

Data Sources

Data Dictionary: [Link](https://docs.google.com/spreadsheets/d/1Gt_Nnit9m4yTpemOcuiL-NzOMbtlrO1R/edit?usp=sharing&ouid=107480654748844323422&rtpof=true&sd=true)

Novice - this data is for Novice division

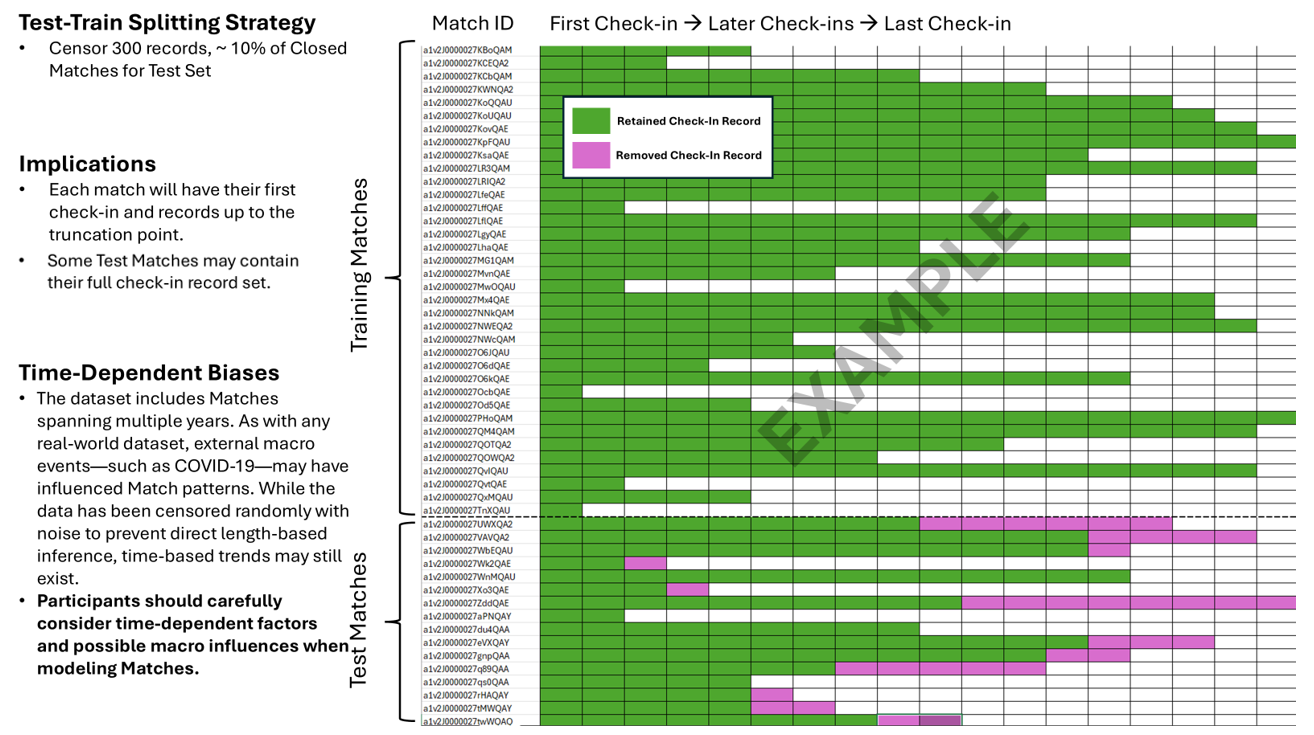
* 3,275 records - each row is a unique match between Big/Little
* 66 fields; 2 Fields excluded
  + Match Support Contact Notes (open text field), and
  + Complete Date (date of Match Support Contact Notes)

Training- this data is for Undergraduate and Graduate divisions

* 39,345 records - each row is a unique match - note instance combination
  + 3275 unique matches
  + Note instances collected over time
* 68 fields

Test - Truncated - this data is for Undergraduate and Graduate divisions

* 2,566 records - each row is a unique match - not instance combination
  + 300 unique matches, ~ 10% of the closed matches
  + Instances of the Match Support Call Notes have been truncated (see below)
* 63 fields; Fields removed include:
  + Stage (all matches are Closed in Test set)
  + Closure Reason
  + Closure Details
  + Match Closure Meeting Date
  + Match Length
* Truncation Scheme
  + Truncation of Match Support Contact Notes was necessary to build a predictive model for Match Length
  + A percentage of the most recent note instances have been deleted for most records
  + A schematic of the truncation process is provided here.



Additional Data Source

MinneMUDAC 2025 participants are encouraged to use additional data sources. The data must be free to use and publicly available. Teams are not allowed to use proprietary/private data.