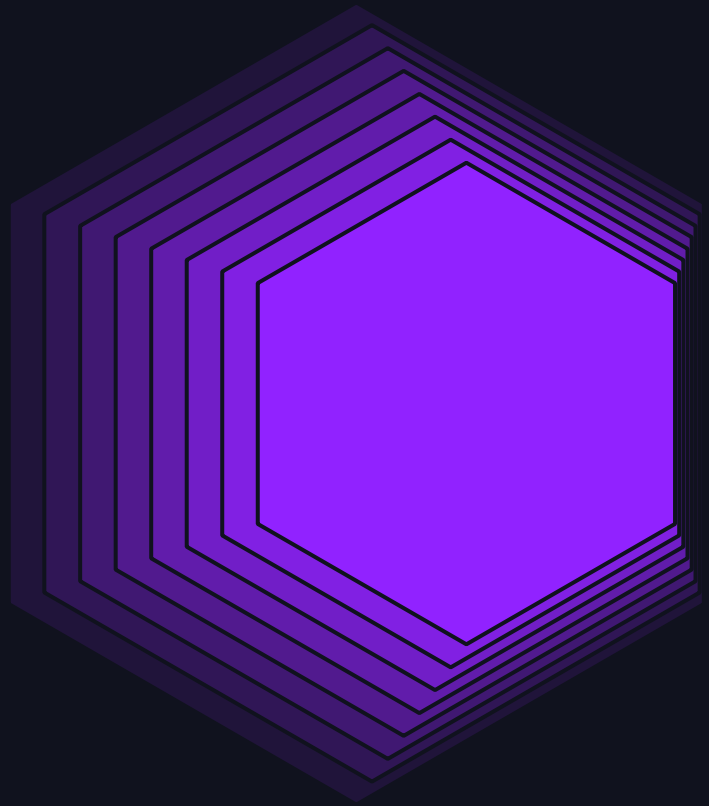
DATA+AI SUMMIT BY databricks

SIMULATING THE SUPERBOWL

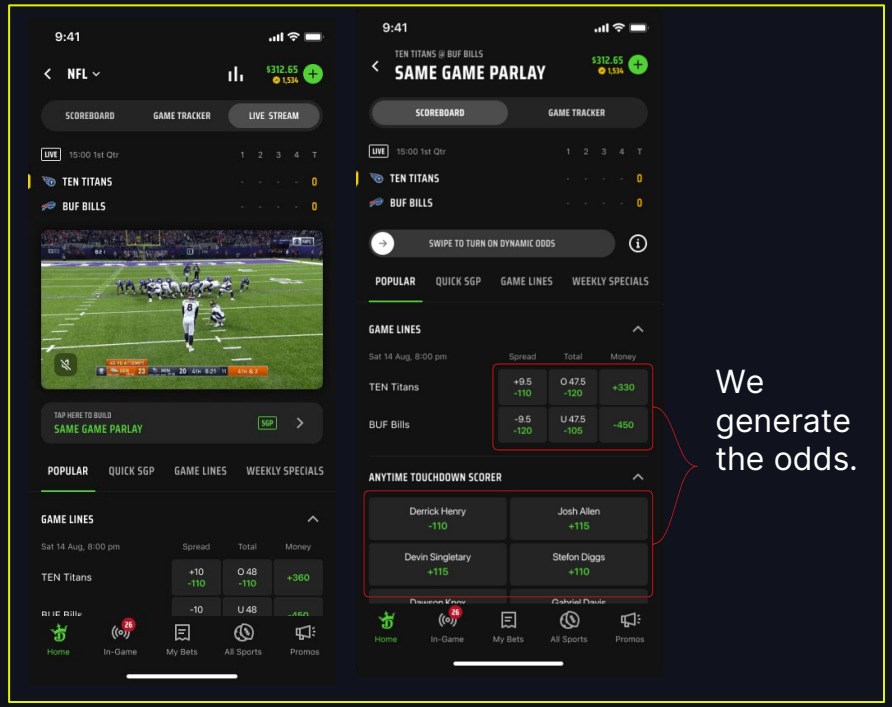
REAL-TIME ML TO PREDICT THE NFL



Rohan Shanker
2024-06-13

Who We Are

- DraftKings is a digital sports entertainment company, with multiple verticals, including a sportsbook.
- Sports Intelligence is responsible for producing sports betting content, primarily for the sportsbook.



