Dynamic Generation of Statblocks for Pathfinder 2nd Edition through Machine Learning Models

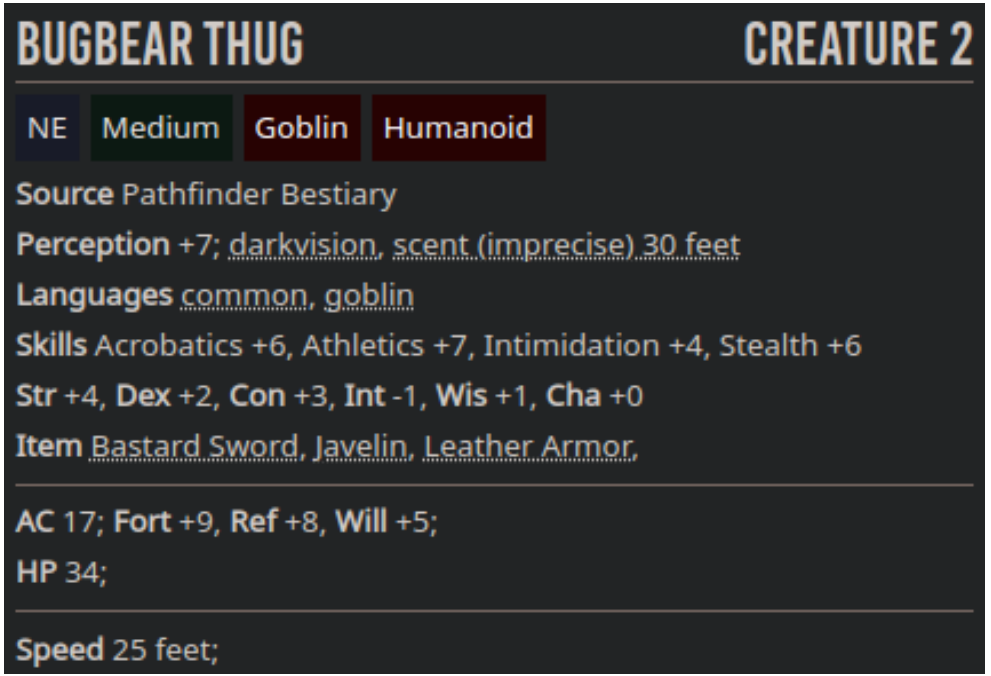
Vess
27 April 2023

Problem Identification, in brief

- Tabletop Roleplaying Games (TTRPGs) are incredibly popular.
- Game Masters / Dungeon Masters spend a lot of time making content for their TTRPG groups.

Problem Identification, in brief

- TTRPG content, statblocks in specific, have a lot of ordinal types or values that fall within a bounded range.



BUGBEAR THUG **CREATURE 2**

NE Medium Goblin Humanoid

Source Pathfinder Bestiary

Perception +7; darkvision, scent (imprecise) 30 feet

Languages common, goblin

Skills Acrobatics +6, Athletics +7, Intimidation +4, Stealth +6

Str +4, **Dex** +2, **Con** +3, **Int** -1, **Wis** +1, **Cha** +0

Item Bastard Sword, Javelin, Leather Armor,

AC 17; **Fort** +9, **Ref** +8, **Will** +5;

HP 34;

Speed 25 feet;

Problem Identification, in brief

- Machine learning models could be trained on these types and values from the wealth of statistical data.
- These trained models could then generate new statblock data based on incomplete suggestions.

Literature and History, in brief

What is DnD

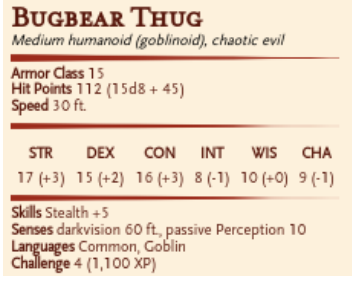
Dungeons and Dragons is the current world leader for the most popular TTRPG system. Wizards of the Coast is the owner of the DnD property and some of its constituents.

What is PF2e

DnD has been released in sequential editions. DnD 5e is the current. Around DnD 3.5e, Wizards of the Coast attempted a greedy license change from open source to not open source. Pathfinder was a split off.

