Predicting Player Passes from Liverpool FC

by:

Kian Peters Mondry

Austin Huang

Christopher Morales

Suyesh Niraula

Introduction

In soccer, the number of passes on the ball for each game is a foundational metric that reflects both individual tactical roles and collective team identity. In our study, we aim to develop a predictive model to estimate an individual player's total attempted passes in an upcoming fixture by analysing historical performance data and specific match-day variables. To do this, we chose players from the Liverpool FC team in the Premier League since they consistently have high possession of the ball during games. Then we picked out specific Liverpool players who also had betting lines for passes attempted on a game day to make our model. A player's attempted passes are a common metric used by betting industries, where people are asked to predict over or under a certain number of passes. We are doing this because we are wondering how effective our predictions are compared to the actual game outcome and versus results from existing models like market efficiency and betting companies.

Prediction models for soccer games have been used for decades; however, models aren't always accurate at predicting statistical outcomes of games. By researching additional ways to predict a player's attempted passes, we can potentially improve the accuracy and see if we can beat betting predictions that are made for the average person. Essentially, the goal of this project is that, in addition to developing improved models for the field of sports statisticians, we can see if our models can give people an edge in outpredicting the betting market.

Different types of factors can influence the accuracy of predictions, such as results from past games, psychological aspects of the game environment, the opponent of the game, unpredictability of game outcomes (such as player injuries or red cards), a player's skillset, and a team's playstyle. These factors can be useful for analyzing our model's advantages and shortcomings compared to other models.

Thus, we will make prediction models accounting for numerous factors that train on previous data for one player in a particular team. These factors include, but are not limited to: average possession rate, home or away team, and the game result at half-time. Then, we will use this data to predict his passes (with 95% confidence intervals) for an upcoming game by using bootstrapping models, which will be used to simulate results from our trained data. Then, we will compare the results from the actual results of the game and market efficiency predictions to determine if our model can make a closer prediction compared to existing models.

Our model generated predictions within the 95% confidence interval for 2 out of 3 games we simulated. For these games, our model's predicted result was relatively close to the actual result. For the other game, our model generated a prediction lower than our 95% confidence interval, likely because of the unpredictability of the game outcome or a recent change in the player's role and performance. Our model did, however, outperform the market line every single time, by correctly identifying which of the over/under options for passes attempted would happen. While our model was able to outperform market models for each game, we can still improve our model by more dynamically adjusting for changes in players' roles.

Literature review

While statistical variables in soccer analytics are well-established, accurately predicting these statistics, such as determining the number of passes attempted for a player, remains a highly nuanced endeavour. Although researchers have evaluated passes through various tracking methods, aggregate passing for individual players and changes in passes from match to match are not commonly explored in current research.

There are, however, many team contextual factors that are already widely recognised in literature. For instance, it is well documented that home teams generally perform better than away teams, with this home-field advantage being even more pronounced for higher-quality teams such as Real Madrid and Liverpool (Wunderlich & Memmert). In addition, a team's playstyle and tactics interact significantly with these contextual factors (Wang et al.). Tactical decisions made by the coach, along with the degree to which a team adopts a defensive or offensive playstyle, heavily influence game outcomes and overall performance. Furthermore, current betting odds and computer models can reliably predict match outcomes, such as who wins and loses (Wunderlich and Memmert). Despite this, these methods mainly focus on team-level factors, rather than individual passes. While they often may affect the number of passes a player attempts, little research has thoroughly examined the variation of passes at the player level.

While not commonly researched, there is some known research and literature about predicting player passes, with some of the known extensive work dedicated to predicting individual passes through tracking data, probability models, path values, and other advanced tracking metrics (Wang et al.). However, forecasting a player's total passing volume, defined as the total passes attempted in an entire 90-minute game, utilising only easily accessible and publicly available data, is a much less explored area. Models relying strictly on pre-match data, such as home versus away dynamics, betting odds, half-time scorelines, and disciplinary actions (yellow or red cards), remain largely uncharted.

By analyzing only publicly available data, our project aims to create a predictive model that accurately incorporates lesser-known contextual variables to determine their impact on an individual player's passes attempted. Our project attempts to address the existing literature gap by attempting to predict a player's passing volume through accessible, contextual game factors rather than through traditional path values or tracking data.