Challenges To Address

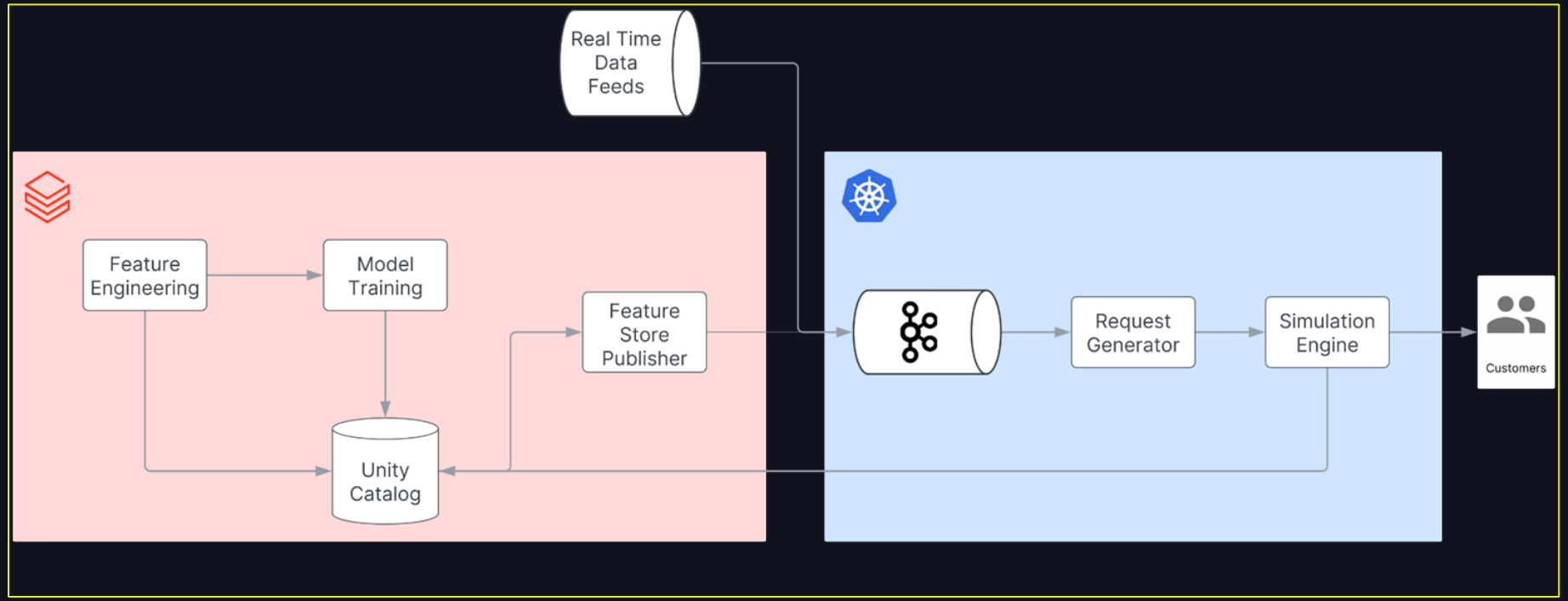
Accuracy and Pricing Performance

- The model must generate ~300 betting markets, which must be accurate (based on financial metrics, e.g. hold percentage).
- Incoming data to the model must be timely. Events in a game must be processed as quickly as possible. Features, such as ratings, must include all completed games.

Production Grade Service

- The model must have a high uptime.
- The model has to handle ~1200 requests during the course of a single NFL game, responding with a P95 latency of 2s.