How DnD applies to PF2e

Because Pathfinder is a split off from DnD 3.5e, it follows much of the same system mechanics and statistic data format that DnD does. PF2e is the second Pathfinder edition released thus far.



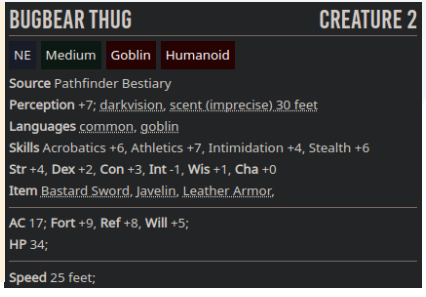
Bugbear Thug

Medium humanoid (goblinoid), chaotic evil

Armor Class 15
Hit Points 112 (15d8 + 45)
Speed 30 ft.

STR	DEX	CON	INT	WIS	CHA
17 (+3)	15 (+2)	16 (+3)	8 (-1)	10 (+0)	9 (-1)

Skills Stealth +5
Senses darkvision 60 ft., passive Perception 10
Languages Common, Goblin
Challenge 4 (1,100 XP)



BUGBEAR THUG **CREATURE 2**

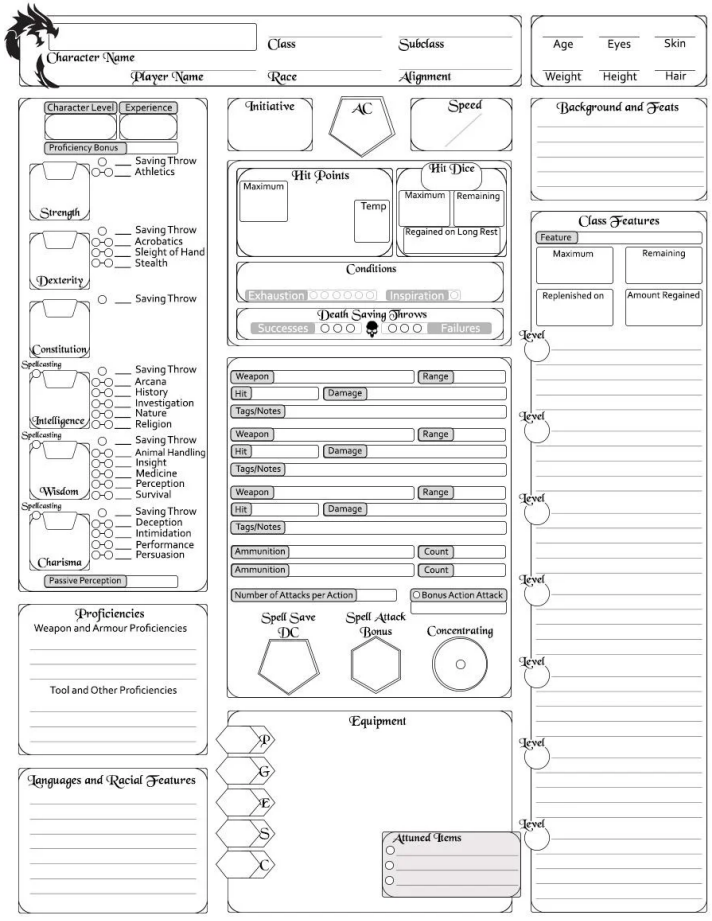
NE Medium Goblin Humanoid

Source Pathfinder Bestiary
Perception +7; darkvision, scent (imprecise) 30 feet
Languages common, goblin
Skills Acrobatics +6, Athletics +7, Intimidation +4, Stealth +6
Str +4, **Dex** +2, **Con** +3, **Int** -1, **Wis** +1, **Cha** +0
Item Bastard Sword, Javelin, Leather Armor,

AC 17; **Fort** +9, **Ref** +8, **Will** +5;
HP 34;

Speed 25 feet;

Literature and History, in brief



DnD Character Sheets

Some research was made into taking in a character biography and training a model to predict the player's preferred race, class, and possible attribute scores.

