D&D Character Predictor

Introduction

In this report, a **D&D character dataset** was used to create a **D&D character predictor** that can be used to predict a character's class from its ability scores.

The **D&D character dataset** consists of approximately 2600 D&D characters obtained from GitHub (https://github.com/oganm/dnddata) with the permission of the repository creator.

The raw data was then filtered, wrangled, visualized, and two models were trained. A final **D&D character predictor** with an **accuracy** of ~**50%** was obtained.

Basic Introduction to Dungeons and Dragons

Dungeons and Dragons is a very popular role-playing game. Each player creates a character that has, among other attributes, a class, a race, and several ability scores. The twelve main classes in D&D are: barbarian, bard, cleric, druid, fighter, monk, paladin, ranger, rogue, sorcerer, wizard, and warlock. The six ability scores are: strength (str), dexterity (dex), constitution (con), intelligence (int), wisdom (wis), and charisma (cha). Ability scores are generally between 1 (very low) and 20 (very high), but there are some exceptions. Ability scores are chosen/assigned when the character is created and can be increased as a character levels up.

A player will usually assign his or her character's ability scores based on the class that they have selected for that character. For example, a fighter's highest ability score is usually strength, while a wizard's highest ability score is usually intelligence. Therefore, it should be possible to predict a character's class by examining its ability scores.

To simplify the dataset, the characters were limited to the main races in D&D, which are: dwarf, elf, halfling, human, dragonborn, gnome, half-elf, half-orc, and tiefling.

Data Analysis and Model Development

Create the Dataset

The raw dataset *dnd_chars_singleclass* was pulled directly from the GitHub repository. The raw dataset was then filtered down into the *character_data* by selecting the class, race, str, dex, con, int, wis, and cha columns.

Character Data

```
##       class       race str dex con int wis cha
## 1    Wizard      Human   8  12  20  20  14  10
## 2   Warlock      Kenku  12  20  16  14  14  21
## 3   Fighter        Orc  20  15  20   6  12  10
## 4   Paladin        Orc  20  13  20  10  12  14
## 5   Paladin Dragonborn  20   8  16   8   8  20
## 6 Barbarian  Aarakocra 103 101 103  99 100  99
```

Looking at the *character_data* dataset, we can see that it contains more classes and races than just the main ones listed above. We can also see that characters with classes or races other than the main ones occur less often.

Character Classes

```
## # A tibble: 19 x 2
##    class                n
##    <chr>            <int>
##  1 Artificer            9
##  2 Barbarian          233
##  3 Bard               192
##  4 Blood Hunter         6
##  5 Cleric             272
##  6 commoner             1
##  7 Crafting Commoner    1
##  8 Druid              175
##  9 Fighter            361
## 10 Gunslinger           1
## 11 Monk               180
## 12 Mystic               8
## 13 Paladin            228
## 14 Ranger             196
## 15 Revised Ranger      13
## 16 Rogue              297
## 17 Sorcerer           168
## 18 Warlock            153
## 19 Wizard             216
```

Character Races

```
## # A tibble: 19 x 2
##    race                     n
##    <chr>                <int>
##  1 ""                       3
##  2 1/2 Elf--Var.            1
##  3 Aarakocra               22
##  4 Aasimar                  7
##  5 Air Genasi               3
##  6 Animal Hybrid            3
##  7 Animal Hybrid (Bear)     1
##  8 Beast                    1
##  9 Bugbear                 16
## 10 Caprien (Goatfolk)       1
## 11 Centaur                 13
## 12 Changeling               8
## 13 Changeling 2.0           1
## 14 Child Of The Old Ones    1
## 15 "Copper Dragon "         1
## 16 Dark Elf                52
## 17 Deep Gnome              21
## 18 Dhampir                  1
## 19 Dhampire                 1
```

To facilitate the model, characters with unusual or uncommon races and classes were removed from the data set. Races in D&D can have sub-races, for example, a character can simply be a dwarf, or a hill dwarf, or a mountain dwarf, etc. Because of this, the dataset has been limited to the main races and their respective sub-races, though race is not used as a predictor for the model.

