

Analysing the Impact of Tackling Techniques on Play the Ball Speed in Rugby League

Project Description

The specific objective of this project is to determine how various tackling techniques and game contextual factors influence the speed of the 'Play the Ball' (PTB) in rugby league games. By analysing data on different types of tackles, positions on the field, times within the game, and the current scores, we aim to predict the PTB speed following a tackle. This analysis will help us understand whether defensive strategies are adjusted to speed up or slow down the PTB based on the game's current status, such as whether a team is leading or trailing, and during critical moments of the match.

The speed of Play the Ball (PTB) in rugby league is a pivotal factor for strategic planning and real-time decision-making on the field. Our project seeks to predict PTB speed post-tackle, examining how different variables influence this critical aspect of the game. Insights from discussions with an ex NRL coach highlighted the significant role of PTB speed in game outcomes, with strategies employed by teams like the 2005 West Tigers and Melbourne Storm illustrating how manipulating PTB can lead to victories.

This analysis will enable coaches to refine their team's tactics and training regimens, focusing on specific tackling techniques that strategically accelerate or decelerate PTB under various game conditions. By integrating data science with practical sports strategy, our project not only enhances the analytical tools available to teams but also provides actionable insights that can transform competitive play.

Previous Work

Previous analyses of play-the-ball (PTB) speed in the NRL have consistently shown a trend of increasing speed over recent years. For example, data from 2024 shows that PTB speeds have improved by one-tenth of a second compared to 2023, and 2023 was faster than 2022 by six-hundredths of a second. However, multiple studies have concluded that there is no direct correlation between faster PTB speed and outcomes, such as winning games or increasing the margin of victory (the difference in points between the two teams). Research has found no relationship between PTB speed and a team's position on the NRL ladder. Additionally, there is no correlation between PTB speeds and referees or PTB speeds and players (individual PTB speed) to impact match results or margin.

However based on the researchers analysis on NRL matches¹, factors that do influence PTB speed include the scoreboard, position from tryline and time in game. The links to the images below displays the results of factors that impact PTB:

- <https://www.rugbyleagueeyetest.com/wp-content/uploads/2024/08/image-32.png>
- <https://www.rugbyleagueeyetest.com/wp-content/uploads/2024/08/image-22.png>

¹ **Reference:** LeagueEyeTest (2024). *If play the ball speed means something but mostly nothing, what is a better metric? - The Rugby League Eye Test.* [online] The Rugby League Eye Test. Available at: <https://www.rugbyleagueeyetest.com/2024/08/19/if-play-the-ball-speed-means-something-but-mostly-nothing-what-is-a-better-metric/> [Accessed 15 Sep. 2024].

Sources

The data used in this project consists of multiple datasets from 2020–2023 NRL men's match events. The key data sources include Event and Match Data. The **Event** Data contains event-level information from NRL matches, focusing on specific in-game actions. Each row represents a unique event with key features including *MatchId* (a unique identifier for each match), *EventName* (the specific type of event with more focus on Tackle and PTB), *X-Y Coordinates* (the location on the field where the event took place), *Elapsed Time* (the time during the match when the event occurred), *Score* and *OppScore* (the current score for both teams, useful for calculating margin), and *Points* (which tracks point-scoring events).

The **Match** Data includes contextual information about each match such as *VenueId* (venue where the match was played), *WeatherConditionId* and *ClubId/OppositionId* (teams involved in the match). By combining these data sources (see Workflow section), this project aims to create a dataset capturing the relevant tackle-PTB events with variables that influence PTB speed ready for the modelling team.

Workflow

Step 1: Filter event data

The Event Data.csv data is the main file that will be used. Since it is a large dataset, the first step is to filter down the rows to only those relevant to the objective. In this case, we only need to consider the rows relating to either a tackle or play-the-ball. Using pandas in Python, we can filter rows based on the variable *EventName* by searching for the keywords 'Tackle' or 'PTB'. This will give us a more concise dataframe that is easier to manipulate.

Step 2: Join data files

The next step is to join the filtered event file data with the Matches.csv file in order to add more contextual information, which can be done by left-joining on the *MatchId* variable.

The Player and Lineup datasets are not included in the final data product as, using previous research of general PTB trends, it is known that these two do not impact PTB performance. This also helps us prevent unneeded complexity in the final product and overfitting during the model creation.

Step 3: Combine tackle events with their corresponding PTB

Next, we need to match each tackle event with its corresponding PTB speed. By combining this data onto a single row, it makes it easier to analyse these factors in correspondence to each other. We can combine these rows by matching using the X and Y coordinates of the event and the time at which the event occurred using the information prevalent in the data.

Step 4: Variable Exclusion

The final step to process the data would be to exclude variables that have little correlation to the PTB. As the dataset has many variables, excluding the ones less likely to meaningfully contribute to the analysis will improve performance of the model and reduce complexity. Variables to be excluded are identified using a combination of partial correlation and regression techniques. Calculating the partial correlation

between PTB speed and other features allows us to calculate the direct linear relationship between these two variables. Variables with each partial correlation are excluded as they do not significantly impact PTB speed when other factors are considered. We can then use regression to find variables that are least predictive for PTB speed using their statistical significance.

Data Description

The data product created in this project will leverage the historic NRL data provided in the most efficient way possible. Each **row** in the dataset will feature a tackle-PTB event pair. This will capture the critical moment related to PTB speed which is the tackle followed by the actual play-the-ball.

The **features (columns)** will include crucial in-game variables of the event pair including where on the field the event took place, time of event, duration etc. External features will be considered on a case-by-case basis based on whether or not they contribute significantly to the PTB speed (see Workflow section). Finally, an additional variable will be added based on the year of event which accounts for the NRL rule change that was implemented in 2020.

Project Status

We have completed our background research, asked experts in the field for what they find important in PTB speed, and explored the given datasets using Python to gain insights into their structure/relationships. Our current work is on joining the event data for tackles, PTB speed, and other peripheral contextual information.

After we complete this data filtering and joining, we will be working on cleaning the data. Removing anything that has inconsistencies, or large missing data points which could negatively affect accuracy of models using this data product. Then we will move onto the steps outlined in the *Workflow* to create the final data product.

Usage

We plan for our product to be used for finding which factors affect PTB speed more in certain scenarios, likely scenarios that coaches would specifically ask about. A few ways this could be handled are using **K-Means Clustering** to get an intuition into how similar tackles occur, and the key differences that affect PTB speed between otherwise “similar” tackles. Or an ML model (e.g. Random Forest Regression) to get a decision tree which optimises for prediction accuracy, and comfortably handles non-linear relationships like the ones we expect to see in this real world data with a high number of variables.

Support information: Contact David on teams or via email david.warton@unsw.edu.au.

Contributors: Aiden Poswell, Ivy Williams-Kell, Nabiha Imran Rajput, Tuvya Jugnarain

If you would like to get involved with this project please contact David above!