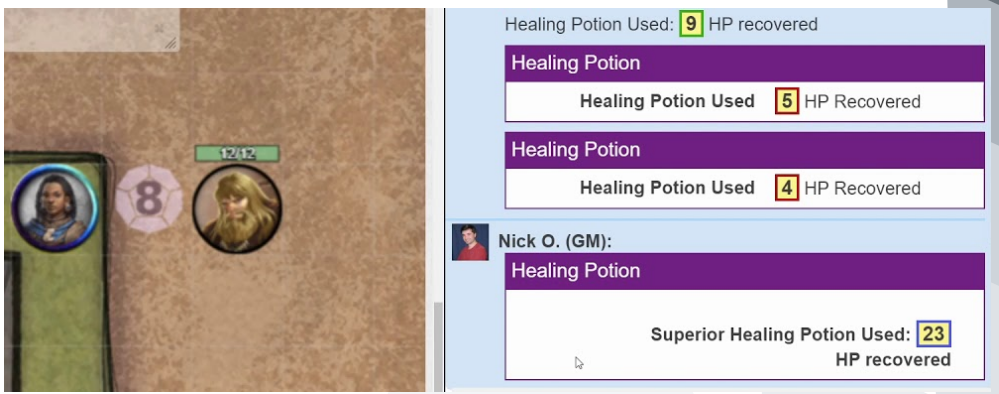
DnD Game Turns

More research was made into using DnD transcripts to train models to generate entire rounds of both combat and roleplay for standard DnD 5e game turns.

DnD AI DMs

With the advent of ChatGPT and popular interest in TTRPG AI applications, Wizards of the Coast has invested attention in developing and investing into AI DMs.

Literature and History, in brief



DnD Character Sheets

Some research was made into taking in a character biography and training a model to predict the player's preferred race, class, and possible attribute scores.