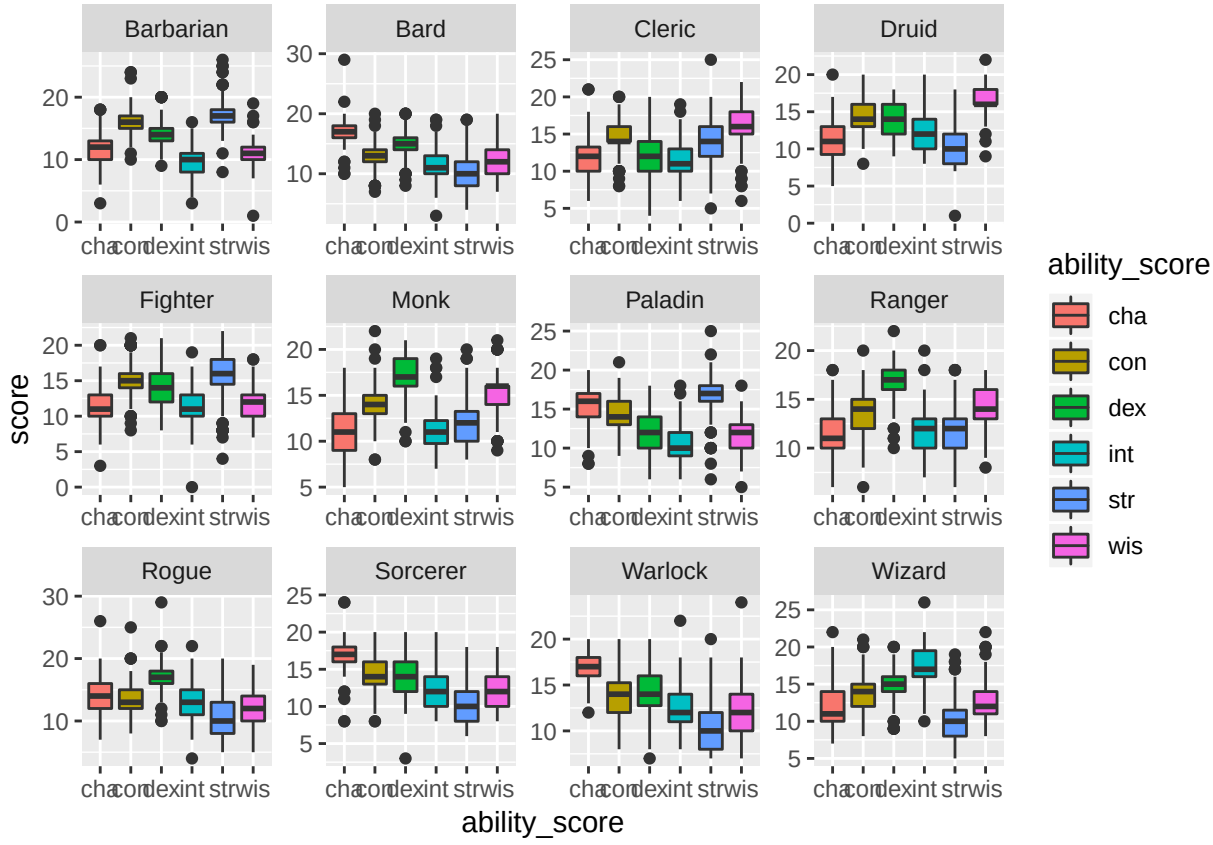
Character Data

```
##     class       race str dex con int wis cha
## 1  Wizard      Human   8  12  20  20  14  10
## 2 Paladin Dragonborn  20   8  16   8   8  20
## 3   Druid      Human   8  18  18  12  20   8
## 4  Wizard      Human  13  19  19  21  15  20
## 5  Wizard      Human  14  17  21  26  22  22
## 6  Wizard      Human  12  13  14  22  20  20
```

