`readme.md`

Data Sources

The data are pulled from the nflfastR package.

Project Structure

- `/data` has the .RData files that can loaded in lieu of downloading the data.
- `/forecast` has the series to predict games for the 2024 season.
- `/series` has the data from the 2020-2023 NFL seasons to train the model.
- `/train` has the files to train the various models and to forecast games for the 2024 season.

Running the scripts

1. Run the parse.R file. This file will download and parse the drive data. It will take a while to download, so don't stop it if it seems to hang. (Alternatively, you can load `data/all_games.RData` into the R environment which has all of the downloaded data.) It will output files into `/data` (the .RData files), `/series` (the series for the 2020-2023 seasons that will be trained on), and `/forecast` (the data for the 2023 and 2024 seasons needed to predict the 2024 season).
2. Open the terminal and navigate to the `/train` directory.
3. To assess each model, run build.py in the terminal. Its interface is `python build.py <model> <td>|<fg> <off>|<def>`. The options for `<model>` are:
    - `lsr`: Least-Squares Regression
    - `mpl`: Multi-Layer Perceptron
    - `rnn`: Recurrent Neural Network
    - `lstm`: Long Short-Term Memory
    - `trans`: Transformer
    - `holt`: Holt Smoothing (takes a while and does poorly)
4. To forecast games for the 2024 NFL season, make sure you're still in the `/train` directory. Then run `python forecast.py <model> <away?> <home?>`, using one of the models above. If `<away>` and `<home>` are provided, the script will print all the data associated with that game if it exists.

These models were run with Python 3.12.4 and R 4.4.2.