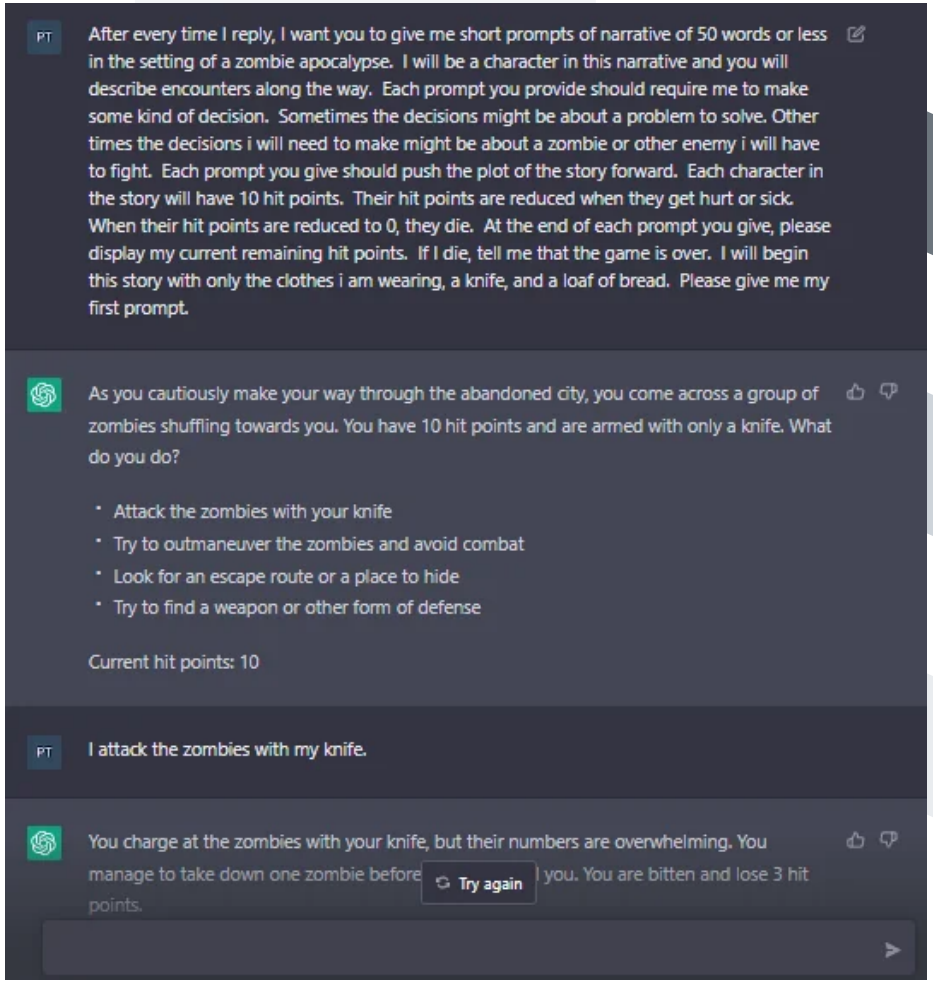
DnD Game Turns

More research was made into using DnD transcripts to train models to generate entire rounds of both combat and roleplay for standard DnD 5e game turns.

DnD AI DMs

With the advent of ChatGPT and popular interest in TTRPG AI applications, Wizards of the Coast has invested attention in developing and investing into AI DMs.

Literature and History, in brief



DnD Character Sheets

Some research was made into taking in a character biography and training a model to predict the player's preferred race, class, and possible attribute scores.

DnD Game Turns

More research was made into using DnD transcripts to train models to generate entire rounds of both combat and roleplay for standard DnD 5e game turns.

DnD AI DMs

With the advent of ChatGPT and popular interest in TTRPG AI applications, Wizards of the Coast has invested attention in developing and investing into AI DMs.

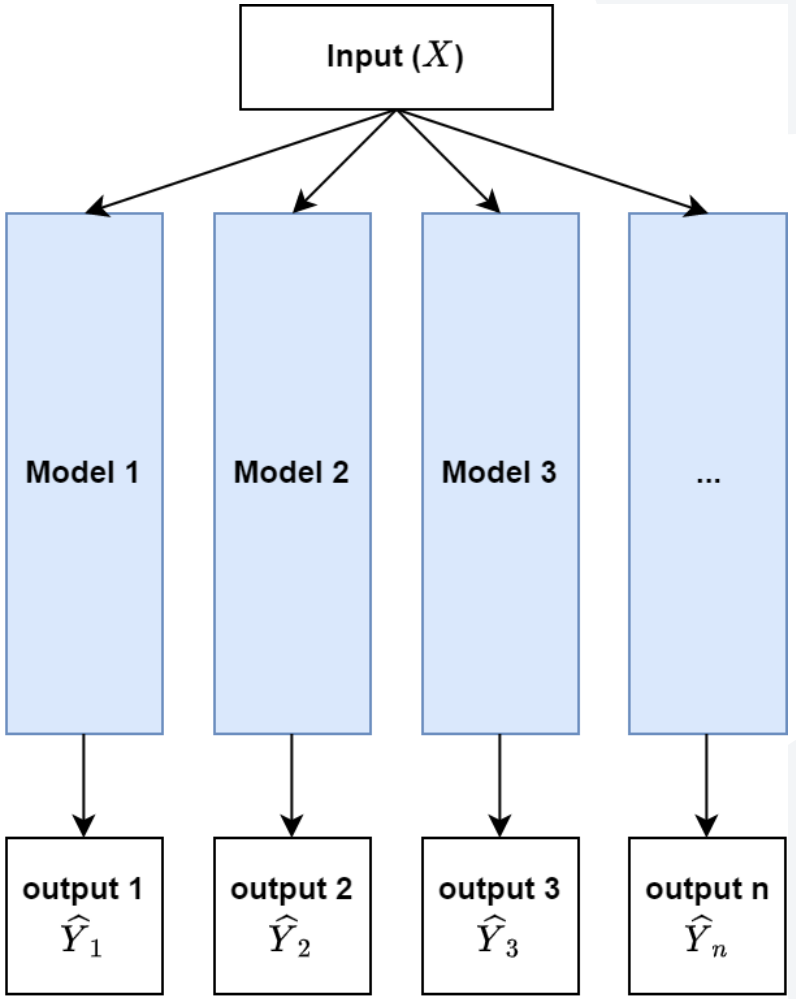
Concepts and Technology, in brief

Multi-Output Models

Most model types only support single-output, where they attempt to predict a single unknown value target.

