***Doruk Özer 70192***

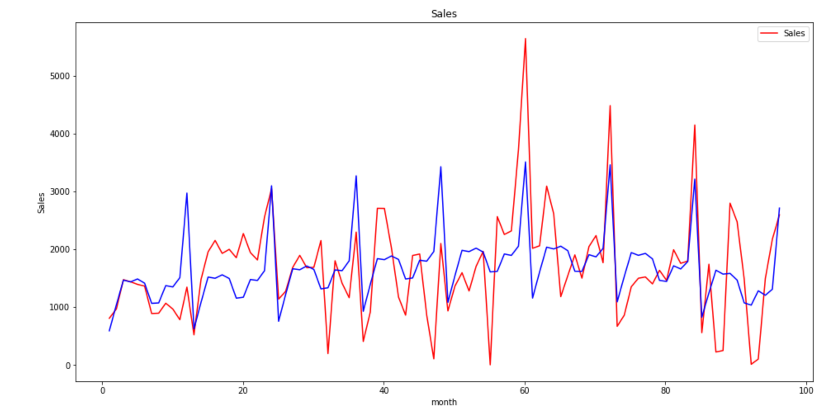
***Part 1: Regression and Model Reduction for Sales Prediction***

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu***Results for full model***

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu



metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Section B***

I fit a least squares regression to training set with the following predictors: monthly dummies at January, February, March, April, June, July, August, September, October, November December ,t, t-squared, t-cube and lastly t-cube. I observed that R^2 value is quite low which is 0.388 and it means that there is a high variability around the regression line. I checked the significance of predictors and observed that t, t-squared, dummies at February, March, April, June, August , September, October, November are insignificant according to null-hypothesis test. That shows that sale values are may not be dependent on those insignificant variables, or other variables in the model can explain those predictors. MAPE on the training set is 5.75, RMSE is 717 and MSE is 514458. RMSE and MSE values are high in the scale of the Sales data. MAPE may be undershooting in low sales period or on high sales period. Also visually we can see that data has strong ups and downs.

***Section C***

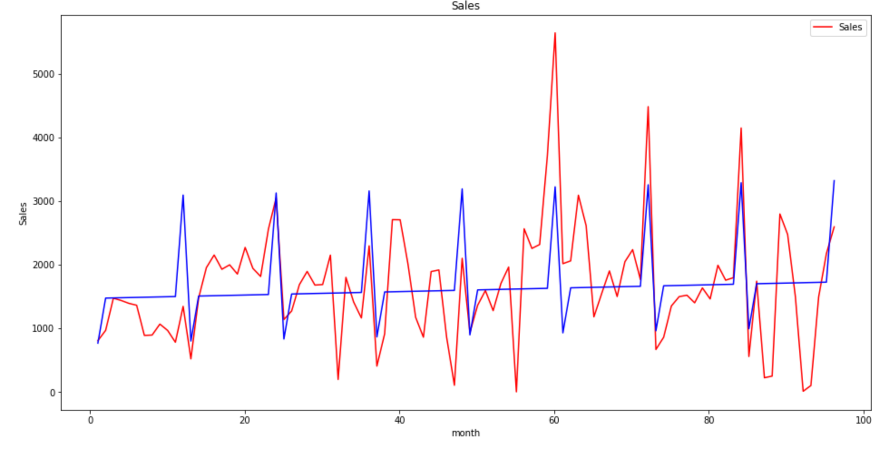
I fit the model to the test set and found out that MAPE value as 23 and MSE as 25055427 and RMSE as 5005. MAPE ,RMSE and MSE increased significantly. This is caused by the overfitting of the model to the training set. The complexity of the model is quite high with the dummies at the each month except May and with the other trend terms such as t , t-squared and t-cube. Those predictors are causing overfitting and many of them are statistically not significant. So the model can be reduced in terms of predictors that are used

***Results of Reduced Model***

***metin, makbuz içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldu***

******

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu ***Section E***

I fit a least squares regression to training set with the following predictors: monthly dummies at January, and December, and trend term t. It can be said that except trend term t all of the predictors are statistically significant. Also it is remarkable that R^2 value is nearly the same as the full-model. MAPE, RMSE, and MSE values of the reduced model is also similar with the full-model. Which indicates that the performance of the reduced model is as well as full model. Since full model is more complex than the reduced model, the reduced model should be preferred to use because their performance is similar.