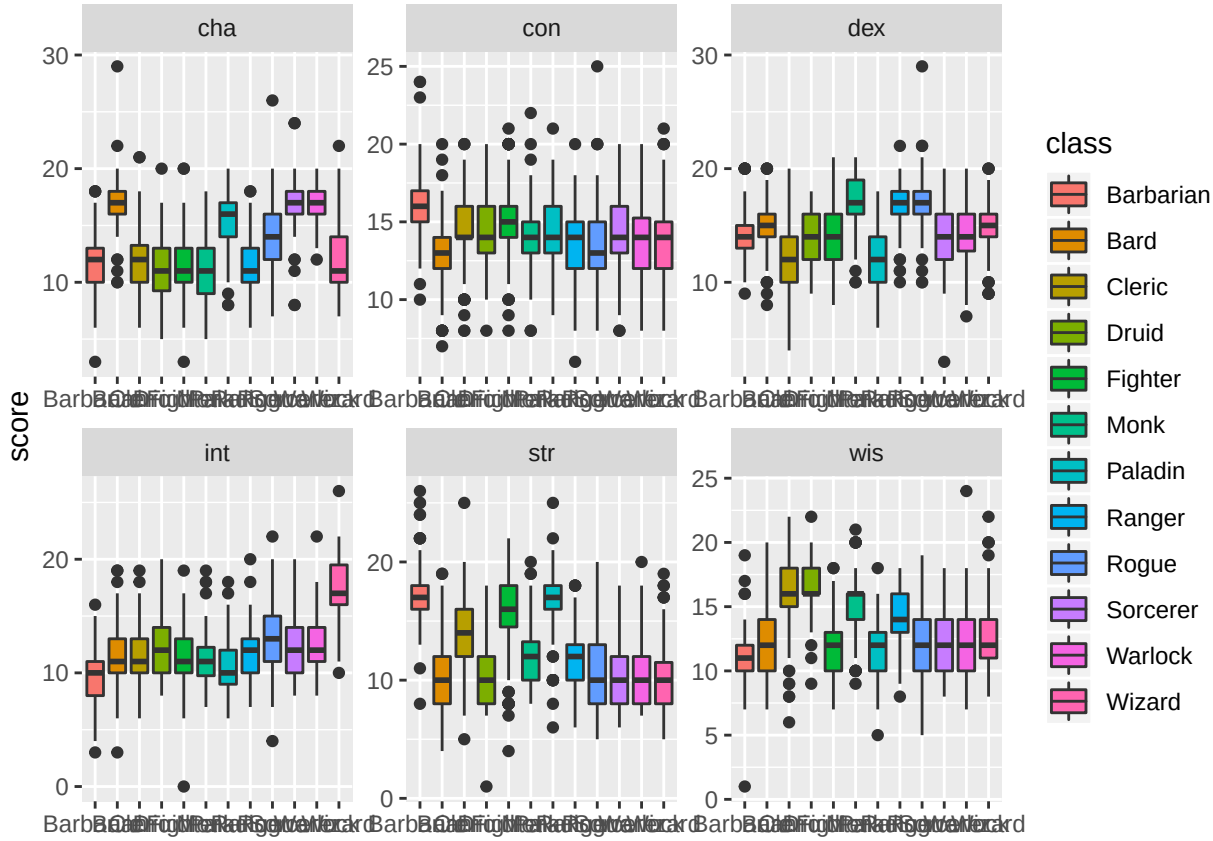
Cursory Data Visualizations and Analysis

Plotting the dataset by class shows that, as expected, different classes tend to favor different ability scores. Barbarians generally have high strength and low intelligence, rogues typically have high dexterity and low strength, and wizards tend to have low strength and high intelligence.

Ability Scores by Class



Class by Ability Scores



A First Model

The first model tested was the k-nearest neighbors (kNN) method tuned for k.

