Grant Starnes

Dr. Liz Wanless

TESM-S501: Advanced Sports Analytics

23 November 2025

Sprint Challenge 3:

A Focus on Predictive & Prescriptive Analytics, Player Side, Business Side, and Model Comparison

Summarize a predictive analytics project of your choice from a trade or peer-reviewed outlet. Cover each phase from the CRISP-DM and how the project did or did not address these phases. Critically assess the project: how would you improve what they constructed?

The predictive analytics project of my choice to look further into would be "A Predictive Metamodel for College Football" by Jay Coleman. Essentially, there are already quite a number of systems that have been developed for the sole purpose of rating and ranking major college football teams. Although there are so many systems, Coleman states that there are only two up to this point (2025) that have tried their hand at a metamodel that are most predictive of the outcomes of future games. Coleman then dives into what makes up his metamodel, which is his use of "29 college football rating systems, data for 5,925 games during 2016-2024, and a k-fold cross-validation process focused on model predictiveness." Coleman's metamodel is then a "five-system metamodel … developed to predict the victory margin for games in the ensuing week." He then goes on to say that the model attains strong results by looking at opening, midweek, and closing betting lines, and is statistically significant regarding the opening betting line in validation and test samples. Coleman then goes on to state that his metamodel could potentially be very useful for different stakeholders, for example, the college football playoff committee, media, fans, the betting market, and oddsmakers, and that his metamodel provides a benchmark against which future rating systems can be assessed.

Business Understanding:

- Addressed:
  - The objective is improving the predictions of game outcomes by using multiple rating systems and adding in game result
  - The scope of the data being used is for games in the range of 2016-2024
  - The problem stated would be improving prediction and its accuracy for college football overall
  - Uses a k-fold cross-validation process to enhance model predictiveness
- Not Addressed:
  - The predictiveness objective is clearly stated, but the business side is a bit ambiguous, for example, who specifically is going to use the metamodel?
  - The metamodel doesn't necessarily state who it can help make informed decisions
  - Coleman doesn't go into detail about things like constraints, the costs and benefits of the metamodel, or how the metamodel will be deployed or used in the real world is someone were to attempt to do so
- Improvements:
  - Coleman could outline specific stakeholders that could use the metamodel a bit more
  - The decisions that can be made based on the metamodel so those using it can act on it

  - Coleman could talk about the costs and benefits of his metamodel, for example, if the metamodel only increases the accuracy by a small percentage, what is the draw to stakeholders to actually use it, and what is the value

Data Understanding:

- Addressed:
  - Coleman uses data from 5,925 games in the range of 2016-2024 and from 29 various college football rating systems
  - Coleman uses cross-validation to split up and validate the data overall
  - Coleman absolutely has a variety of data since he utilizes different rating systems as well as the data from the 5,925 games
- Not Addressed:
  - There isn't exactly a conversation about the quality of the data Coleman uses. It would be better to include what the data actually looks like, for example, missing data (NAs), outliers, and biases if there are any
  - Coleman could have done a bit more of a deeper dive into the data, for example, areas we've looked at in class like home/away team or conference matchups
- Improvements:
  - Coleman could include more descriptive data, such as aggregations like means and medians, standard deviation, how often missing values occurred (NA count)
  - Coleman could also document any data handling or cleaning he conducted in the process

Data Preparation:

- Addressed:
  - Coleman uses cross-validation and combines the 29 college football rating systems
- Not Addressed:

- Coleman's research article is a rather light regarding data preparation overall as there isn't much regarding the cleaning and handling of all of the data Coleman was pulling in from his various sources
- Coleman doesn't refer to creating any tables or transforming data if necessary

- Improvements:
    - A simple improvement Coleman could make would be to simply document how he arranged and aligned his data from the numerous sources he was pulling from, how he transformed the data (if needed), and how he cleaned the data to prepare it for his model creation
    - Coleman could include actual information he got while preparing his data, like the frequency of missing values or what metrics/variables/features he utilized or threw out