***Section F***

The MAPE , RMSE and MSE performance of the reduced model is much better than the MAPE, RMSE and MSE performance of the full-model. The decrease in the predictor variables reduced the model complexity and solved the overfitting issue so it performed better in the test set. To sum up, reduced model should be preferred since the it’s complexity is lower, gives better performance at test set, solves overfitting issue

***Part 2: Logistics Regression and Model Classification for Market Directions:***

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu ***Results for Training Set of Full Model***

******metin, makbuz içeren bir resim

Açıklama otomatik olarak oluşturuldutablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

AUC result

Confusion Matrix

***Results for Test Set of Full Model***

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldu

AUC result

Confusion Matrix

***Section B***

I fit a logistic regression to the market direction data (up or down) in month t where the predictors are Lag 1 = yt−1 − yt−2, Lag 2 = yt−2 − yt−3, Lag 3 = yt−3 −yt−4, and Jan(indicator for January) and Dec (indicator for Dec.). It came up that the predictors of January and December are statistically insignificant. That is to say those predictors doesn’t explain ups and downs. They can be considered to be removed from the model.

***Section D***

Based on the model that I fit to the training set with the threshold 0.5 I found the Confusion Matrix, AUC values and the performance metrics. It showed up that it correctly predicted 42 of the up periods correctly out of 52, and it predicted 25 down periods out of 40 correctly. The sensitivity of the model is calculated as 0.81 (42 /52) and specificity as 0.62 (25/40). The AUC is 0.71 which is relatively far to 1 and it indicates it is somehow close to the perfect classifier. Error Rate is 25/92 which is 0.27.

***Section E***

Based on the model that I fit to the test set with the threshold 0.5 I found the Confusion Matrix, AUC values and the performance metrics. It showed up that it correctly predicted 36 of the up periods correctly out of 39, and it predicted 16 down periods out of 34 correctly. The sensitivity of the model is calculated as 0.92 (36 /39) and specificity as 0.47 (16/34). The AUC is 0.69 which is relatively far to 1 and it indicates it is somehow far to the perfect classifier. Error Rate is 21/73 which is 0.28 which is similar to the error performance at the training set..

***Results for Reduced Model on Training Set***

metin içeren bir resim

Açıklama otomatik olarak oluşturuldutablo içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Confusion Matrix

AUC result

metin, makbuz içeren bir resim

Açıklama otomatik olarak oluşturuldu

Results for Reduced Model on Test Set

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldu



AUC result

Confusion Matrix

***Section F***

I fit a logistic regression to the market direction data (up or down) in month t where the predictors are Lag 1 = yt−1 − yt−2, Lag 2 = yt−2 − yt−3 and Dec (indicator for Dec.). It showed that only the predictor December is statistically insignificant and others are statistically significant. Based on the model that I fit to the training set with the threshold 0.5 I found the Confusion Matrix, AUC values and the performance metrics . It showed up that it correctly predicted 42 of the up periods correctly out of 52, and it predicted 28 down periods out of 40 correctly. The sensitivity of the model is calculated as 0.81 (42 /52) and specificity as 0.70 (28/40). The AUC is 0.75 which is relatively better than the full-model and it is closer to the perfect classifier. Error Rate is 22/92 which is 0.23. This model is far better than the previous model. It increased the sensitivity and the specificity on training set. It also decreased the error rate on training set.

Based on the reduced model that I fit to the test set with the threshold 0.5 I found the Confusion Matrix, AUC values and the performance metrics. It showed up that it correctly predicted 33 of the up periods correctly out of 39, and it predicted 14 down periods out of 34 correctly. The sensitivity of the model is calculated as 0.84 (33 /39) and specificity as 0.41 (14/34). The AUC is 0.63 which is worse than the previous model’s test set performance. Error Rate is 26/73 which is 0.35 which is worse than the error performance full-model on training set. Sensitivity and the specificity performance of the reduced model is worse than full-model’s test set performance. Also the error rate is higher than the full-model on test set. This indicates that reduced model has overfitting issues since it performed better in training set and worse in the test set compared to the full-model.

***Section G***

AUC of the reduced model on the test set is computed as 0.63 and AUC of the full-model on the test set is computed as 0.69. That shows that reduced model is worse than full model. As in the previous section I indicated that reduced model has overfitting issue it performs worse on test set. Since it has overfitting issue full-model should be preferred instead of reduced model. Since it has lower error rate and better specificity and sensitivity on the test set.