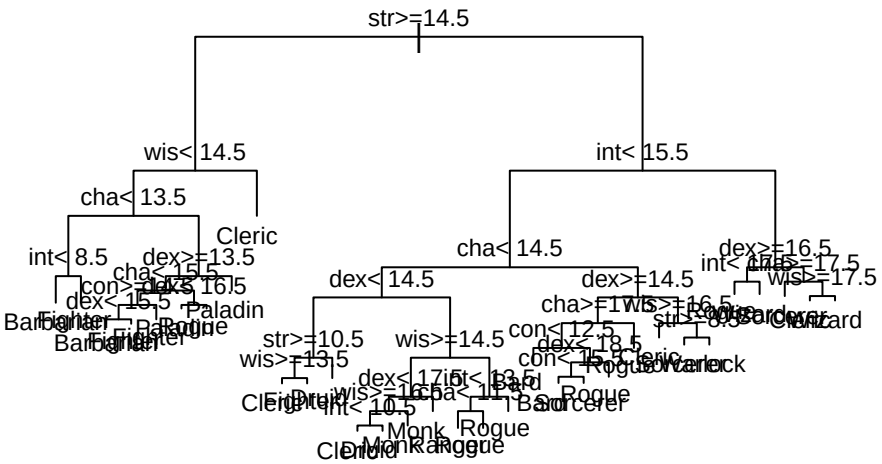
kNN Model

The accuracy of the **tuned kNN** model is 0.49.

A Second Model

The second model tested was a cp-optimized regression tree.

Optimized Regression Tree



The accuracy of the **optimized regression tree** model is 0.45, which is a loss in accuracy from the previous model. Also, this regression tree is very crowded, almost impossible to read, and has way too many nodes. The excess of nodes seems to stem from the fact that even within a class, the values of the ability scores vary greatly. To accommodate this, the ability score data was ranked.

Ranking the Ability Scores

A user defined function was created using the rank() function to rank a character's ability scores from highest (6) to lowest (1). The apply() function was used to apply the function to the entire dataset, and the resulting list was compiled back into a data frame for further analysis

Ranked Character Data

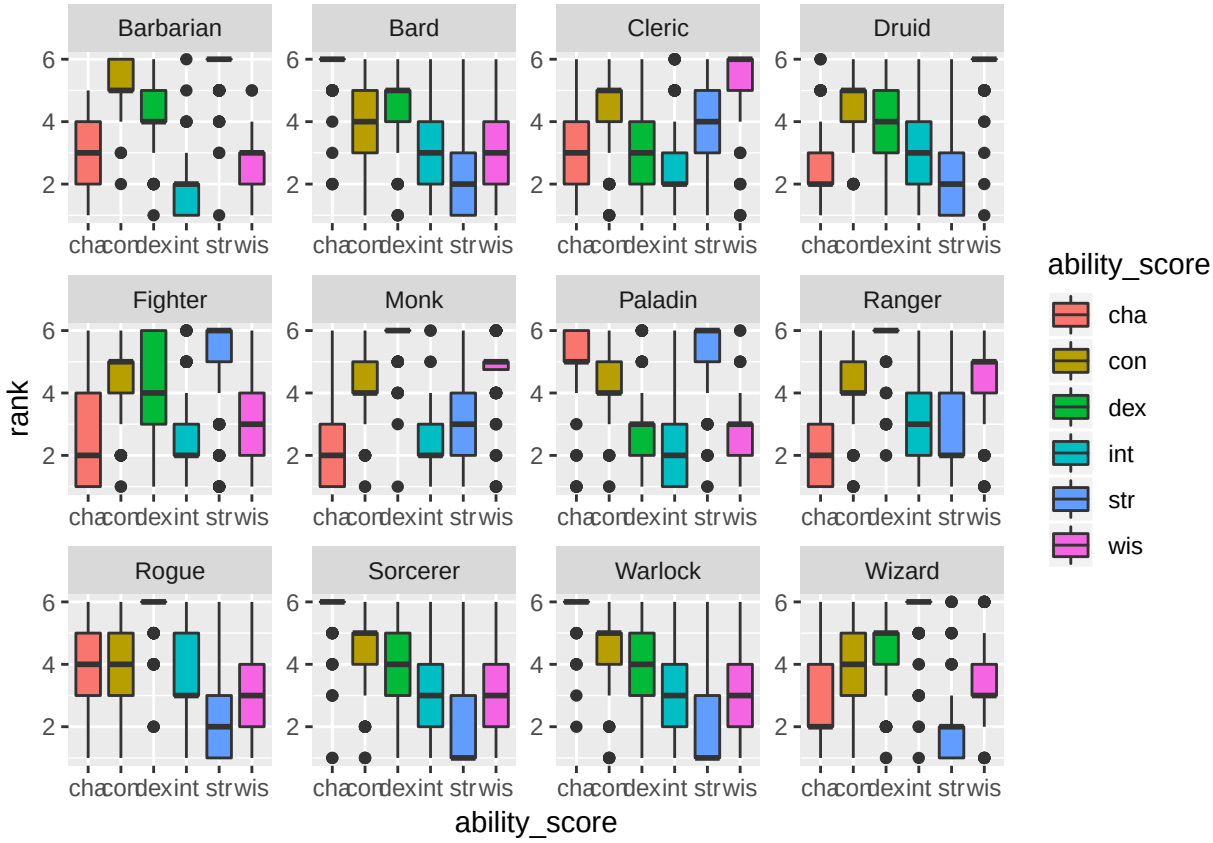
```
##     class       race str dex con int wis cha
## 1  Wizard      Human   1   3   6   6   4   2
## 2 Paladin Dragonborn   6   3   4   3   3   6
## 3   Druid      Human   2   5   5   3   6   2
## 4  Wizard      Human   1   4   4   6   2   5
## 5  Wizard      Human   1   2   3   6   5   5
## 6  Wizard      Human   1   2   3   6   5   5
```