Materials and methods

The primary objective of this study is to predict player pass volume within the specific tactical environment of the Premier League. We intentionally excluded past seasons of data due to changes in players and tactics over time that would lead to too much random and unpredictable variation. Therefore, we restricted our scope to this current 2025/2026 season. We also focused exclusively on Liverpool FC due to its league-leading average possession rate of 61.3%, which provides a stable environment for analysing high-volume passing metrics. Having

a constant high possession in each game leads to a lot less variation in possession for each game, also keeping the players' average passes more stable. For case 1, we gathered data for the past 26 Premier League matches that were extracted from Footmob and Football-Data.co.uk. In case 2, we have 27 games to train on, and in case 3, we have 28, as more games have been played by then. We intentionally excluded regional tournaments and Champions League fixtures, as the varying possession archetypes in European competition and tournaments do not align with the domestic league's playstyles. Since there are no draws in those competitions past the group stage, more teams are focused on a stronger defensive structure that may be different than how they play in the leagues. The predictive power of our model relies on the identification of specific environmental and performance-based metrics that influence a player's passing volume. This data includes a player's passes attempted for each of those previous games, a player's minutes played, position, the team's average possession, and a binary indicator for venue (home/away) status. To record a player's more recent tactical role in the team, we also specifically looked at the past 4-5 games' passes attempted. In our first test case, we simply kept his tactical form in mind while making the model, but for test cases 2 and 3, we tried to apply it statistically to our model.

We further plan to refine the model by incorporating in-game dynamics, such as the scoreline at half-time (Winning, Losing, or Drawing) and the occurrence of red cards. These factors may be critical, as red cards on a team's side can often negatively affect their possession and players' pass attempts. In addition, mid-match tactical shifts caused by the changing scoreline can lead to increased or decreased possession for players based on a team's playstyle. For example, often when winning, teams will keep possession by passing it around their backline, giving defenders the ball more, whereas when losing, attackers will get the ball more in the hopes of attacking rather than keeping it. Thus, to account for tactical shifts, we will also make an additional 3 predictions on top of one overall one: one for the case of winning at half-time, one for losing at half-time, and one for drawing at half-time. We also looked into adding red cards as a factor in our model; however, Liverpool games didn't have many games with red cards, so we removed them from our data. It would be more applicable if we did our study on a team like Chelsea, who have had many games recorded with red cards.

By integrating our identified environmental and performance-based factors into a linear model framework, the models train the performance-based metrics and in-game dynamics on the previous games to establish the baseline relationship between match conditions and player pass volume. We then apply a bootstrapping methodology to these regression results to generate robust point predictions and 95% confidence intervals for both the overall match and specific halftime game states. This dual-layered approach allows us to quantify the statistical uncertainty of a player's output across different tactical scenarios, ensuring our "edge" is grounded in a distribution of probable outcomes rather than a single, static average.

While bootstrapping cannot create new information, it addresses the small sample size by simulating the sampling distribution of player performance 5000 times and taking the mean result, which provides a more robust estimate of the model's standard error than a single linear

regression. This process allows the model to generate 95% confidence intervals that reflect the true stability of the data. In test cases 2 and 3, bootstrapping also helped to integrate weighted sampling for high-volatility roles. For example, for Ryan Gravenberch's midfield position, we applied a 5:1 or later for Van Dijk, a 10:1 bias towards the more recent games, so that the bootstrap model would be more likely to sample recent results over the rest of the season data. This helped statistically apply how recent form affects the players' passing into our model, rather than using guesswork and logic.

While our model may produce multiple results for each tactical scenario of half-time results, we need to predict the most likely one of these scenarios before the game starts in order to make a prediction for the betting line (which you cannot adjust once the game starts). This was done by analyzing online probabilities, past head-to-head results, and the form of both teams.

The three test cases were chosen based on whatever fixture was next and which player's betting line was available at the time on a betting site called Prizepicks. For the first game, only Van Dijk was available. For the second game, only Gravenberch was available. In the third game, Van Dijk and Allison were available, but we chose Van Dijk since we already had his data recorded. In each test case, we looked at the players' betting line that had an Over/Under that is supposed to be even (50/50 likelihood of picking either option line) for a certain number of passes attempted. Then, we made our prediction of what would occur before the game happened by using our model as our reasoning. Then, following the match, we noted down our model's performance by discussing how our model and reasoning fared in predicting the betting line, as well as how far off or close it was to the actual result and why that might have happened.