Model Generation:

- Addressed:
    - Coleman uses k-fold cross-validation to evaluate the model performance
    - Coleman then built a metamodel that essentially combines the 29 rating systems and games from 2016-2024 to enhance prediction outcomes
    - Coleman includes multiple tables: College football rating systems used in analysis, Ordinary least squares metamodel from forward selection, Performance of metamodel (MM) and adjusted metamodel (MM*), Adjusted metamodel (MM*) in presence of betting lines, and Adjusted metamodel (MM*) win rates against the spread (ATS), excluding pushes and MM* = line.
- Not Addressed:
    - Coleman himself doesn't necessarily go into what modeling analyses or techniques he used personally, and doesn't compare his model to that of the other rating systems

    - There aren't any actual visualizations created by Coleman to compare his metamodel to the utilized rating systems, rather just a sequence of tables where he calculates things like R-Square, Root Mean Squared Error, Adjusted R-Square, mean absolute error, mean error (bias), and accuracy (correct winner)
- Improvements:
    - Coleman could document or even add in the techniques he used/could use (regression, random forest, etc.)
    - Provide a comparison of some of the rating systems he used to his new metamodel to show that there was an improvement of some kind and validate that his metamodel is accurate and should be used in the future
    - Include context regarding all of the metrics he used, like which metric(s) are most important or significant, and which rating systems are most important

Model Evaluation:

- Addressed:
    - Again, Coleman uses k-fold cross-validation to evaluate how well the metamodel is at predictive performance on the data included
    - Coleman uses what is most likely the most important metric, MM* (correct winner), which was roughly 75% accurate regarding prediction, and applies it to opening, midweek, and closing betting lines, where all three remain hovering around the 75% accuracy mark
- Not Addressed:
    - Again, Coleman doesn't include predictive accuracy of previous rating systems and compare the accuracy of those to his new metamodel to determine and provide support that his metamodel is superior to that of the existing rating systems, as well as leaves out the unique benefits and drawbacks of his metamodel and why it should be used over others

- Improvements:
    - Include improvement metrics regarding the metamodel and how it compares to that of existing rating systems, for example, the percentage increase in accuracy of prediction to drive stakeholders to actually use the metamodel
    - Include information about metamodel limitations at the current point and what goes in to the upkeep of the metamodel

Solution Deployment:

- Addressed:
    - Coleman's research article mainly just includes information focusing on the development and evaluation of his metamodel, but doesn't go in depth regarding the actual deployment of his metamodel
- Not Addressed:
    - How Coleman plans to deploy his metamodel, more specifically which stakeholders will use this metamodel, how can they integrate it, and how will it stay up to date over time as new data comes in over the coming seasons
    - There's no real discussion about how an organization would use the metamodel
- Improvements:
    - Simply include a section in the research article regarding how the metamodel could be integrated/reproducable for those in the sports analytics field and how it would stay up to date with the latest college football data or statistics

Upgrade the baseline win probability model we created in class via the creation of new variables, model comparison, and/or other creative ideas. Your new specification should improve the outcome in a meaningful way.

a. Compare the accuracy with the model we created in class.

Baseline Model Accuracy:

```
Coefficients:
                         Estimate  Std. Error z value            Pr(>|z|)    
(Intercept)            0.89275940  0.03658654  24.401 < 0.0000000000000002 ***
yardline_100          -0.00763883  0.00040080 -19.059 < 0.0000000000000002 ***
game_seconds_remaining -0.00001602  0.00000888  -1.804            0.071265 .  
down2                 -0.07915185  0.02318035  -3.415            0.000639 ***
down3                 -0.18641295  0.02710871  -6.876    0.00000000000613 ***
down4                 -0.36232814  0.03365466 -10.766 < 0.0000000000000002 ***
ydstogo               -0.01285578  0.00245470  -5.237    0.00000016302146 ***
posspread              0.11799042  0.00163143  72.323 < 0.0000000000000002 ***
score_differential     0.18526554  0.00152773 121.268 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 108611  on 78349  degrees of freedom
Residual deviance:  68617  on 78341  degrees of freedom
AIC: 68635

Number of Fisher Scoring iterations: 5
```