The ranked data was revisualized and the models were re-run to see if there was any improvement.

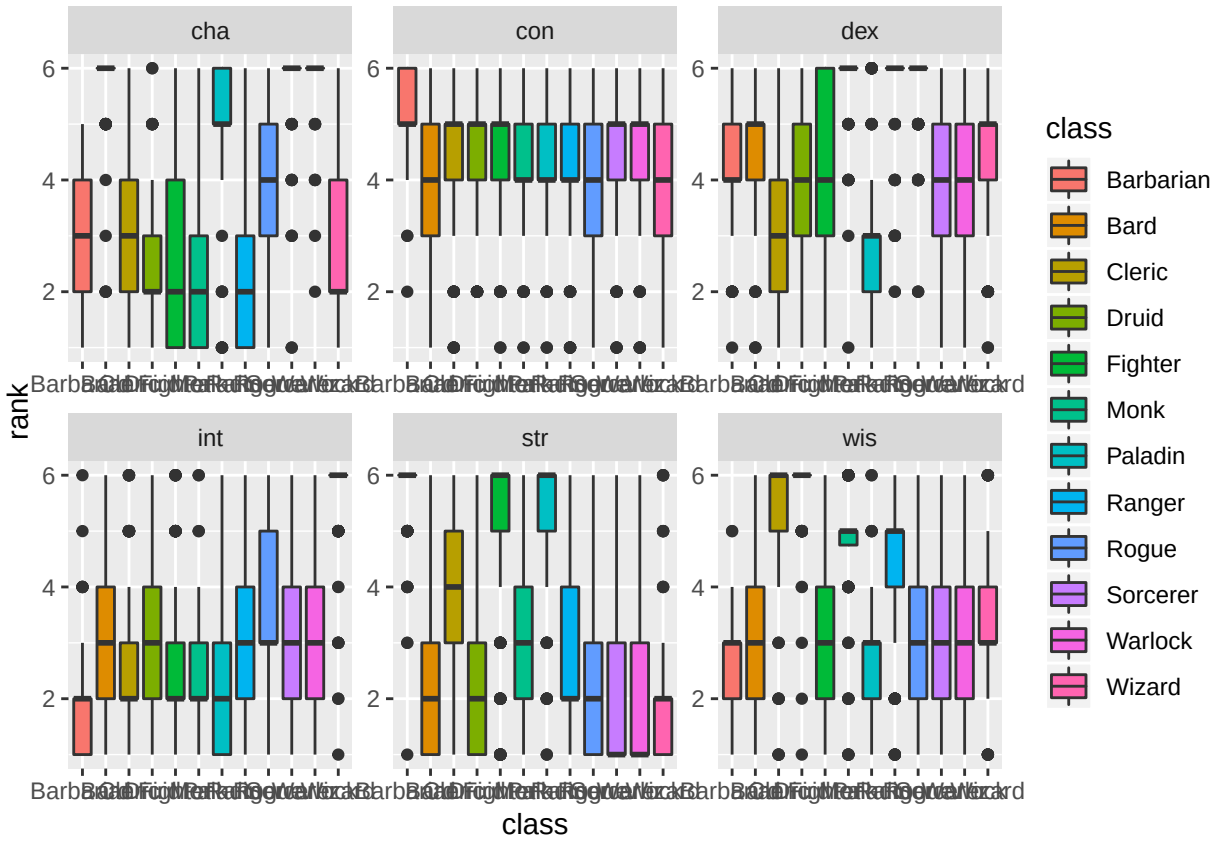
More Data Visualizations

As expected, the visualizations with the ranked ability scores look almost identical to the visualizations with the raw ability scores, aside from the scales.

Ranked Ability Scores by Class



Class by Ranked Ability Scores



A Third Model

The third model tested was the k-nearest neighbors (kNN) method tuned for k with the ranked ability scores.

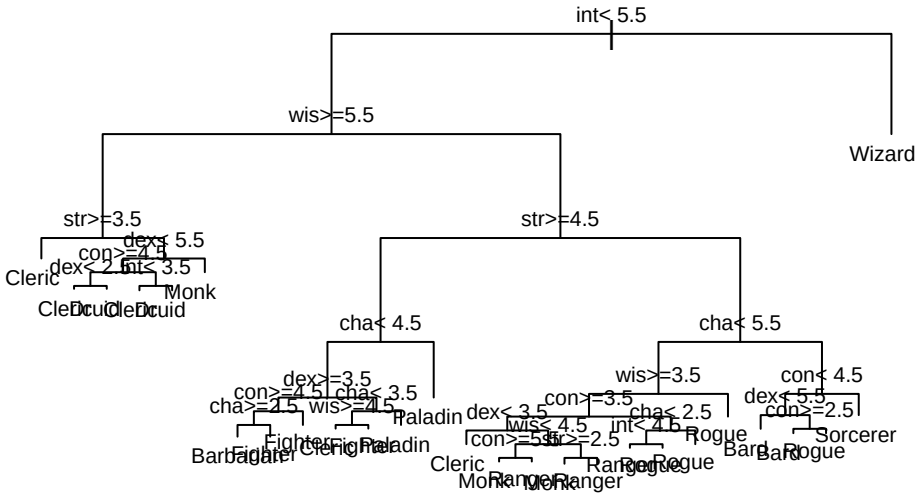
kNN Model

The accuracy of the **ranked tuned kNN** model is 0.49.

A Fourth Model

The fourth and final model tested was a cp-optimized regression tree with the ranked ability score data.

Optimized Regression Tree



The accuracy of the **Ranked Optimized Regression Tree** model is 0.49.

Results - The Best Model

Three of the models developed, the **Tuned kNN** model, the **Ranked Tuned kNN** model, and the **Ranked Optimized Regression Tree** have accuracies of ~50%. The easiest model to visualize is the **Ranked Optimized Regression Tree**.

Conclusions

After visually analyzing and examining the data and testing several models, a model to predict a D&D character's class from its ability scores with an **accuracy** of **0.49** was developed.

One limitation of these models stems from the relatively small size of the dataset. Organized, publicly available data on D&D character stats is difficult to find. A larger dataset, if it could be accessed, would most likely lead to a more accurate model.

That being said, one of my characters is a wood elf ranger named Pylia with the following ability scores: str 11, dex 18, con 12, int 12, wis 14, and cha 12, which ranked is str 1, dex 6, con 4, int 4, wis 5, and cha 4. Following the decision tree for these ability scores, we find that her intelligence rank is less that 5.5 (left at node 1), her wisdom rank is less than 5.5 (right at node 2), her strength rank is less that 4.5 (right at node 3), her charisma rank is less than 5.5 (left at node 4), her wisdom rank is greater than 3.5 (right at node 5), her constitution rank is greater than 3.5 (left at node 6), and her dexterity rank is greater than 3.5 (right at node 6), giving the accurate class of ranger!