Project Overview

This project explores generative modeling and outcome prediction of NFL plays using tracking data.

Directory Structure

Data/
- Contains the data used for this project (not included in GitHub due to size/privacy).
- `preprocess.py`: Loads raw CSVs, extracts frame-level data, and normalizes player coordinates.
- `run_preprocessing.py`: Pipeline script to run preprocessing and split data into train/val/test sets.
- `split_batches.py`: Splits data into smaller memory-efficient batches.

Models/
- `vae.py`: Variational Autoencoder (VAE) model for generating player trajectories.
- `outcome_predictor.py`: LSTM-based model for predicting play outcomes (e.g., yards gained).

Train/
- `train_vae.py`: Trains the VAE for up to 50 epochs and saves `trajectory_vae.pt`.
- `train_predictor.py`: Trains the outcome predictor for up to 50 epochs and saves `outcome_predictor.pt`.

Test/
- `experiment.py`: Evaluates models using metrics, visualizes real/generated plays, and generates `predictions.csv`.
- `plot_predictions.py`: Creates scatter plots from `predictions.csv`.