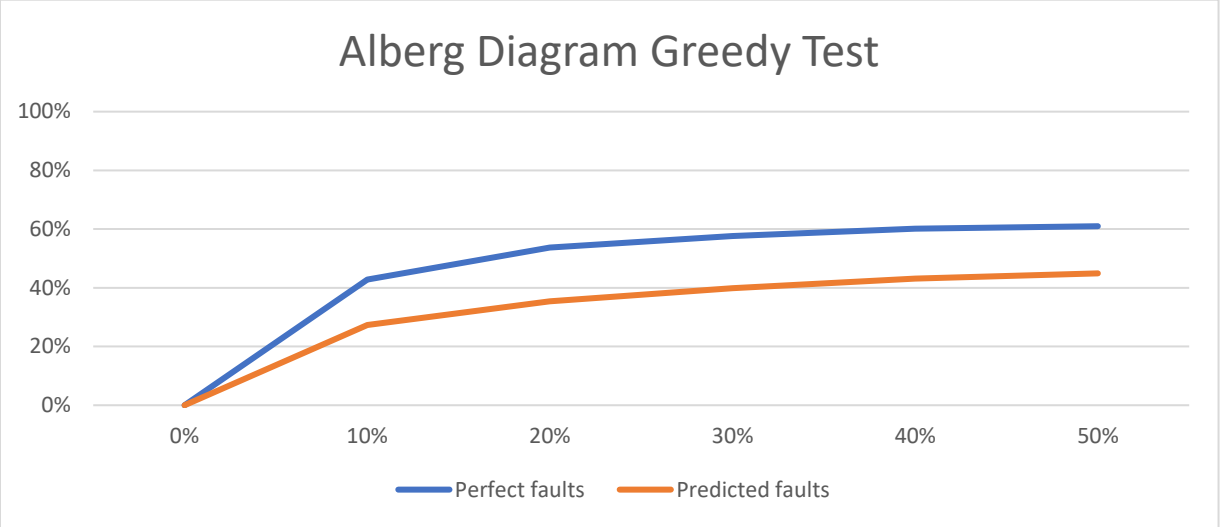
Greedy Fit

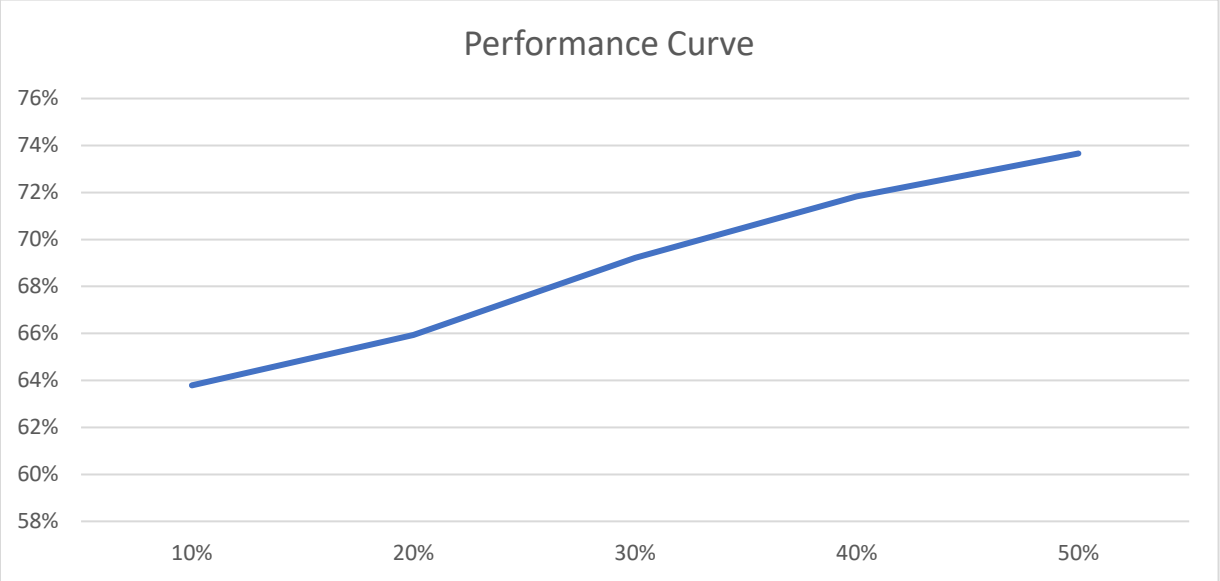




Note: At 50% the performance is at 100%.

