Hackathon Data Dictionary

Tennis, one of the most popular professional sports around the world, still uses manual coding of point outcomes. The purpose of this cute is to find a better approach.

Every tennis match is made up of a sequence of points. A point begins with a serve and players exchange shots until a player makes an error or is unable to return a shot in play.

Traditionally, the shot ending a point in tennis has been had been described in one of three mutually exclusive ways: a winner, an unforced error, or a forced error. A winner is a shot that was in play, not touched by the opponent, and ends with the point going to the player who made the shot. The other two categories are two distinct types of errors where both end with the point going to the player who did not make the shot.

The dataset includes point outcomes of rallies only (where the number of shots hit exceeds two, which represents the serve and return). All points were played at a past Australian Open.

Two datasets are provided for men and women. Both the men’s and women’s data have 5,000 points for training. There will be 2,000 points for testing the men’s and 1,000 points for testing the women’s solutions.

|  |  |  |
| --- | --- | --- |
| Variable | Description | Value Range |
| rally | The number of shots in the point counting serves and point-ending shot | An integer from 1, 2, 3... |
| serve | A number indicating whether the point was played on a first or second serve. | 1 = First  2 = Second |
| hitpoint | Shot category for point-ending shot | F = Forehand  B = Backhand  V = Volley  U = Unknown |
| speed | Speed of point-ending shot | Continuous (m/s) |
| net.clearance | Distance above the net as point-ending shot passed the net | Continuous (cm) distance above net. Can be negative if shot did not pass above the net. |
| distance.from.sideline | Lateral distance of the point-ending shot bounce from the nearest singles sideline. | Perpendicular distance in meters (always positive even if out) |
| depth | Distance of the point-ending shot bounce from the baseline | Perpendicular distance in meters  (always positive even if out) |
| outside.sideline | Logical indicator of whether point-ending shot landed outside of the in-play singles sideline | TRUE, FALSE |
| outside.baseline | Logical indicator of whether point-ending shot landed beyond the in-play baseline | TRUE, FALSE |
| player.distance.travelled | Distance player who made the point-ending shot travelled between the impact of the penultimate shot and the impact of the point-ending shot | Euclidean distance in meters |
| player.impact.depth | Distance of player who made point-ending shot from the net at the time the point-ending shot was made | Perpendicular distance along the length of court from net in meters |
| player.impact.distance.from.center | Distance of player who made point-ending shot from the center line at the time the point-ending shot was made | Perpendicular distance from the center line in meters |
| player.depth | Distance of player who made point-ending shot from the net at the time the penultimate shot was made | Perpendicular distance along the length of court from net in meters |
| player.distance.from.center | Distance of player who made point-ending shot from the center line at the time the penultimate shot was made | Perpendicular distance from the center line in meters |
| opponent.depth | Distance of opponent from the net at the time the at the time the penultimate shot was made | Perpendicular distance along the length of court from net in meters |
| opponent.distance.from.center | Distance of opponent from the center line at the time the penultimate shot was made | Perpendicular distance from the center line in meters |
| same.side | Logical indicator if both player and opponent were positioned on the same side of the center line (ad or deuce court) at the time the penultimate shot was made | TRUE, FALSE |
| previous.speed | Speed of penultimate shot | Continuous (m/s) |
| previous.net.clearance | Distance above the net as penultimate shot passed the net | Continuous (cm) distance above net. Can be negative if shot did not pass above the net. |
| previous.distance.from.sideline | Lateral distance of the penultimate shot bounce from the nearest singles sideline. | Perpendicular distance in meters (always positive even if out) |
| previous.depth | Distance of the penultimate shot bounce from the baseline | Perpendicular distance in meters  (always positive even if out) |
| previous.hitpoint | Shot category for penultimate shot | F = Forehand  B = Backhand  V = Volley  U = Unknown |
| previous.time.to.net | Time for penultimate shot to be hit and pass the net | Continuous number in seconds |
| server.is.impact.player | Logical if player who made point-ending shot was the server of the point | TRUE, FALSE |
| outcome | Target variable, character with three categories indicating the type of shot that ended the point | W (Winner), FE (Forced Error), UE (Unforced Error) |
| id | A 10-character unique identifier for the point | Character |

# The actual distance from the sideline

1. {'true.distance.from.sideline, distance.from.sideline, outside.sideline}

# The actual distance from the baseline

2. true.distance.from.baseline = depth outside.baseline

# Sum of distance of Opponent & player from net

3. player.distance.from.other.length = player.depth + opponent.depth

# Sum of distances of Opponent & player from center

4. player.distance.from.other.width = {player.distance.from.center

opponent.distance.from.center

same.side}

# Straight line distance between players

5. bw.player.distance.penultimate ={player.distance.from.other.length, 'player.distance.from.other.width}

# Distance penultimate shot was made from net- previous to net \* penultimate shot

6. player.from.net.penultimate = {previous.speed, previous.time.to.net}