Upgraded Model Accuracy:

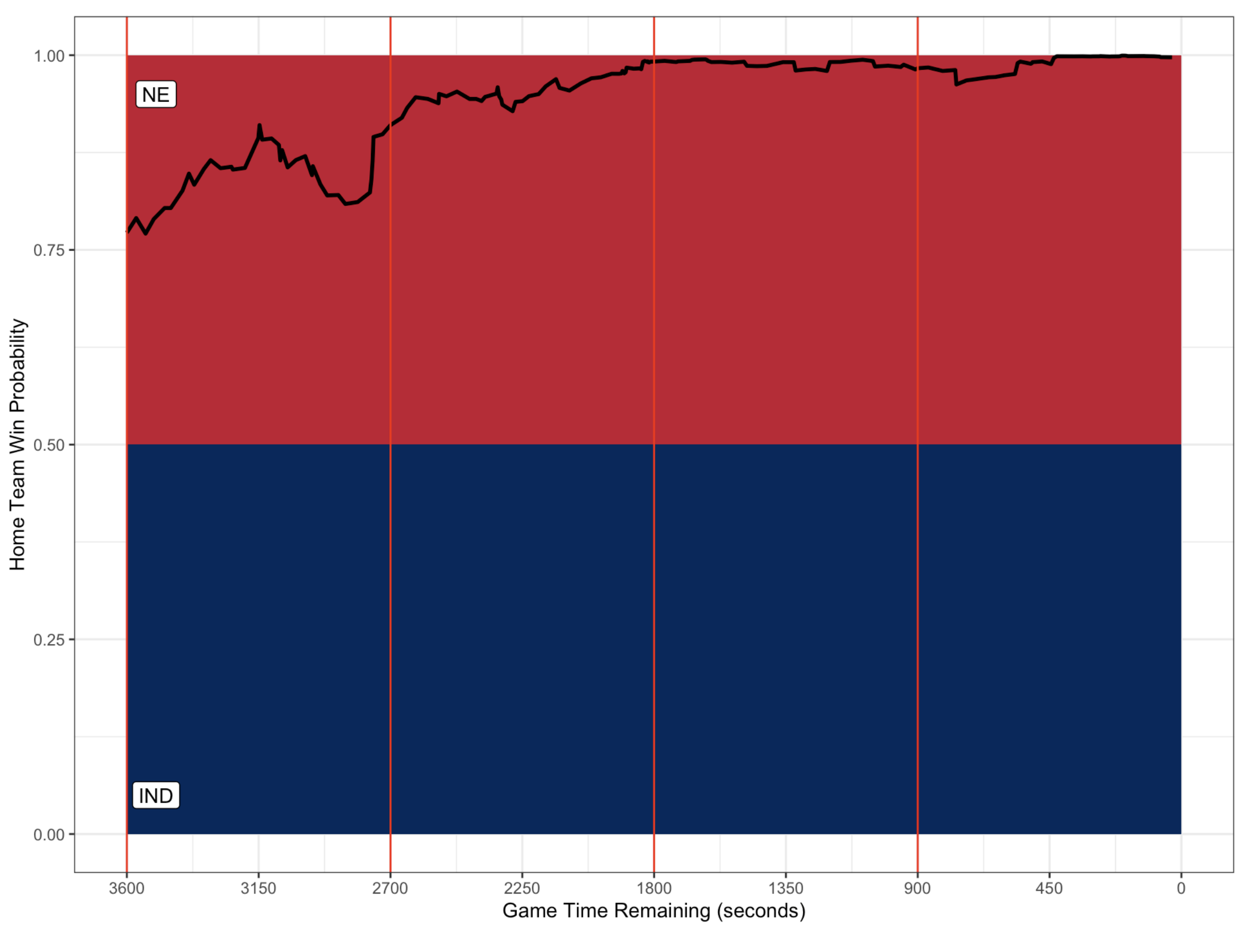
```
Coefficients:
                                           Estimate   Std. Error z value            Pr(>|z|)
(Intercept)                             1.201156022  0.115712414  10.381 < 0.0000000000000002
yardline_100                           -0.008271496  0.000406490 -20.349 < 0.0000000000000002
game_seconds_remaining                 -0.000126545  0.000035124  -3.603            0.000315
down2                                  -0.089953228  0.023342471  -3.854            0.000116
down3                                  -0.210488148  0.027342799  -7.698   0.0000000000000138
down4                                  -0.395032917  0.034068904 -11.595 < 0.0000000000000002
ydstogo                                -0.013501983  0.002475002  -5.455   0.0000000488784912
posspread                               0.117129496  0.001636735  71.563 < 0.0000000000000002
score_differential                      0.298899551  0.003824663  78.151 < 0.0000000000000002
qtr2                                   -0.052036929  0.040580041  -1.282            0.199728
qtr3                                   -0.054239153  0.068495463  -0.792            0.428439
qtr4                                   -0.065710106  0.100405973  -0.654            0.512826
game_seconds_remaining:score_differential -0.000067371  0.000001867 -36.094 < 0.0000000000000002

(Intercept)                            ***
yardline_100                           ***
game_seconds_remaining                 ***
down2                                  ***
down3                                  ***
down4                                  ***
ydstogo                                ***
posspread                              ***
score_differential                     ***
qtr2                                      
qtr3                                      
qtr4                                      
game_seconds_remaining:score_differential ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 108611  on 78349  degrees of freedom
Residual deviance:  67174  on 78337  degrees of freedom
AIC: 67200
```

