Works Cited

Barger, Colton. "Trying to Predict NHL Game Outcomes with ML and Why It's Difficult." *CodeX*, Medium, 6 June 2022, medium.com/codex/trying-to-predict-nhl-game-outcomes-with-ml-and-why-its-difficult-aaac4d2a690b. Accessed 26 Jan. 2023.

The article is about using machine learning to predict the outcomes of NHL (National Hockey League) games. The author collected data from 17 seasons and cleaned it up to extract features from each team's performance. They calculated rolling averages of different statistics over different timeframes and compared those to their opponents' same stats for each game. They then used three different ML models to predict the outcomes of games and found that logistic regression was the most accurate, with an accuracy of 57.3%. The article goes into detail about the data cleaning process and how the ML models were applied. Overall, this article is helpful for anyone interested in using machine learning to analyze sports data.

Bunker, Rory P., and Fadi Thabtah. "A machine learning framework for sport result prediction." *Applied computing and informatics* 15.1 (2019): 27-33.

The research paper talks about how machine learning can be used for predicting sports results. It says that machine learning has shown promising results in the domains of classification and prediction, and sports prediction is one of the expanding areas that necessitate good predictive accuracy. The authors identify the learning methodologies utilized, data sources, appropriate means of model evaluation, and specific challenges of predicting sport results. They propose a novel sport prediction framework using artificial neural networks as a learning strategy. The paper provides a critical analysis of the literature in machine learning, focusing on its application to sport prediction. It is hoped that this research will be informative and useful for future studies in this area.

Carbonell, Jaime G., Ryszard S. Michalski, and Tom M. Mitchell. "An overview of machine learning." *Machine learning* (1983): 3-23.

This chapter explains what machine learning is and what researchers focus on in this field. Machine learning is the study of learning processes using computer models. Researchers in machine learning have three primary areas of focus: developing and analyzing learning systems to improve performance in specific tasks, investigating and simulating human learning processes, and exploring different learning methods and algorithms that can be applied across different domains. A key objective of machine learning is to discover new ways of learning, develop better algorithms, understand their limitations, and create general techniques that can be applied in many different situations.

Gu, Wei, et al. "A game-predicting expert system using big data and machine learning." *Expert Systems with Applications* 130 (2019): 293-305.

The National Hockey League (NHL) is a lucrative sports organization with stakeholders interested in league competitiveness and team performance. To better predict NHL game outcomes and improve recruiting and salary decisions, an expert system is proposed using player and team data from various web sources. The system employs principal components analysis, nonparametric statistical analysis, a support vector machine (SVM), and an ensemble machine learning algorithm to predict game outcomes with a testing set accuracy exceeding 90%. The system is user-friendly, allows for easy input of data sources, and addresses challenges in predicting hockey game wins.

Gu, Wei, Thomas L. Saaty, and Rozann Whitaker. "Expert system for ice hockey game prediction: Data mining with human judgment." *International Journal of Information Technology & Decision Making* 15.04 (2016): 763-789.

This paper presents an expert system that predicts NHL game outcomes with a new method using data and judgments. The system combines the support vector machine with other factors to achieve an accuracy of 77.5% in predicting win-lose outcomes, which is the most accurate reported to date for hockey game prediction.

Hubáček, Ondřej, Gustav Šourek, and Filip Železný. "Exploiting sports-betting market using machine learning." *International Journal of Forecasting* 35.2 (2019): 783-796.

This research paper presents a new sports-betting system that uses machine learning to predict match outcomes and generate profits. Unlike previous models that only focus on predictive accuracy, this system also reduces correlation with bookmaker predictions, which allows for better profit generation. The system also uses convolutional neural networks to leverage a vast number of player-related statistics and adopts elements of modern portfolio theory to design a strategy for bet distribution. The experiments with NBA data from seasons 2007-2014 showed positive cumulative profits, making this method superior to alternative methods tested.

Keshtkar Langaroudi, Milad, and Mohammadreza Yamaghani. "Sports result prediction based on machine learning and computational intelligence approaches: A survey." *Journal of Advances in Computer Engineering and Technology* 5.1 (2019): 27-36.

The use of data mining techniques in sports has increased in recent years, allowing coaches and managers to predict game outcomes, evaluate player performance, identify talent, and assess game strategy. Predicting sports results is a challenging task due to various factors such as player skills, coaching strategy, and morale. This study reviews previous research on data mining systems used to predict sports results and evaluates the pros and cons of each system.

"March Machine Learning Mania 2021 - NCAAM." *Kaggle*, www.kaggle.com/competitions/ncaam-march-mania-2021/code. Accessed 19 Feb. 2023.

Morrell, Eric. "How to Use Machine Learning to Predict NCAA March Madness." *Analytics8*, 2 Feb. 2023, www.analytics8.com/blog/how-to-use-machine-learning-to-predict-ncaa-march-madness/. Accessed 2 Jan. 2023.

The author of this blog shares his experience of winning NCAA March Madness bracket pools by using a machine learning algorithm for predicting which team will win each game. The author explains that March Madness is an annual college basketball tournament, and every year millions of people participate by filling out their own bracket, attempting to predict all 63 games, and enter their selections in bracket pools. The author shares that machine learning can help people fill out their bracket by using data from prior NCAA games to create a model that predicts the winner of each game, and he covers all the steps needed to win March Madness using machine learning. The author explains that machine learning is a method of data analysis that automates analytical model building and is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention. The author uses a classification model to predict individual games for the NCAA Tournament, and he used a dataset provided by Kaggle.com that contains game-by-game data for every college basketball game since 2003. The author calculated each team's average of statistics for the previous 14 games, providing a snapshot of each team's current form for every game, which he uses for the predictions for the NCAA Tournament games. In addition to using statistics for each team, the author also created an Elo rating system for each team, which is a rating system updated on a game-by-game basis showing how powerful a team is at the time of each game. The author uses the combined dataset to train his machine learning model, using the Gradient Boosting Classifier algorithm to create a model to predict the winner of each game in the NCAA tournament. The author concludes that machine learning can provide an edge when filling out a bracket, and even though it is not a surefire way to win, it can get pretty close.

