Marissa841 / **phase_5_project**   Public

☆ **0** stars   ⑂ **0** forks

| ☆ Star | ◉ Unwatch ⌄ |

<> **Code**   ⊙ Issues   ⑂ Pull requests   ▷ Actions   ⊞ Projects   📖 Wiki   ⚠ Security   📈 Insights   ⚙ Settings

⑂ **main** ⌄   ⋯

🟣 **Marissa841** Update readme   ⋯   2 minutes ago   ⟳ 18

View code

≡ **README.md**   ✎

Business Case

Head injuries and concussions have become a serious issue in professional sports, affecting the health of players and the winning potential of teams. The goal of this project is to predict which NHL players are more likely to suffer head injuries based on their past performance and other relevant information. This analysis is targeted towards NHL teams and team managers to help them take proactive measures to prevent head injuries and minimize the impact of such injuries on the team's performance.

Data

The data used in this project is from the Eliteprospect_scraper package which contains data of NHL players and https://nhlinjuryviz.blogspot.com/2015/11/nhl-injury-database.html which I corowsnvert to a csv file to view injuries over the past 20 years. The final dataset contains over 18,000 rows.

Methods

The methods used for this project were based on the OSEMN method and iterative modeling. For processing the data I cleaned the data, changed datatypes to integers/floats and made new features. This was class imbalance problem, where the target variable had the least amount of values:
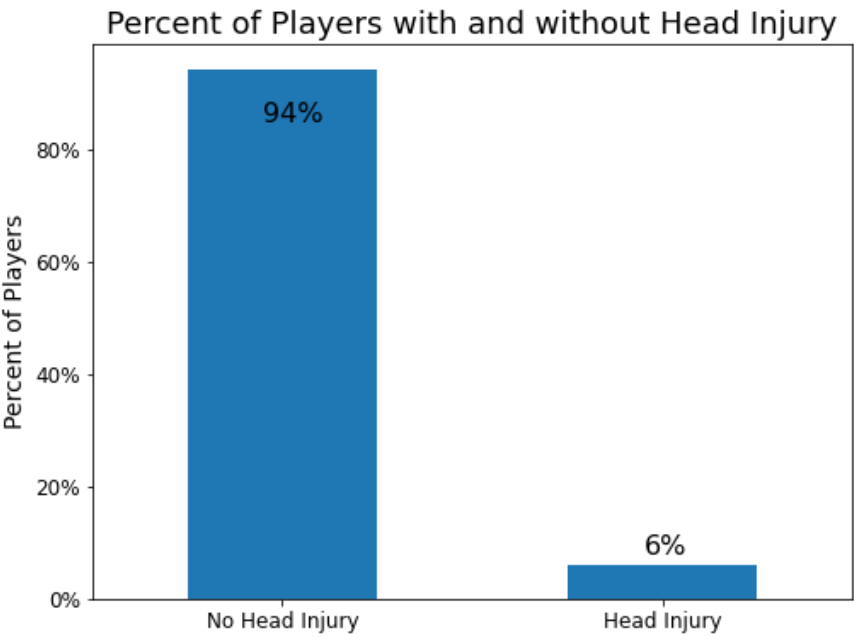
For the modeling portion, I used a function called model_helper to run through five machine learning modes: Logistic Regression, Decision Tree Classifier, Random Forest Classifer, Support Vector Machine (SVM), and Gradient boosting.

Findings

Model	F1 Score
Dummy Classifier	0.0
Logistic Regression	0.03
Decision Tree	0.16
Random Forest	0.17
Support Vector Machine (SVM)	0.13
Gradient Boosting	0.10
Random Forest (3 years)	**0.25**

Conclusion

Using machine learning, we are able to predict whether someone has a head injury at a higher rate, than by random guessing.

However, due to the randomness of head injuries, this is still a challenging problem.

Future Work

- More feature engineering

- Validating the target variable with another NHL injury dataset (injury data is not always accurately reported)

For more information

See the full analysis in the Jupyter Notebook or review this presentation. For additional info, my email is Marissabush.02@gmail.com

Repository structure

- data
- img
- main_notebook.ipynb
- data_notebook.ypynb
- phase_5_project_nontechnical.pdf
- model_helper.py
- README.md

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 99.5%   ● **Python** 0.5%