## horizontal line



Golf Course Analyzer

10.18.2022

**─**

Jared Ponmakha

CDS 403

# Introduction

As an avid golfer, I have favorite golf courses in my area and I have disliked courses. I am able to make these preferences through years of playing all the courses in the area. However, if I were to move to an area that I am not familiar with or I am simply vacationing to, I would have a hard time to find golf courses that would fit my preferences in a timely manner.

The purpose of this project is to use publicly available data to give golf course recommendations to a person based on their preferences on peak price, distance, difficulty and reviews.

This problem has not been solved before as golfing is a niche sector of luxury activity. There are also no free datasets that contain golf courses on a national scale or even on a state level, most datasets will be contained by the municipality or county.

# Goals

1. Compile datasets of courses in the DMV area as well as one other target location (New York City)
2. Use KNN and Decision Trees to give recommendations of golf courses based on user preferences.

# Lit Review

This project is original, however the idea behind it is common with some of the largest companies such as Amazon and Netflix. There is also a website called GolfNow that compiles courses in your area however it does give any preference on where to play. This was short of answering my question because utilizing Machine Learning in this problem will allow courses and companies to set pricing for courses more efficiently based on demand as well as increase consumer satisfaction by playing courses they are more likely to enjoy.

# Specifications

The current dataset has 86 entries total entries, with 58 being from the DMV and 28 from New York. I have also compiled addresses, Google reviews, and Slope Rating for each of the courses. I have not however, compiled pricing because it fluctuates based on the time of day. The datasets I did find also contained phone numbers and longitude and latitude coordinates of the course itself. I plan to take most interest in just the addresses, Google reviews, Slope Rating and price.

I plan to use K-nearest neighbor algorithm because the datasets of all golf courses costs hundreds of dollars to have access to. By not having to train the data with larger datasets beforehand, I can quickly add in other cities or areas if I were to expand the project.

This project would benefit immensely in the real world with predictiive pricing for customers, such as a discount in price to a course they are likely to play golf at. In this regard, utilizing decision trees and the CART algorithm would be advisable here.

## Setbacks

This project does not take into account other factors for why a golfer will want to go to a course. Limitations such as amenities and personal experiences previously are not accounted for in the project. Additionally, K-nearest neighbor may not be fully accurate in make a prediction based on the parameters given.

# Sources

<https://data.ny.gov/Recreation/State-Golf-Courses/2wzg-7b48>

<https://www.golfnow.com/>

<https://catalog.data.gov/dataset/golf-courses>