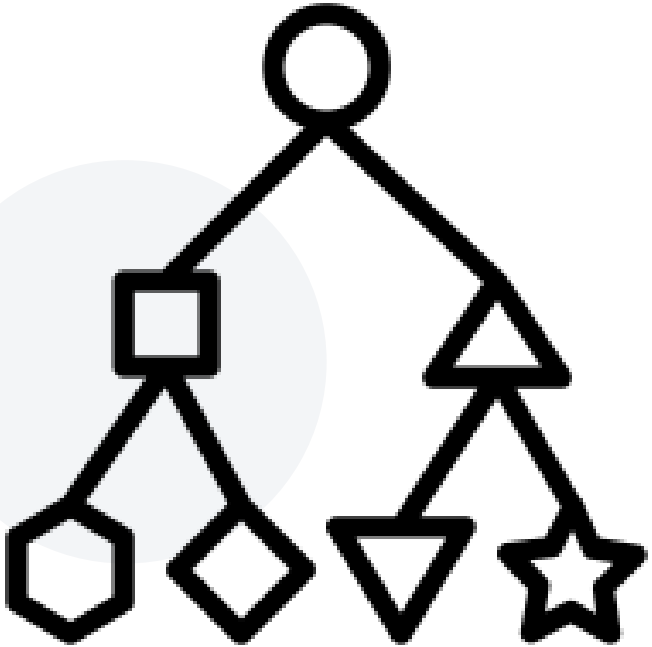
Some models support, or can be extended to support, multi-output, where they attempt to predict multiple unknown value targets.

To do so, a distinct model of the model type is trained to guess each unknown value.



Concepts and Technology, in brief

Classifier and Regressor Models

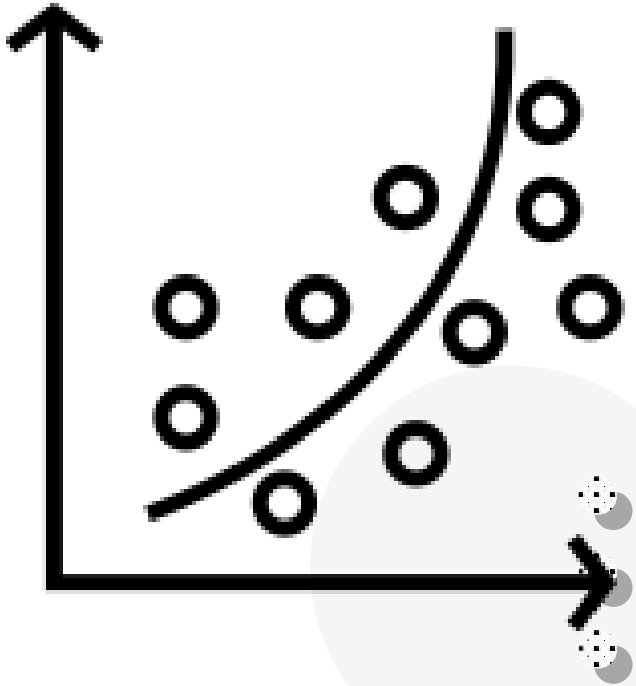


Both classifiers and regressor model types attempt to solve for fundamentally different problems.

Classifiers work better with ordinal types, such as IS, or IS NOT-classification.

Regressors work better with data that represent trends, such as X, X+1, X+2, etc. Curve fitting.

Unfortunately, my data contains much of both types.



