**Introduction**

I do not have a particular business problem for this project but I want to work with spatiotemporal data as I am very interested in working with such datasets in team sports analytics. The role of deep learning in sports analytics has been massively expanded by the technological breakthrough which allows organizations to efficiently (cost-wise) and accurately collect near continuous data on players’ locations during a game. There are two primary end users for this sort of data: 1) Gambling Books & 2) Sport Organizations.

The rise in online gambling companies has obviously made sports betting more accessible and popular. But perhaps the bigger change from this deregulation trend is the rise in in-game or live betting. Judging by the amount of advertising showcasing these live betting opportunities, it is safe to say that it is a big part of gambling companies’ business plans[[1]](#endnote-1). However, setting the odds for these bets can be very difficult for sports like soccer, basketball and hockey because of the dynamic/fluid pace of play. Having a model that can take a sequence of game states, each defined by the location of all players (& ball/puck), to predict future game states in the short-term can be funneled into less reactive models for predicting medium- and/or long-term outcomes.

For sport organizations, the potential opportunities are much more robust:

* Identifying general strategies that are successful
* Identifying/Predicting opponent strategies
* Optimal personnel for different approaches
* Optimal training and practice regiments
* Post-game analysis of player and team efficiencies/deficiencies

This is especially true for hockey analytics which has a unique amount of complexity, even compared to other fluid-action team sports like basketball and soccer. This is due to, among others, the following factors:

* Fluidity of Offensive Possession
* Rapidly Changing Personnel
* Unique Role of Goalie

**Possible Datasets**

*Hockey Play-by-Play from BigDataBall.com*

The data covers the regular season and playoffs for the seasons which began in the calendar years of 2017 through 2023.

* A total of 9,075 games.
* With an average of 314 plays / game.
  + 209 plays at a minimum
  + 767 plays at a maximum
* A total of >2.85mm plays: 19 different types & 172 subtypes
  + Led by Faceoffs (18%), Shot (18%), & Hit (14.5%)

Pros:

* Target Sport
* Lots of observations
* Clearly identified events
* Player-level detail
* Could be useful complement to raw spatial data
* Not Deep Learning Model Maybe?
  + Or Feature Engineering to create sequences.

Cons:

* Messy dataset
* Uneven frequencies
* Lacks a lot of spatial information
* Likely requires raw spatial data to be valuable
* Unclear what target variable could/should be

*Race Tracking from Kaggle*

Pros:

* Multiple Modeling Targets:
  + Trajectory/Path
  + Results (Time to Finish)
  + Results (Placement)
* Multiple Modeling Options/Approaches
* Clean dataset w/ Lots of Feature Engineering Possibilities
* Interaction Between Horses (although not as direct as team sport examples)
* Clear end results
* Possible to engineer horse pre-race attributes
* Unique External Interactions
  + Jockey Information
  + Race Course/Distance
  + Weather Info

Cons:

* Only one result
* Enough Observations?
* Not a team sport & no puck/ball to track
* Complicated space limitations
* At first glance, not much strategic nuance for race path
* Sequence length is capped
* Live-betting angle does not apply

**Resources**

*Datasets*

[Kaggle-NYRA Data](https://www.kaggle.com/competitions/big-data-derby-2022)

[Kaggle-Kentucky Derby Data](https://www.kaggle.com/datasets/sadhliroomyprime/kentucky-derby-tracking)

[Big Data Ball\_NHL Play-by-Play](https://www.bigdataball.com/datasets/nhl-data/historical/)

[Sportsradar NHL API](https://developer.sportradar.com/ice-hockey/reference/nhl-overview)

*Spatiotemporal Machine Learning*

[Generating Long-Term Trajectories Using Deep Hierarchical Networks](https://arxiv.org/abs/1706.07138)

[A Graph Attention Based-Approach for Trajectory Predictions in Multi-Agent Sports Games](https://arxiv.org/abs/2012.10531)

*Spatiotemporal, ML & Sports*

[A Spatial Perspective on Sports Analytics](https://www.amazon.com/Spatial-Perspective-Sports-Analytics-analyzing-ebook/dp/B09BDGKB1X/ref=sr_1_2?crid=1UID6TTOFU8QA&dib=eyJ2IjoiMSJ9.dhNJbNd-2sJnkg6IhyycU89r9RV-EciJih_jor8wtEbj3d6SHdsSCmjRRRZRHRdbv12TNAP-YqfzMqRJcydN6eeVXUI2KjmsI3VsuL7QM7K_D-531YORFunQpy3Y2rnuuZ0_A21JlYt_Jw3h7ptn3Q.TChYG5j9iXZtHNwmGFEnnZE9evKzbWoN75ovRZUATPU&dib_tag=se&keywords=sports+spatial+data&qid=1750792180&sprefix=sports+spatial+data%2Caps%2C96&sr=8-2)

*Hockey Analytics/Strategy*

[Markov Game Model](https://www.researchgate.net/publication/315634981_A_Markov_Game_model_for_valuing_actions_locations_and_team_performance_in_ice_hockey)

[Hockey Analytics-Game Changing](https://www.amazon.com/Hockey-Analytics-Game-Changing-Stephen-Shea/dp/1977533493/ref=pd_bxgy_d_sccl_1/145-8979297-5255560?pd_rd_w=ZJaUY&content-id=amzn1.sym.dcf559c6-d374-405e-a13e-133e852d81e1&pf_rd_p=dcf559c6-d374-405e-a13e-133e852d81e1&pf_rd_r=1DGK5X2G)

[Tape To Space: Hockey Tactics](https://www.amazon.com/dp/1790480493/ref=mes-dp?_encoding=UTF8&pd_rd_w=VR001&content-id=amzn1.sym.a8908360-3609-476b-8c64-1eef634998b7&pf_rd_p=a8908360-3609-476b-8c64-1eef634998b7&pf_rd_r=CQNVNDN0ZDJ0AGTSQ6WE&pd_rd_wg=XXeJM&pd_rd_r=9a47b774-c9ed-4b70-a151)

*Horseracing Analytics/Strategy*

[*https://royalsocietypublishing.org/doi/10.1098/rsbl.2011.1120*](https://royalsocietypublishing.org/doi/10.1098/rsbl.2011.1120)

*https://www.biorxiv.org/content/10.1101/2020.06.11.145797v1*

[*https://pubmed.ncbi.nlm.nih.gov/33264298/*](https://pubmed.ncbi.nlm.nih.gov/33264298/)

*Gambling*

https://help.draftkings.com/hc/en-us/articles/4405230615699-What-is-a-live-bet-US

<https://www.risk.inc/blog/how-sportsbooks-make-money---a-look-inside-the-online-betting-business>

1. [Per Google: recent quotes from DraftKings CEO](https://www.google.com/search?client=safari&rls=en&q=draftkings+ceo+on+live+betting&ie=UTF-8&oe=UTF-8) [↑](#endnote-ref-1)