Architecture Overview



Handling An Incoming Event

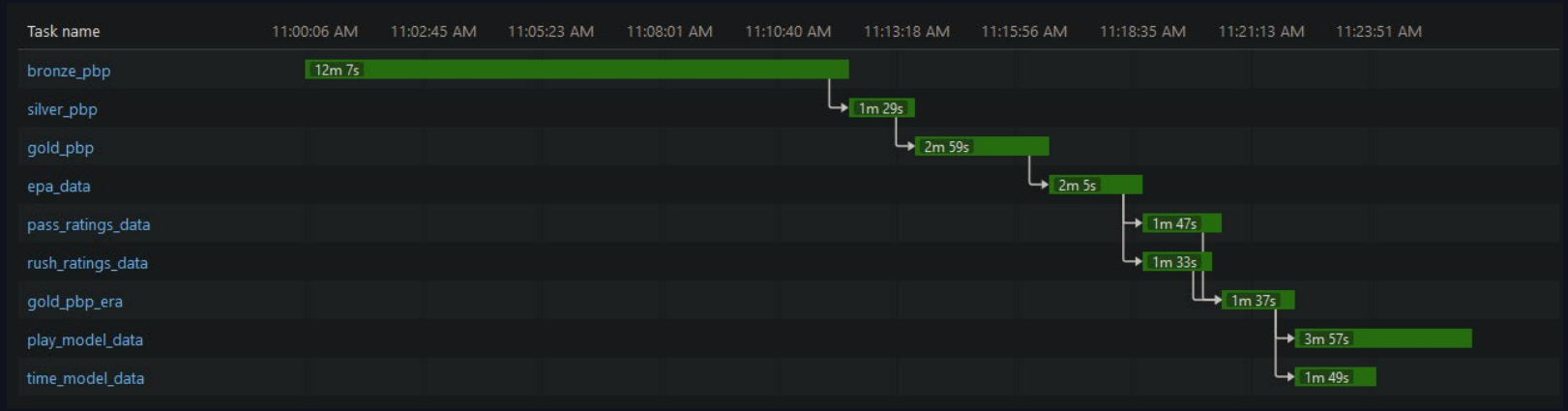

```
{
          "category": "touchdown",
          "description": "C.McCaffrey for 21
yards, TOUCHDOWN.",
          "end_location": {
                    "alias": "KC",
                    "yardline": 0
          },
          "result": "touchdown",
          "sequence": 2,
          "start_location": {
                    "alias": "KC",
                    "yardline": 0
          },
          "yards": 21
}
```

Feature Pipelines

100s of features for our data points

Snippet of the same play

	game_id	game_date	clock_end_event	touchdown	air_yards	yards_after_catch	ydstogo	blitz	complete_pass
1	2023_22_SF_KC	2024-02-11T23:30:00.000+00:...	"4:23"	1	-5	26	10	0	1