Concepts and Technology, in brief

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

Technical Details, Scores

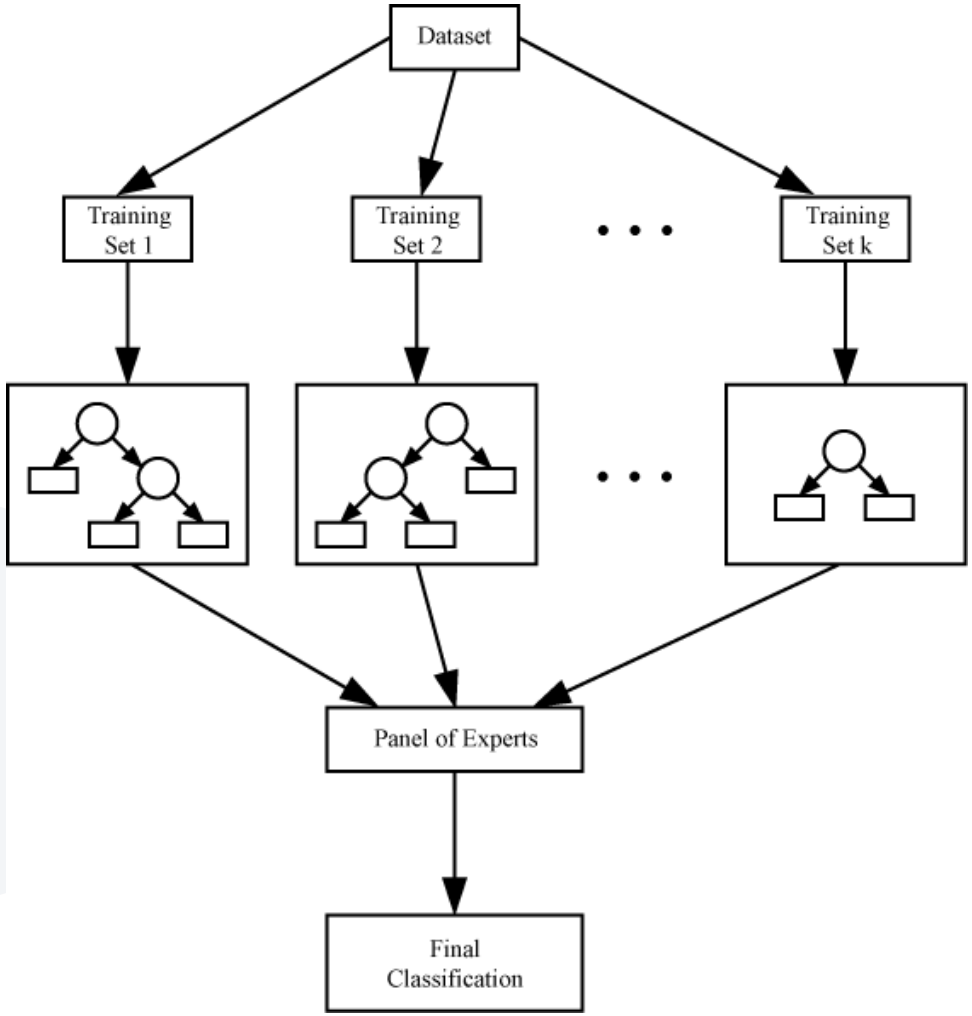
Accuracy score was used to evaluate the classifiers.

Accuracy is a measure of correct guesses against the total of correct and incorrect guesses.

R^2 score was used to evaluate the regressors.

R^2 is a measure of variance from the predicted trend line against the unseen data.

Concepts and Technology, in brief



Technical Details, Voting

The top three classifier and regressor models were put into a voting model.

A voting model is inherently self explanatory- these models vote on what they believe the classification / trend of the unseen data is and make predictions from these votes.

Concepts and Technology, in brief

01 Core Stats

Level
Size
Armor Class
Health Points
Fortitude, Reflex, Willpower

02 Stats

Strength, Dexterity, Constitution
Intelligence, Wisdom, Charisma

03 Skills

Acrobatics, Arcana, Athletics,
Crafting, Deception, Diplomacy,
Intimidation, Lore, Medicine,
Nature, Occultism, Performance,
Religion, Society, Stealth,
Survival, Thievery

04 Extra

Size
Alignment
Perception Bonus

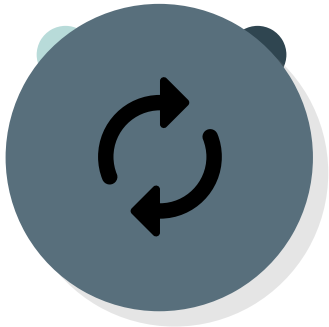
Methodology and Process, in brief

Scraping

Assessed a few websites for possible scraping opportunity.

Settled on one that maintained well structured HTML files that tagged most datapoints of interest.