Results & analysis

Test Case 1

For the first game, we are testing the prediction on Van Dijk versus Nottingham Forest, a team with a relatively medium-low possession team, having a 48.4% average possession per game. The data consisted of the past 26 games of the season and were put in a data frame to train the data for this game.

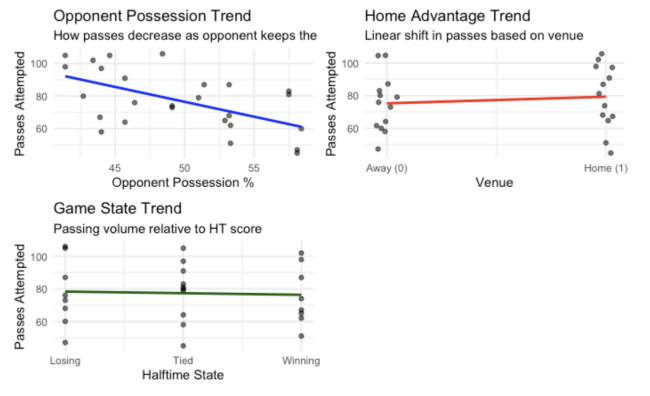
Van Dijk played every game and played a full 90 minutes in each one, which is rare for a player. This made calculations easier since we did not have to worry about converting passes per 90 and trying to predict his expected minutes. We also didn't have to worry about his position, as he played CB or centre back in all of them. This left no variation for that factor, so we excluded it from the data frame. Looking at the rest of the data, opponent possession seemed to heavily reflect variations, as playing a lower possession team often led to higher passes. Home

Van Dijk data Table:

Passing Performance Table

Passes	Minutes	Opponent Poss %	Home	HT State (-1=L,0=T,1=W)	Passes per 90	Passes per Min
58.00	90.00	44.00	0.00	0.00	58.00	0.64
45.00	90.00	58.10	1.00	0.00	45.00	0.50
51.00	90.00	53.30	1.00	1.00	51.00	0.57
73.00	90.00	49.10	0.00	−1.00	73.00	0.81
98.00	90.00	41.40	1.00	1.00	98.00	1.09
83.00	90.00	57.50	0.00	0.00	83.00	0.92
87.00	90.00	51.40	0.00	−1.00	87.00	0.97
91.00	90.00	45.70	1.00	0.00	91.00	1.01
102.00	90.00	43.40	1.00	1.00	102.00	1.13
79.00	90.00	51.00	0.00	0.00	79.00	0.88
65.00	90.00	52.90	1.00	1.00	65.00	0.72
64.00	90.00	45.70	0.00	0.00	64.00	0.71
97.00	90.00	44.00	1.00	0.00	97.00	1.08
80.00	90.00	42.70	0.00	0.00	80.00	0.89
106.00	90.00	48.40	1.00	−1.00	106.00	1.18
47.00	90.00	58.10	0.00	−1.00	47.00	0.52
87.00	90.00	53.20	1.00	1.00	87.00	0.97
76.00	90.00	46.40	0.00	−1.00	76.00	0.84
68.00	90.00	53.20	1.00	−1.00	68.00	0.76
60.00	90.00	58.40	0.00	−1.00	60.00	0.67
105.00	90.00	44.60	0.00	−1.00	105.00	1.17
67.00	90.00	43.90	1.00	1.00	67.00	0.74
105.00	90.00	41.40	0.00	0.00	105.00	1.17
81.00	90.00	57.50	1.00	0.00	81.00	0.90
62.00	90.00	53.30	0.00	1.00	62.00	0.69
74.00	90.00	49.10	1.00	1.00	74.00	0.82

versus away games barely showed any differences in his passes attempted, with home being very slightly more associated with higher passes. In addition, halftime results didn't have much of a change (will also see in final model results). We also noticed that his recent form (noted by the first couple of data entries corresponding to the last couple of games) indicated a much lower average than the rest of the season. This may be due to a change in playstyle and Liverpool wanting to change tactics after getting some struggling results. In this specific test case, we have not yet implemented weighting sampling to account for form in our model. We only noted that we should expect to get a little bit on the lower end of our confidence interval. This will change in test cases 2 and 3.