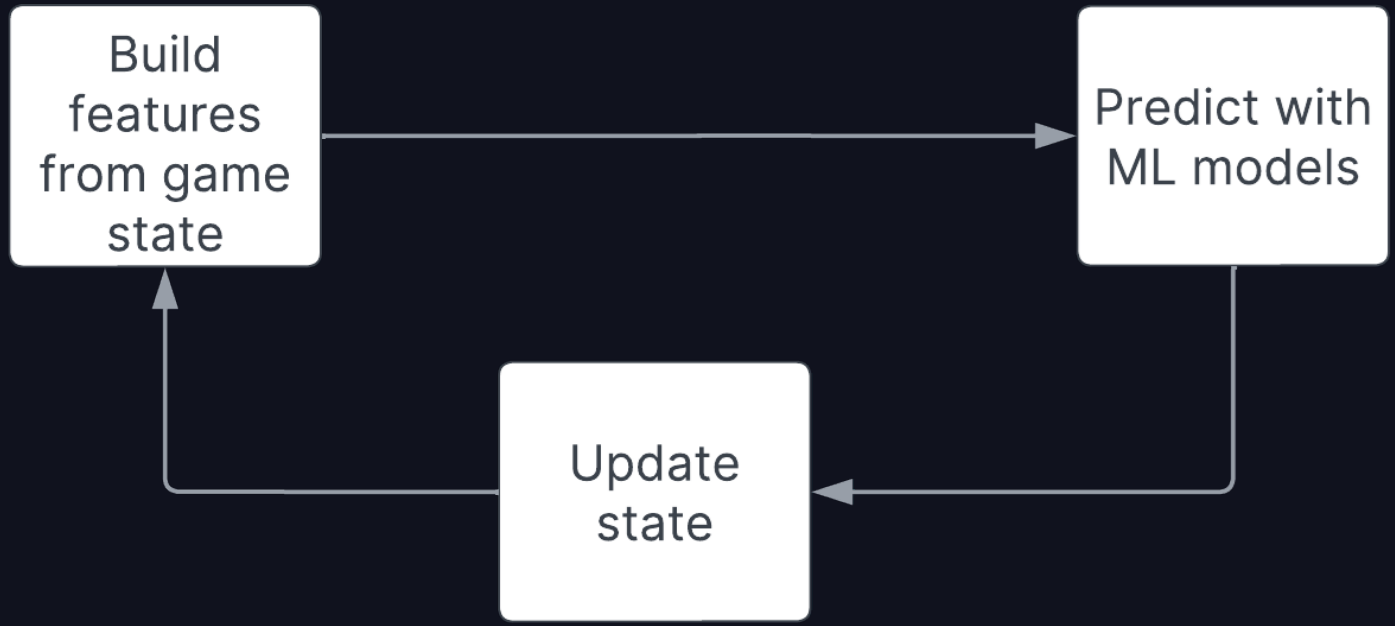


Model Training

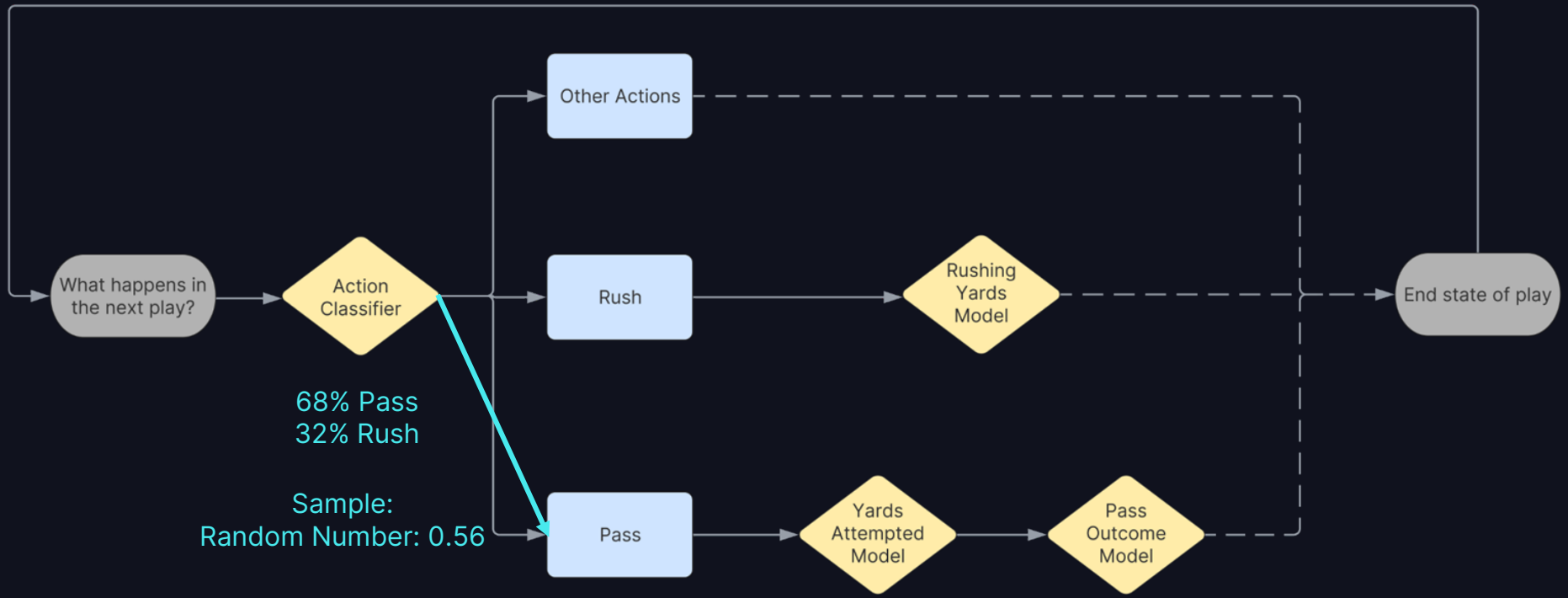
- Individual ML models used to predict probabilities for various events in a game.
- For example:
  - Action Classifier - what action (rush, pass, punt etc.) will the offense choose at a point in the game?
  - Pass Outcome Model - given a play is a pass, what's the probability it is successful?

```
import mlflow
model = mlflow.sklearn.load_model("models:/<catalog>.football.regular_action_model")

model.predict_proba(<features>)

[0.6757097 , 0.32429025] # For McCaffrey TD, 68% pass, 32% rush
```

