**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

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

**Possible Datasets**

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

Pros:

* Target Sport
* Lots of observations
* Clearly identified events
* Player-level detail

Cons:

* Messy dataset
* Uneven frequencies
* Lacks a lot of spatial information
* Unclear what target variable should be

*Race Tracking from Kaggle*

Pros:

* Clean dataset
* Interaction Between Horses
* Clear end results
* Possible to engineer horse pre-race attributes

Cons:

* Only one end result
* 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*

*Spatiotemporal Machine Learning*

*Hockey Analytics/Strategy*

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=1DGK5X2G954DJ7ZKFP9F&pd\_rd\_wg=MT2n5&pd\_rd\_r=a0c2868d-2ae5-4d79-be67-a530a9a4dafd&pd\_rd\_i=1977533493&psc=1

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-a151-1a6f3a8c30dd

*Horseracing Analytics/Strategy*

*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)