We then ran our bootstrapping model and applied these trends in past data to predict an away game at Forrest. Our model predicted 77 passes attempted overall, with a 95% confidence interval of around 69 to 86 passes. The tied at half time and losing at HT results are similar, but the winning at Half time prediction was a bit lower than the overall prediction. In terms of predicting the half-time scenario, we noted that despite Liverpool still being favorites to win (at least probability-wise on online sites), they have historically struggled to beat Forrest since Liverpool lost 3-0 at home last time they played. Also, Forrest plays at home, who can get good results, such as drawing to Arsenal at home (the current best team in the league). Thus, our logic would suggest us to predicting a competitive game, so we leaned more towards the tied at half-time model. The Over/Under line was at 77.5 on PrizePicks, a line that is very similar to our middle estimate. However, we noted earlier that our current model doesn't account for recent data, and that Van Dijk has been averaging significantly fewer passes than normal recently. Because of this, we estimated on the lower end of our confidence interval, predicting a passing range of low 70's and thus predicted under 77.5 passes for the betting line.

Prediction model:

```
Rate of passing: 0.8582104
Rate of passing CI: 0.7646101 to 0.9543477
Expected number of passes for full game: 77.23894
CI for expected full-game passes: 68.81491 to 85.89129
Winning at HT:
Expected passes: 74.68894
CI: 66.26491 to 83.34129

Tied at HT:
Expected passes: 77.23894
CI: 68.81491 to 85.89129

Losing at HT:
Expected passes: 76.68894
CI: 68.26491 to 85.34129
```

The actual result for Van Dijk's passes versus Nottingham Forest was 74 passes attempted. Our model ended up being very close to the actual result, as it was well within the confidence interval and only 3 off from our mean prediction (though this is without form weighting). Using logic and our models' statistics, we also made an accurate prediction that bested the betting lines model. Our assessment of taking the under against models was accurate

(even if it was very close) and shows a slight edge on our model and logic beating market efficiency, at least for this game. With a low sample size, we cannot conclude that this wasn't just by chance.

Test case 2

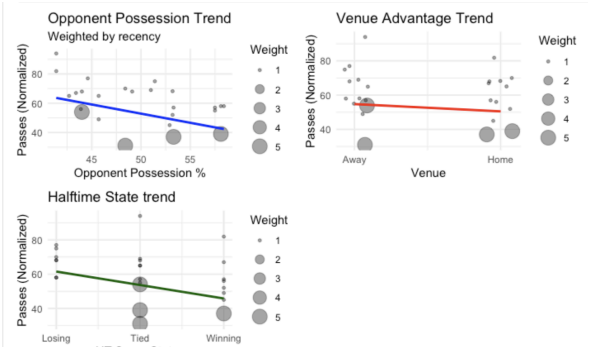
This time, we are looking at Gravenberch playing against West Ham, a low-possession team with an average possession of 42.4%, the second lowest in the league. The data here consists of only the past 25 games of the season and was put in a data frame to train the data for this game. It's only 25 because he missed 2 games during the season, and we are adding the last fixture versus Nottingham Forest (so he was in 25/27 possible games).

Some noticeable comments about the data are that he plays center defensive midfielder, or CDM, which he did for all the games he played. Thus, again, we didn't have to worry about his position, which is why it's not in the data frame. He played the majority of games (mostly 90 minutes), but missed 2 through injury and also was subbed early in blowouts or when Liverpool also had an important midweek game coming up (like a Champions League game). To account for this, games he didn't play were removed, and the few games with low minutes are converted to per 90 passes. Since he is not always guaranteed 90 minutes per game, we have to try to predict his expected minutes. For this specific game versus West Ham, we are simply going to assume another full game of 90 minutes, since there is no following Champions League game or injury. Like the Van Dijk data, average possession was especially important since high variation was seen in his passes, it usually being more when the opponents had a relatively lower average possession. However, unlike Van Dijk, the prediction for the half-time state, especially passes when losing at half-time, was especially higher than when tied or winning, which we will see emphasized in our final model. One important trend we also noticed was that his last 4 games of passes attempted were considerably lower than the rest, with a 40.25 average of passes attempted

Gravenberch data table:

Passing Performance Table

Passes	Minutes	Opponent Poss %	Home	HT State (-1=L,0=T,1=W)	Passes per 90	Passes per Min
31.00	90.00	48.40	0.00	0.00	31.00	0.34
54.00	90.00	44.00	0.00	0.00	54.00	0.60
39.00	90.00	58.10	1.00	0.00	39.00	0.43
37.00	90.00	53.30	1.00	1.00	37.00	0.41
68.00	90.00	49.10	0.00	−1.00	68.00	0.76
71.00	78.00	41.40	1.00	1.00	81.92	0.91
55.00	90.00	57.50	0.00	0.00	55.00	0.61
75.00	90.00	51.40	0.00	−1.00	75.00	0.83
65.00	90.00	45.70	1.00	0.00	65.00	0.72
67.00	90.00	43.40	1.00	1.00	67.00	0.74
69.00	90.00	51.00	0.00	0.00	69.00	0.77
45.00	90.00	52.90	1.00	1.00	45.00	0.50
49.00	90.00	45.70	0.00	1.00	49.00	0.54
68.00	90.00	44.00	1.00	0.00	68.00	0.76
65.00	90.00	42.70	0.00	0.00	65.00	0.72
70.00	90.00	48.40	1.00	−1.00	70.00	0.78
58.00	90.00	58.10	0.00	−1.00	58.00	0.64
52.00	90.00	53.20	1.00	1.00	52.00	0.58
47.00	62.00	53.20	1.00	−1.00	68.23	0.76
58.00	90.00	58.40	0.00	−1.00	58.00	0.64
77.00	90.00	44.60	0.00	−1.00	77.00	0.86
56.00	90.00	43.90	1.00	1.00	56.00	0.62
94.00	90.00	41.40	0.00	0.00	94.00	1.04
57.00	90.00	57.50	1.00	0.00	57.00	0.63
57.00	90.00	53.30	0.00	1.00	57.00	0.63



versus 60.3 for the rest of the season. This was also while playing a mixture of high, medium, and low average possession teams, which may indicate a new role for him in recent games that we need to account for. This time, to account for this, we applied bootstrap weighting to our model and placed a higher weighting on the last 4 games. Specifically, the last 4 games were weighted by the ratio 5:1, meaning they were 5 times more likely to be sampled when bootstrapping, so that our end mean and Confidence Interval are slightly skewed more to the recent results.

We ran our bootstrapping model and applied these trends in the past data to predict a home game against West Ham. Our expected mean passes are close to 60, with 95% a confidence interval of 44 to 75 passes attempted. We see a large variability of the scenarios at the expected mean passes. For losing, the expected passes are around 78, while the expected passes for tied and winning at halftime results are closer to 60. This is likely because midfielders are responsible for driving forward when losing and the team needs a goal, so they are on the ball a lot more. To predict the half-time scenario, we looked at online computers, which have Liverpool at a 71% chance of winning before the game and a 14% chance of drawing, totaling an 85% chance of not losing. In addition, Liverpool has historically always beaten West Ham, home or away. Thus, we shift our prediction to lean towards winning or drawing at half-time predictions. We predicted around mid to high 50's. Betting models on Prizepicks have their line at 61.5 for the over/under. This one can be harder to predict, since if Liverpool is losing at halftime, our model would suggest the higher, though if Liverpool is winning or drawing, we would take the under. However, given the likelihood that they are heavy favorites not to lose this game, we suggest the under.

Prediction Model:

```
Rate of passing (per min): 0.6644956
Rate of passing CI: 0.4912559 to 0.8337967
expected passes (Full Game): 59.80461
CI for expected full-game passes: 44.21303 to 75.0417
Winning at HT:
Expected passes: 59.53304
CI: 43.94147 to 74.77014

Tied at HT:
Expected passes: 59.80461
CI: 44.21303 to 75.0417

Losing at HT:
Expected passes: 77.86933
CI: 62.27776 to 93.10643
```

The actual result of Gravenberch's passes attempted versus West Ham was 40 passes attempted (of which Gravenberch played 86 minutes, which is close to a full game). Thus, our model predicted much higher than the actual amount. We noticed that Gravenberch performed similarly to his average from the past 4 games (40.25). This may mean that we may have failed to fully account for his role change and position volatility in Liverpool's recent shifted tactics. Or, we may potentially need to weigh his recent form even higher for big role changes. It's also possible that this could just be an outlier or due to random chance, as a key note is that this game was an unusual blowout in favor of Liverpool, being 3-0 up by halftime, meaning Liverpool didn't need to attack with the ball or even have it much at all. Regardless, our logic of taking

under the betting companies' line of 61.5 was still accurate, as they too predicted a high amount of passes. Thus, our model continues to show signs that it can still provide a slight edge on beating market efficiency.

