NFL Division Winner Predictor
Evan Nuss
Level of Interest: 8/10

1. **Description**

   As I've paid more attention to the many different major sports leagues around the world, I've been amazed at the parity in the NFL. It's possible for a team to do absolutely horribly one year, make it to the Conference Championship the next year, then go back to having a losing record the year after that. That kind of sharp fluctuation is pretty common in the NFL, and I want to see if the winner of any given NFL division can be determined based on all available information about the team at the beginning of the season. This may be impossible (especially since minds much smarter than me have probably tried), but I thought it was worth a shot!

2. **Data Features**

   There are hundreds of potential features that could be used. However, I believe the most important features of the dataset will be: All-Pro players who have remained on the team, offseason season-ending injuries to starters, win percentage from the previous year, total win percentage under the current general manager, and total win percentage under current head coach. I also strongly believe that these features will change as we discover that certain ones don't actually contribute much while others contribute more than we initially thought. Some other potential features that could be employed would be: defensive rank from the previous season, percentage of defensive starters returning for the season, offensive rank from the previous season, and percentage of offensive starters returning for the season.

   An example instance from a dataset using the 5 features first mentioned above would look like the following:

Team	Returning all-pro players	Injuries	Win % prev season	Win % current GM	Win% current head coach	Target: Division winners
2019 Baltimore Ravens	3	1	62.5%	N/A. GM's first year.	62%	Yes

3. **Gathering Data**

   All information for the proposed features is readily available, but not all in one place. The win % under the current GM would need to be compiled by averaging the win percentage for each individual season as it did not seem readily available anywhere. The win % under the current coach was easily able to be googled using the coach's name and team along with "record". The injuries, all-pro players, and records for each season are found on pro-football-reference.com. However, the injury reports do not say whether the player was a starter, so that would have to be cross-referenced with the roster. Since I don't have any experience with web scraping or anything like it, the data would be gathered manually.