To upgrade the baseline win probability model we created in class, I compared the predictive accuracy of the baseline win probability model with my upgraded win probability model where I added quarter (qtr) and an interactive kind of variable that works with score_differential : game_seconds_remaining. Based on the baseline model, the confusion matrix states that the overall accuracy is about 68%, with an AIC of 68,635 and a Residual Deviance of 68,617. This baseline model performs well and outputs a great visual, but it seems to treat time and score as linear and independent factors and doesn't take into account the actual game context, like the quarter, so predictions late in the game aren't as realistic. As for the upgraded model, again I incorporated qtr for game quarter and score_differential : game_seconds_remaining where the accuracy increased, AIC dropped to 67,200, and residual deviance dropped to 67,174. A lower AIC and low deviance helps indicate that the upgraded model fits to the data a bit better than the baseline model did and truly captures a more real NFL game dynamic with these new variables.

b. Explain the logic and reasoning behind your change(s) with a narrative explaining your approach.

To upgrade the win probability model from the baseline model from class, I mainly focused on some of the key limitations of the baseline model, which was that it treated all game situations as equally important and didn't account for when they were occurring. That said, in football games, a play and its impact changes dramatically depending on the quarter, time left, and in the context of the game, so my main goal was to refine the baseline model a bit more. By adding in the qtr metric, the model was able to learn different probabilities based on the different stages of the game. Implementing this change allows the model to adjust the win probability a bit more realistically as time goes on in the game, and adds in a bit more of a contextual awareness that one would experience in a real life game. The next metric I added in was score_differential * game_seconds_remaining as the overall impact of score differential varies over time, for example, if someone is leading by 3 with 12 minutes left, win probability won't change too much, but if someone is up 3 with 12 seconds left, there's a huge change in probability. The baseline model didn't exactly capture this and this is something that was talked about in class. By including this metric, the upgraded model is able to scale the effects of a score margin dependent on the remaining time. This makes the upgraded model more realistic, mainly in the 4th quarter or late game situations where one score can swing win probability potentially.

