

Tangible Benefits with Data Science



Objective

Data Science encompasses many diverse tools to help with decision making. It is often easy to get sucked into the ecosystem that is accuracy scores when the end goal is to deliver value, not be the most accurate.



Finding a Problem

The most direct way to create value is to solve a problem directly involving money. For an individual the most accessible problems are the stock market and betting. I've chosen betting for variety.



Tennis

FREQUENTCY

Tennis has no off season this is important if generating income.

SIMPLICITY

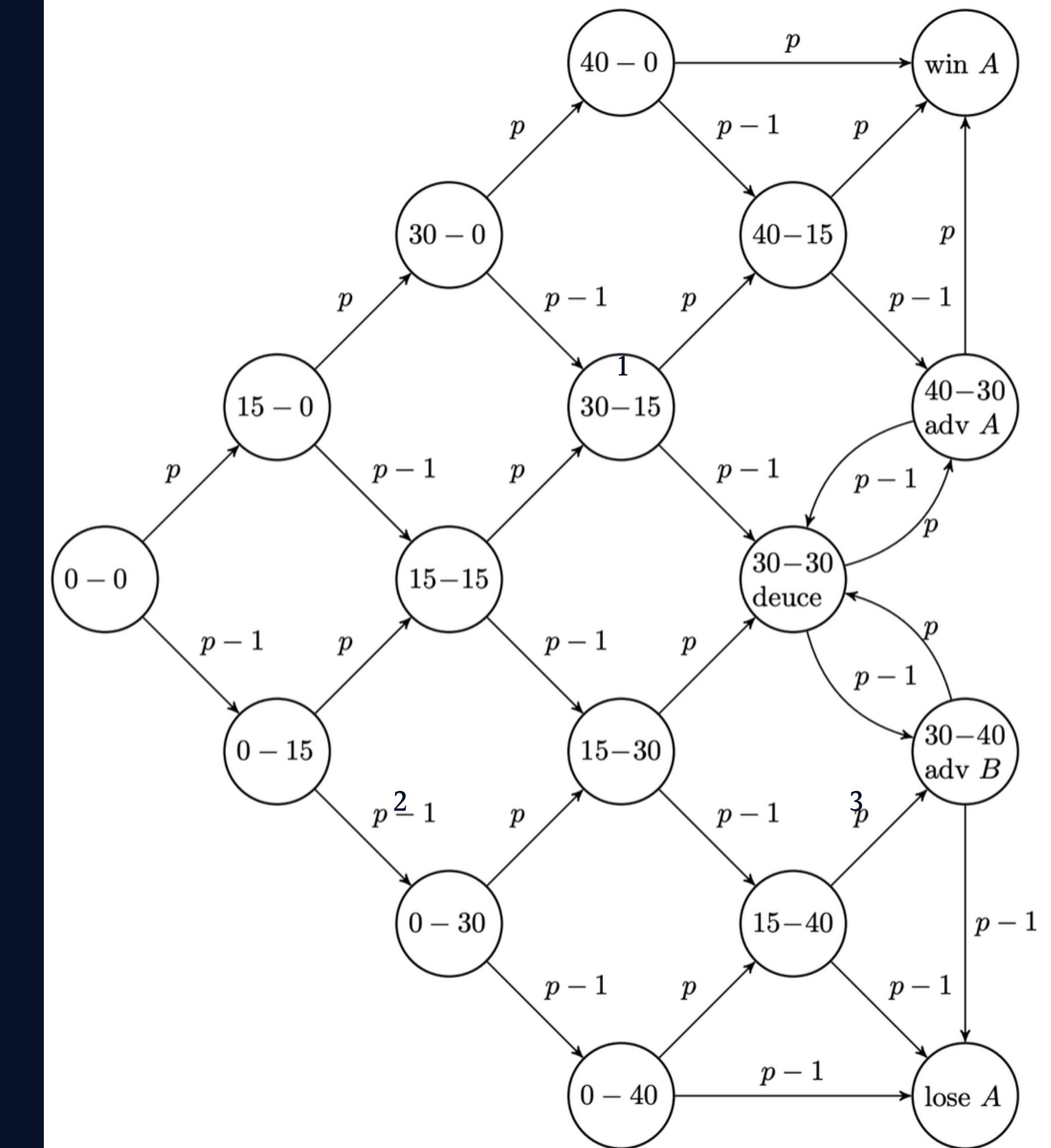
There are only two players to account for and a finite amount of outcomes.

AVAIABILITY

A large proportion of tennis matches played are about to be bet on.

This is Tennis

In essence it is a series of games to five, but you must win by two.



What are our features?

The main features are the players stats averages at the time of the match for the purposes of modeling. Going forward simple historical averages will suffice.

The Accuracy Trap

Example

Yoshihito Nishioka vs Novak Djokovic

Moneyline

+1000 to -10000

Meaning a \$10000 wager wins \$100

The model is very certain about certain matches, however we can't bet on these matches because sports books are certain about these matches as well.

Embrace Uncertainty

Betting odds are determined by sportsbook models. There is little profit to be had in high probability predictions.

The simple solution is to only look at semi confident predictions, this is somewhat counterintuitive.

Target Range

I am looking at .22-.30 and .7-.78 only.

Approximately 18% of test set predictions fall in this range

I tested this range using the 20 most recent matches and a betting strategy.



Betting Strategy

Start with your \$100 bankroll

Bet 20% of bankroll each match



RESULTS

\$100 —————> \$452.18

the matches were played 10-07-19 to 10-28-19

The model was wrong on 35% of the 20 matches

Questions?