Created a script that scraped once, to minimize load.



Parsing

Once all of the HTML files were scraped, parsing began.

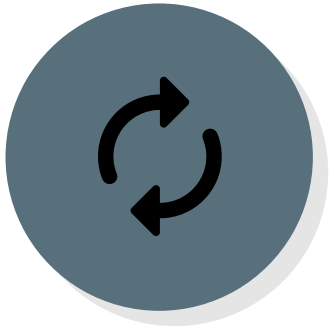
Various unique HTML tags were targeted within the HTML in-order to extract out the necessary data for each statblock.

Cleaning

Various statblocks were identified as non-necessary, duplicates, or as containing data that didn't match the general statblock format.

These statblocks were excluded from the final collection of data.

Data Mining and Generation, in brief



Importing

51 regressor models from SciKit were imported. Of these, 24 were excluded, and 27 were kept.

37 classifier models from SciKit were imported. Of these, 30 were excluded, and 7 were kept.

Exclusion reasons: speed, horrible scoring, non-support for multi-output.

Wrapping

The models that were kept were wrapped in multi-output models for regressors and classifiers.

The wrapping allowed all models to flexibly support either 31 data types as input, or just 1 data type as input.

Finalize

All models were trained on the data in under 10-20 seconds total.

The trained models were then evaluated, before being further wrapped into two voting models.

These voting models were trained in under 5-10 seconds.

Result Presentation, in brief

```python
# Target keep is what we are providing to our models.
target_keep = {
    "level": None, # 2
    "size": 3, # 3 | 1 Tiny, 2 Small, 3 Medium, 4 Large, 5 Huge,
    "law": 2, # 2 | 1 Chaotic, 2 Neutral, 3 Lawful.
    "moral": 1, # 1 | 1 Evil, 2 Neutral, 3 Good.
    "ac": 17, # 17
    "hp": None, # 34
    "perception": 7, # 7
    "fortitude": 9, # 9
    "reflex": 8, # 8
    "willpower": 5, # 5
    "strength": 4, # 4
    "dexterity": 2, # 2
    "constitution": 3, # 3
    "intelligence": -1, # -1
    "wisdom": 1, # 1
    "charisma": 0, # 0
    "acrobatics": 6, # 6
    "arcana": 0, # 0
    "athletics": 7, # 7
    "crafting": 0, # 0
    "deception": 0, # 0
    "diplomacy": 0, # 0
    "intimidation": 4, # 4
    "medicine": 0, # 0
    "nature": 0, # 0
    "occultism": 0, # 0
    "performance": 0, # 0
    "religion": 0, # 0
    "society": 0, # 0
    "stealth": 6, # 6
    "survival": 0, # 0
    "thievery": 0, # 0
}
```

```python
target_keep = {
    "level": 2, # 2
    "size": None, # 3 | 1 Tiny, 2 Small, 3 Medium, 4 Large, 5 Huge, 6 Gargantuan.
    "law": None, # 2 | 1 Chaotic, 2 Neutral, 3 Lawful.
    "moral": None, # 1 | 1 Evil, 2 Neutral, 3 Good.
    "ac": None, # 17
    "hp": 34, # 34
    "perception": None, # 7
    "fortitude": None, # 9
    "reflex": None, # 8
    "willpower": None, # 5
    "strength": None, # 4
    "dexterity": None, # 2
    "constitution": None, # 3
    "intelligence": None, # -1
    "wisdom": None, # 1
    "charisma": None, # 0
    "acrobatics": None, # 6
    "arcana": None, # 0
    "athletics": None, # 7
    "crafting": None, # 0
    "deception": None, # 0
    "diplomacy": None, # 0
    "intimidation": None, # 4
    "medicine": None, # 0
    "nature": None, # 0
    "occultism": None, # 0
    "performance": None, # 0
    "religion": None, # 0
    "society": None, # 0
    "stealth": None, # 6
    "survival": None, # 0
    "thievery": None, # 0
}
```