c. **Create a new win probability model for a game from a season/game date of your choice. Explain the trajectory of the curve based on the game itself.**



NFL · Oct 4, 18 Final

Indianapolis Colts (1 - 4) 24 - 38 New England Patriots (3 - 2)

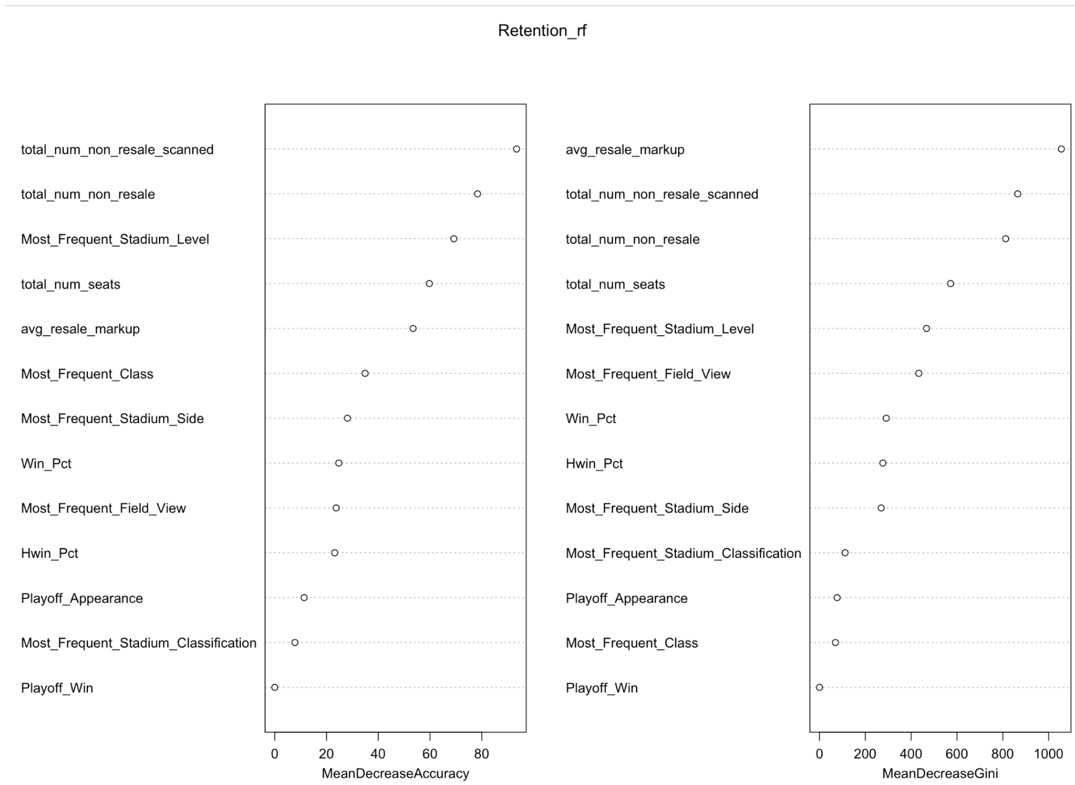
Team	1	2	3	4	T
Indianapolis Colts	0	3	7	14	24
New England Patriots	7	17	0	14	38

Change the business understanding focus of the customer retention model we created in class. This could be analyzing a different customer base or the same focus customer base with a different set of variables or the same base/variables with a model that supports interpretation instead of prediction accuracy. Respond to the following:

a. **What types of descriptive results might be important to the business perspective? Demonstrate the descriptive results with visuals in the software of your choice. Explain the meaning of your visuals.**