Greedy Test


Alberg Diagram Greedy Test
100%
80%
60%
40%
20%
0%
0%
10%
20%
30%
40%
50%
Perfect faults
Predicted faults


Performance Curve
76%
74%
72%
70%
68%
66%
64%
62%
60%
58%
10%
20%
30%
40%
50%

MOM Results

Use tables to summarize the results of MOM.

M5 Fit

MOM Cutoff Percentile	Perfect Ranking # Faults	Predicted Ranking # Faults
10%	275	246.3382
20%	96	86.5869
30%	44	48.034
40%	19	33.2232
50%	17	24.914

Percentages		
Cutoff Percentile	Perfect faults	Predicted faults
0%	0%	0%
10%	61%	55%
20%	82%	74%
30%	92%	84%
40%	96%	92%
50%	100%	97%

Performance	
Cutoff Percentile	Perfect faults/Predicted faults
10%	90%
20%	90%
30%	92%
40%	95%
50%	97%

M5 Test

MOM Cutoff Percentile	Perfect Ranking # Faults	Predicted Ranking # Faults
10%	193	119.1118
20%	49	35.7944
30%	18	19.3691
40%	11	13.7411
50%	4	6.9755

Percentages		
Cutoff Percentile	Perfect faults	Predicted faults
0%	0%	0%
10%	43%	26%
20%	54%	34%
30%	58%	39%
40%	60%	42%
50%	61%	43%

Performance	
Cutoff Percentile	Perfect faults/Predicted faults
10%	62%
20%	64%
30%	67%
40%	69%
50%	71%

Greedy Fit

MOM Cutoff Percentile	Perfect Ranking # Faults	Predicted Ranking # Faults
10%	275	251.1796
20%	96	88.9358
30%	44	50.2208
40%	19	35.5285
50%	17	26.5117

Percentages		
Cutoff Percentile	Perfect faults	Predicted faults
0%	0%	0%
10%	61%	56%
20%	82%	75%
30%	92%	87%
40%	96%	94%
50%	100%	100%

Performance	
Cutoff Percentile	Perfect faults/Predicted faults
10%	91%
20%	92%
30%	94%
40%	98%
50%	100%

Greedy Test

MOM Cutoff Percent	Perfect Ranking # Faults	Predicted Ranking # Faults
10%	193	123.115
20%	49	36.4537
30%	18	20.4159
40%	11	14.6742
50%	4	7.9069

Percentages		
Cutoff Percentile	Perfect faults	Predicted faults
0%	0%	0%
10%	43%	27%
20%	54%	35%
30%	58%	40%
40%	60%	43%
50%	61%	45%

Performance	
Cutoff Percentile	Perfect faults/Predicted faults
10%	64%
20%	66%
30%	69%
40%	72%
50%	74%

MOM Summary

Provide analysis of your MOM results summary.

Based on the results from each MOM comparison, the most optimal model is Greedy Fit. When finding the predicted faults by % cutoff, each cutoff (10%-50%) is the overall highest. The second-best model is the M5 Fit. When finding the predicted faults by % cutoff, each cutoff (10%-50%) is around 2% lower than the results from Greedy Fit. Regarding performance curve, Greedy Fit also has the best results. In Greedy fit, the performance results range from 91% to 100% while in M5 Fit the performance results range from 90% to 97%.

Note: I believe the Fit performance results are so high because out of 192 samples around 50% are fault prone. This means for MOM to catch **all** faults the cutoff must be at least 50%. For both Fit MOM comparisons at the 50% cutoff all fault prone samples have been accounted for resulting in 100% perfect faults. Then with the weights from the models in HW I-B the Fit models from MOM can correctly organize the top 50% highest predictions.

Raw Excel Data

Due to sheer amount of data, the raw excel data will be included separately in the file titled, “Tudor_hwk_4_data.xlsx”.