Patel, Savan. "Chapter 2 : SVM (Support Vector Machine) — Theory." *Machine Learning 101*, Medium, 3 May 2017, medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72. Accessed 20 Feb. 2023.

This chapter explains the theory behind Support Vector Machine (SVM), which is a discriminative classifier that separates labeled training data by an optimal hyperplane. SVM finds a line/hyperplane that separates the data into classes, and it uses kernels to transform data into a higher-dimensional space to create a clear separation. In real-world applications, perfect classifying millions of training datasets takes a long time, so SVM has tuning parameters, such as regularization, gamma, kernel, and margin, which can be adjusted to achieve a reasonable amount of accuracy in a shorter time.

Pischedda, Gianni. "Predicting NHL Match Outcomes with ML Models." *International Journal of Computer Applications*, www.researchgate.net/publication/284457066\_Predicting\_NHL\_Match\_Outcomes\_with\_ML\_Models. Accessed 15 Feb. 2023.

The paper explores the use of machine learning techniques to predict the outcome of NHL games using performance data. The study repeats an experiment conducted by a University of Ottawa team and compares the results with their own ML models. The data is also parsed to create a new dataset, and models are built to compare with the results of the original one. The paper proposes a framework for using this data in a practical application (betting) and uses three ML techniques: Decision Trees, Artificial Neural Networks, and ClusteR.

Richter, Chris, Martin O'Reilly, and Eamonn Delahunt. "Machine learning in sports science: challenges and opportunities." *Sports Biomechanics* (2021): 1-7.

This article discusses the impact of machine learning on the field of sports science. Machine learning has been applied to the optimization of data acquisition devices, extraction of information from acquired data, and processing of acquired data. Machine learning can facilitate objective decision-making with respect to rehabilitation practices and injury prevention. The article outlines the advantages and challenges of implementing machine learning models and emphasizes the importance of understanding key concepts such as data capture/processing, feature selection, modeling, and evaluation.

Valero, C. Soto. "Predicting Win-Loss outcomes in MLB regular season games–A comparative study using data mining methods." *International Journal of Computer Science in Sport* 15.2 (2016): 91-112.

This paper discusses the use of sabermetrics statistics to predict the outcome of Major League Baseball (MLB) games using four data mining methods. The study uses ten years of historical data to create a dataset and assesses the predictive capabilities of the classification and regression based methods. Results show that the classification scheme performs better than the regression scheme, and SVMs produce the best predictive results with nearly 60% prediction accuracy for each team. The evaluation is done using stratified 10-fold cross-validation, and the inherent difficulties of predicting sports outcomes are confirmed using two geometry or topology-based measures of data complexity.

Vellido, Alfredo, José David Martín-Guerrero, and Paulo JG Lisboa. "Making machine learning models interpretable." *ESANN*. Vol. 12. 2012.

Weiner, Josh. "Predicting the outcome of NBA games with Machine Learning." *Towards Data Science*, Medium, 5 Jan. 2021, towardsdatascience.com/predicting-the-outcome-of-nba-games-with-machine-learning-a810bb768f20. Accessed 15 Feb. 2023.

The article discusses a project aimed at predicting the outcome of NBA games using machine learning techniques. The project involves steps such as scraping relevant data, cleaning and processing the data, feature engineering, data analysis, and predictions. The article acknowledges previous research on the same topic and mentions that the best published model had a prediction accuracy of 74.1% for playoff outcomes, with most others achieving an upper bound between 66-72% accuracy. The article states that the project aimed to develop a model that could reach and beat a 67.9% accuracy, which is the average upset rate across the entire season in the NBA.

Zufelt, Nicholas. "Example of a Modeling Process." *GitHub*, Microsoft, 1 Oct. 2021, github.com/nzufelt/jupyter\_notebooks\_ML\_2018/blob/master/lessons/Example%20of%20a%20Modelling%20process.ipynb. Accessed 2 Feb. 2023.

The notebook is a complete solution to fitting a machine learning model to a dataset, with the aim of predicting whether or not a new loan applicant would default on their loan. The author first imported standard data libraries and loaded the data. The data set contains features such as student, balance, income, rent, and default, and has been modified for the exam. The data was cleaned up by making the binary columns 1 and 0 and by dealing with any missing data. The rent column is immediately suspect. The author also binarized the student and default columns by creating new columns.

---. "An Introduction to Jupyter Notebooks and the PyData Stack." *GitHub*, Microsoft, Oct. 2021, github.com/nzufelt/jupyter\_notebooks\_ML\_2018/blob/master/lessons/Intro\_to\_jupyter\_and\_pydata\_updated.ipynb. Accessed 3 Feb. 2023.

The article explains the basics of Jupyter Notebooks and how to use them. Jupyter Notebooks are interactive environments for Python, commonly used for data manipulation. The article explains the three types of cells - Markdown, Code, and Raw. The article also explains the two modes of working with notebooks, Edit mode and Command mode, and how to navigate between cells in each mode. The article also discusses the linearity of code, which is governed by the kernel attached to the notebook.