I used a very similar structure to that of the customer retention model we created in class, but analyzed a different customer base. My customer base was focused on the fans who simply attended 2 or more games, and I applied a filter to "Retention_cust" in order to do so, utilizing ">= 2". I then utilized a confusion matrix similar to the model from class to compare the predicted vs. the actual customer retention for fans who attended 2 or more games. Focusing on the specificity, the confusion matrix is rather strong regarding identifying renewing customers, as the specificity is roughly 96%. Although specificity is high, focusing more on sensitivity, it is far lower and roughly around 12% and is much more difficult to detect when looking at churn. Based on my rendition of the random forest model, some key findings could be that attendance is absolutely the best predictor. If we look at the random forest model, total_num_non_resale_scanned and total_num_non_resale hold the top of the model, suggesting that fans who actually use their tickets and attend games are more likely to renew their season tickets. We can also look at seating locations like level and seats. Most_Frequent_Stadium_Level and total_num_seats also contribute heavily, suggesting that seating location/quantity provide a lot of value to season ticket holders. Next we can look at the resale results where avg_resale_markup is decently high, which informs the team that ticket resellers behave differently than the actual ticket users and helps identify this. Lastly, we can look at results regarding team performance like win percentage and playoffs where, in this instance, they are not the best predictors based on the model. Team success can vary season to season, but individual fan engagement seems to matter far more to season ticket holders.

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0   356   399
         1  2631 10064

               Accuracy : 0.7747
                 95% CI : (0.7676, 0.7818)
    No Information Rate : 0.7779
    P-Value [Acc > NIR] : 0.8167

                  Kappa : 0.1106

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.11918
            Specificity : 0.96187
         Pos Pred Value : 0.47152
         Neg Pred Value : 0.79275
             Prevalence : 0.22208
         Detection Rate : 0.02647
   Detection Prevalence : 0.05613
      Balanced Accuracy : 0.54052

       'Positive' Class : 0
```



b. Explain your process and results – how does your new analysis compare to the class construction?

As stated above, I followed the same steps as the class construction, but made a customer base adjustment to where I filtered the customer base to fans who attended at least 2 games per season. These customers who go to 2 or more games each season behave differently and allow the model to focus on engagement patterns far more than outliers or those who don't attend the games. In the class construction, we used a segmentation that wasn't focused on attendance, but rather type of account/customer. The class model produced a great model, but as we talked about, there is a heavy class imbalance. I utilized a random forest model very similar to the class example, but on the customer base of those who attend 2 or more games per season. My confusion matrix yielded results like a roughly 78% accuracy, 96% specificity for renewal detection, 12% sensitivity for the difficulty to identify churn, 54% balanced accuracy, and still has a bit of a class imbalance. Even changing the customer base, my results were very similar to that of the class model where attendance is the strongest predictor of renewal, followed by seat location and actual team performance (win percentage and playoff appearance). My results remove the low engagement outliers and the model can more clearly separate the high engagement, moderate engagement, and on the fence renewals, and enhances the metric importance as seat characteristics/locations increased greatly.

c. **What would you tell business personnel about the short-term and long-term planning that could be associated with model results?**

There are a number of suggestions I would make to business personnel about the short-term and long-term planning that could be associated with the model results for fans that attended 2 or more games each season in this data set. For the short-term, since attendance is so important in this model, I could tell business personnel to target the low attendance customers early, and potentially the customers that have attended one or less games by week 4 into the season. Business personnel could reach out to them personally and offer them incentives to get them to return, such as concession discounts or parking validation. Business personnel could also improve their in-season service by focusing on the customers who may be on the fence and could disperse mid-season surveys to try and identify why these customers may be so on the fence. This makes sense because on the sensitivity, these customers are harder to predict, so actually doing a check in could be beneficial. For the long-term, teams could make improvements to the stadium itself, for example, the extracurriculars/entertainment within the stadium while customers are there, have flexible concessions options as trends may change over the years, and look for more efficient ways to get customers in the building, like entrances, parking, and metal detectors and ticket scanning. Teams could also introduce season tickets that can be bought for years in advance at one time, for example, season tickets that will last 2 or 3 seasons. Loyalty programs can be implemented, and the team could put on community/loyalty events to deepen the actual emotional attachment of customers.