Result Presentation, in brief

```
Training:
[(0.26575540680069787, 'TransformedTargetRegressor', TransformedTargetRegressor()), (0.26575549999720877, 'Ridge',
Ridge()), (0.2999308291212176, 'GradientBoostingRegressor', GradientBoostingRegressor())]
0.28512717899704065, 2.75s.
```

```
Training:
[(0.0, 'KNeighborsClassifier', KNeighborsClassifier()), (0.0, 'LabelPropagation', LabelPropagation()), (0.0, 'Rando
mForestClassifier', RandomForestClassifier())]
0.0, 19.16s.
```

```
Training:
[(0.9716045185772781, 'HuberRegressor', HuberRegressor()), (0.9808225060009667, 'BaggingRegressor', BaggingRegressor
()), (0.9837911169746241, 'GradientBoostingRegressor', GradientBoostingRegressor())]
0.9851150929845455, 1.51s.
```

```
Training:
[(0.34375, 'BaggingClassifier', BaggingClassifier()), (0.34375, 'ExtraTreesClassifier', ExtraTreesClassifier()),
(0.40625, 'RandomForestClassifier', RandomForestClassifier())]
0.40625, 4.2s.
```

Result Presentation, in brief

```
[('level', 0.038197730508405 95), ('hp', 41.75786077924345)]

[('level', -1), ('hp', 9)]
```

Conclusions

The regressor models provided values closer to our expectations, but also crazier values outside of normal ranges.

The classifier models provided values far outside our expectations, but the values were more sane and within normal ranges.

Neither models fell within the range of “success”. Conclusion for this project: General failure.

```
[   ('size', 2.2333777680533244),
    ('law', 1.8161292020506214),
    ('moral', 1.8840834215654538),
    ('ac', 62.35127332707301),
    ('perception', 49.34511125307656),
    ('fortitude', 37.368398688594084),
    ('reflex', 53.44839391132714),
    ('willpower', 50.41679644347613),
    ('strength', -3.4736877739651404),
    ('dexterity', 14.865276412725336),
    ('constitution', 0.14154101635152008),
    ('intelligence', 8.719729126330156),
    ('wisdom', 8.330296001793208),
    ('charisma', 10.12413242163626),
    ('acrobatics', 37.41820822597882),
    ('arcana', 7.926468840028012),
    ('athletics', -8.029282198926436),
    ('crafting', 12.06431883573184),
    ('deception', 30.78055365313196),
    ('diplomacy', 18.356661879102347),
    ('intimidation', 16.727747626600202),
    ('medicine', 1.8137597829117265),
    ('nature', -1.065886729834822),
    ('occultism', 31.651116603284333),
    ('performance', 4.390949346312963),
    ('religion', 9.39003872719804),
    ('society', 6.90061577218573),
    ('stealth', 35.535450402516325),
    ('survival', 5.748415640610702),
    ('thievery', 14.69776993934788)]
```

```
[   ('size', 1),
    ('law', 2),
    ('moral', 2),
    ('ac', 1),
    ('perception', 0),
    ('fortitude', 0),
    ('reflex', -1),
    ('willpower', 0),
    ('strength', -5),
    ('dexterity', -5),
    ('constitution', 0),
    ('intelligence', -5),
    ('wisdom', 0),
    ('charisma', 0),
    ('acrobatics', 0),
    ('arcana', 0),
    ('athletics', 0),
    ('crafting', 0),
    ('deception', 0),
    ('diplomacy', 0),
    ('intimidation', 0),
    ('medicine', 0),
    ('nature', 0),
    ('occultism', 0),
    ('performance', 0),
    ('religion', 0),
    ('society', 0),
    ('stealth', 0),
    ('survival', 0),
    ('thievery', 0)]
```

Future Iteration, in brief

Future work that iterated on mine could:

1. Package my work so that it is more transportable and accessible for people to run.

2. Fine tune each of the models for the problem, generally, or for each unique problem target set.

3. Seek out further PF2e datasets to utilize as reference. Or, relax my restrictions on filtering statblocks from my dataset.

What did I learn?

1. Set reasonable expectations.

2. Cut down on goals to deliver.

3. Machine learning isn’t a panacea.

4. Garbage data in, garbage data out.

5. Python is great.

6. Be aware of required solution space.

7. Iterative development is great.

8. Problem identification is great.

9. Comments are great.

Thank you for attending.