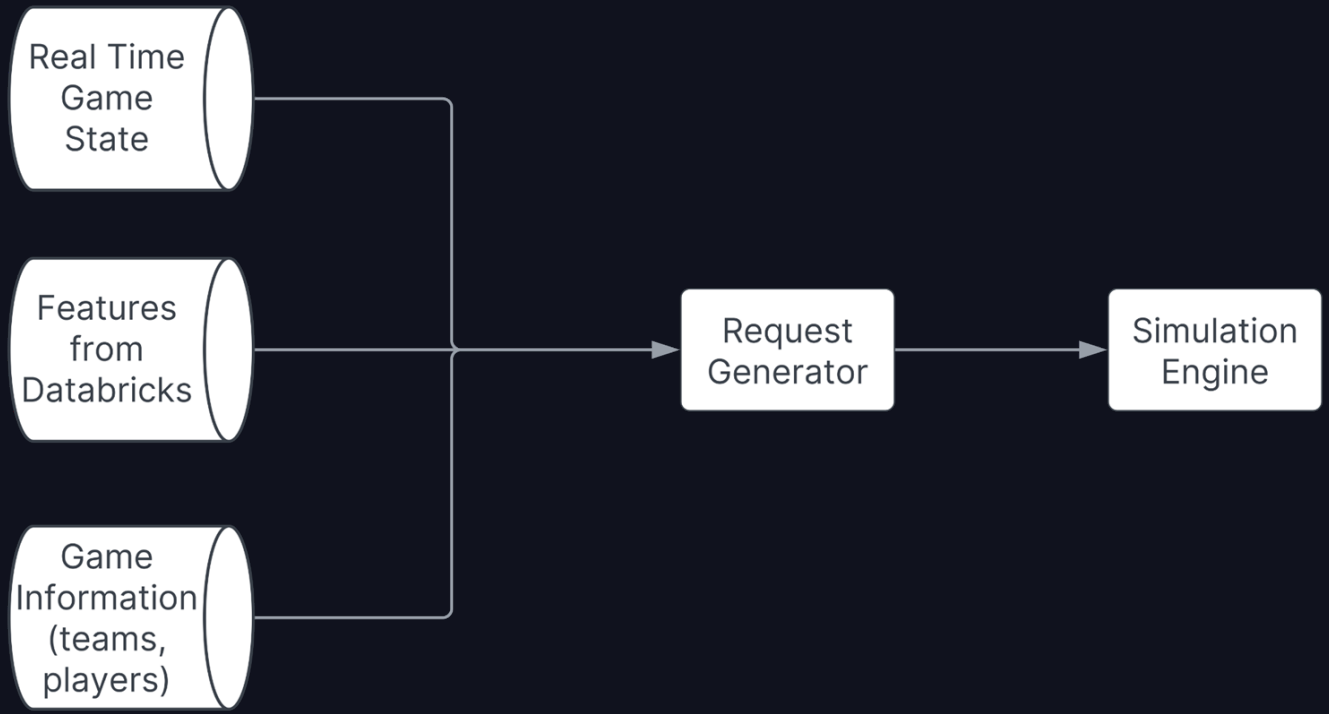
Monte Carlo Simulation



ML Models In A Monte Carlo Simulation



Bringing It All Together



Results

Production Reliability

- DraftKings have integrated Databricks over the last 18 months.
- This system was in production for the last NFL season, meeting latency and throughput SLAs.

Financial Results

- The system was accurate and performant.
- In New York this system was able to help deliver a 16% hold for the Superbowl.

Wrap Up

- We discussed:
  - How to use Monte Carlo simulations to predict sports.
  - Chaining ML models together within a simulation.
  - DraftKings' use of Databricks for feature and model building pipelines.
  - How DraftKings build production grade ML services.

DATA+AI SUMMIT BY databricks

Thank you!

Resources

Modeling Football Blog Post

DK Engineering Medium