Test Case 3

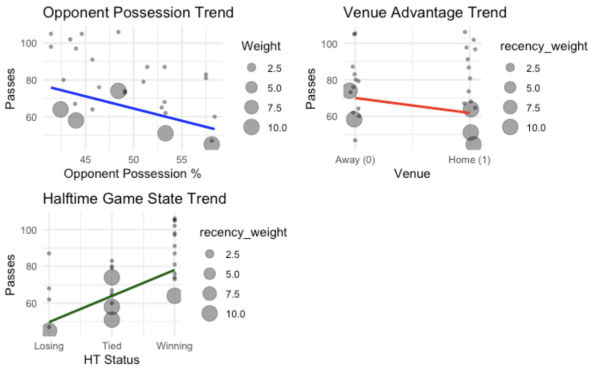
For our last case, we are once again doing Van Dijk, this time against the Wolverhampton Wolves away, a 43.4% possession team (one of the lowest in the league). After adding the recent Forrest and West Ham games, our new data consisted of the past 28 games of the season and were put in a data frame to train the data for this game.

Van Dijk Data Table:

Passing Performance Table

Passes	Minutes	Opponent Poss %	Home	HT State (-1=L,0=T,1=W)	Passes per 90	Passes per Min
58.00	90.00	44.00	0.00	0.00	58.00	0.64
45.00	90.00	58.10	1.00	0.00	45.00	0.50
51.00	90.00	53.30	1.00	1.00	51.00	0.57
73.00	90.00	49.10	0.00	−1.00	73.00	0.81
98.00	90.00	41.40	1.00	1.00	98.00	1.09
83.00	90.00	57.50	0.00	0.00	83.00	0.92
87.00	90.00	51.40	0.00	−1.00	87.00	0.97
91.00	90.00	45.70	1.00	0.00	91.00	1.01
102.00	90.00	43.40	1.00	1.00	102.00	1.13
79.00	90.00	51.00	0.00	0.00	79.00	0.88
65.00	90.00	52.90	1.00	1.00	65.00	0.72
64.00	90.00	45.70	0.00	0.00	64.00	0.71
97.00	90.00	44.00	1.00	0.00	97.00	1.08
80.00	90.00	42.70	0.00	0.00	80.00	0.89
106.00	90.00	48.40	1.00	−1.00	106.00	1.18
47.00	90.00	58.10	0.00	−1.00	47.00	0.52
87.00	90.00	53.20	1.00	1.00	87.00	0.97
76.00	90.00	46.40	0.00	−1.00	76.00	0.84
68.00	90.00	53.20	1.00	−1.00	68.00	0.76
60.00	90.00	58.40	0.00	−1.00	60.00	0.67
105.00	90.00	44.60	0.00	−1.00	105.00	1.17
67.00	90.00	43.90	1.00	1.00	67.00	0.74
105.00	90.00	41.40	0.00	0.00	105.00	1.17
81.00	90.00	57.50	1.00	0.00	81.00	0.90
62.00	90.00	53.30	0.00	1.00	62.00	0.69
74.00	90.00	49.10	1.00	1.00	74.00	0.82

A lot of the noticeable comments about the data remain the same as in test case 1. Van Dijk is still getting full minutes and only playing CB, so we don't have to worry about those factors. The opponent's possession also remained an important factor, with low possession indicating higher passes. Van Dijk's recent form has also been lower than the rest of the season data, a trend that we saw starting in test case 1. This time, we applied bootstrapping weights to account for this change in recent form, much like we did in case 2. We weighted the past 5 games in a 10:1 ratio, making the past 5 games much more likely to be sampled by the model. We decided to increase the weighting since we felt we didn't properly account for Gravenberch's form and role change in test case 2, and we hoped that placing a greater emphasis on recent form might help. One small difference from test case 1 about Van Dijk's data is that his halftime result has more of an effect on his probable end passes. This is different from last time, where the difference in the half-time result barely mattered for Van Dijk. This could likely be due to how the weighting impacted the data.



We ran our bootstrapping model and applied these trends in the past data to predict an away game at Wolves. The expected passes are 69 passes with a 62 to 77 confidence interval. If they are winning at half-time, our model predicts a huge increase in passes for Van Dijk, whereas when losing, this number is a lot lower. For predicting this live match

Prediction Model:

```
Rate of passing (per min): 0.7731745
Rate of passing CI: 0.6962801 to 0.8639358
expected passes (Full Game): 69.58571
CI for expected full-game passes: 62.66521 to 77.75422
Winning at HT:
Expected passes: 85.53043
CI: 78.60993 to 93.69894

Tied at HT:
Expected passes: 69.58571
CI: 62.66521 to 77.75422

Losing at HT:
Expected passes: 57.66679
CI: 50.74629 to 65.8353
```

state, we actually had a bit of a dilemma. Based on league position and the most probable scores of winning by supercomputer models, Liverpool is the heavy favorite. This is because Wolves are bottom of the league table. That being said, in past games, Wolves tied the best team in the league, beat third place, and have performed well recently at home. Despite Liverpool being heavy favorites by league position (5th place versus 20th), Wolves could easily make it a competitive game, and Liverpool likely won't have an easy game, especially considering their own recent lackluster results. If we are expecting a competitive game, we should lean towards the tied or maybe even losing at half-time range, which is the mid to high 60s. Betting models on Prizepicks have their line at over/under 80.5. Our prediction of mid 60's (due to the likelihood of a competitive game) would suggest taking the under, but the model would suggest the over if Liverpool has a comfortable victory.

The actual result of Van Dijk's passes versus Wolves was 63 passes attempted. As predicted, it was a competitive game with the half-time result being tied and the end result being 2-1 in Wolves' favor. This meant our assessment of predicting to be around the tied half-time result scenario ended up being right, and our logic of the mid-60s wasn't too far off from the final result of 63 passes. The market model prediction was over/under 80.5, which was way higher than the actual result. The market models were likely expecting an easy win for Liverpool or relied too much on full-season data rather than recent form for passes attempted. Regardless, we continue to outperform the market and show that our methods are giving us an edge in prediction results.

Discussion & conclusion

In our three tests, the actual result was near our model's predicted passes 2 out of 3 times, with the test for Gravenberch's game having a much lower result compared to our prediction. We outperformed the market models each time, showing that our model and logic can give you an edge in outpredicting the market models. The probability of getting it right all 3 times in the supposed "even lines" created by the market is 1/8, or .125 (assuming the even lines are 50/50). This is not statistically significant under common alpha thresholds like .05, so we need more tests before we can claim our model's approach to be conclusively better. Also, the struggles with the Gravenberch prediction compared to the actual result are a concern that needs to be addressed, but with more test results, we will have a better understanding if this result is an outlier or if there's a flaw with our prediction model.

Thus, one area for improvement in this project is to conduct more tests and comparisons, which would yield more reliable results than the 3 tests we conducted. In addition, we could look for more potential factors, such as determining if there are any other underlying factors that could give us more precise calculations. For example, how could we assess role changes better, like in the Gravenberch case?

One limitation is the scope to which our model can apply. Because of the methods we used (especially due to our source of data collection), we could only look at teams in one

domestic league. This meant we could not predict Champions League matches (a competition of the best teams across Europe) since possession stats from differing leagues don't carry over well.

Also, there is still a degree of randomness. We can attempt to predict a player's expected minutes, but we will never actually know what will occur in a game, especially if he gets injured, receives a red card, or is an early substitution in a blowout game. Even if we try to account for it, we will never be right 100% of the time. This is why this study would be better if this study was longitudinal to account for randomness in a single game, as it could be averaged out and/or seen as an outlier when it's mixed in with lots of results.

Lastly, there are 9 games left in the season. Even though this class will be well over by the time they are all played, our group will continue to predict results to increase the sample size and determine more conclusive results on our own time.

Sources:

Football Betting, www.football-data.co.uk/notes.txt. Accessed 9 Mar. 2026.

Wang, Si Hang, et al. “A Systematic Review about the Performance Indicators Related to Ball Possession.” *PloS One*, U.S. National Library of Medicine, 17 Mar. 2022, pmc.ncbi.nlm.nih.gov/articles/PMC8929629/.

Wunderlich, Fabian, and Daniel Memmert. “The Betting Odds Rating System: Using Soccer Forecasts to Forecast Soccer.” *PloS One*, U.S. National Library of Medicine, 5 June 2018, pmc.ncbi.nlm.nih.gov/articles/PMC5988281/.