

Kip McCharen (cam7cu)

Clair McLafferty (cm2rh)

## DS 6014 Project Proposal

For our project, we plan to use Bayesian Machine Learning to detect fraud in Travis McElroy's dice rolling during games of Dungeons and Dragons in the podcast *The Adventure Zone*. Bayesian Machine Learning has been used extensively for industrial fraud detection, as evidenced by more than 20,000 research paper results in Google Scholar<sup>1</sup>. Since fraud is usually a human choice that aims to evade detection by imitating normal human behavior, using a probabilistic approach makes sense. To apply the theory, we must therefore quantify expectations of "normal" human behavior and update them using specific instances of fraud to come up with a posterior distribution that isolates instances of fraud.

Dice are often used as simple, academic probability exercises, but are generally not the subject of real world analysis. However, there are many games that use dice rolling as a mechanism for success. In our COVID-affected world, remote games of Dungeons and Dragons<sup>2</sup> are more common, and unless a tool is employed to broadcast all rolls to all players, we must trust our fellow players to tell the truth about the results. The risk of fraud is high, especially when the group is in danger of being defeated by an opponent or if there is pride on the line. There are complications to this analysis such as advantage, player stats, and dungeon master choices, all of which may modify roll results on the fly. However, this could make for an interesting and nuanced model.

The McElroy family makes a Dungeons and Dragons podcast called *The Adventure Zone*<sup>3</sup>. As co-host Justin McElroy<sup>4</sup> notes on Twitter, his brother and co-host Travis's unusually successful dice rolls over the course of the show have caused some controversy. So much so, in fact, that members of the community have manually documented rolls made by Travis over several seasons<sup>5</sup>. The dataset (more than 150 observations) is presented in raw text that we will parse using regular expressions to extract data which may serve as our likelihood estimation (Bernoulli), as well as a basis to approximate a prior distribution notionally independent of Travis' bias. These are our initial impressions of the data, but the expected distribution and exact modeling technique used may change depending on exploratory data analysis.

In this project, we will explore several modeling options including MCMC analysis, naive Bayes modelling, and potentially Bayesian belief networks<sup>6</sup> to model this data. Using this method, we hope to quantify the probability that Travis McElroy is cheating.

---

<sup>1</sup> <https://scholar.google.com/scholar?hl=en&q=%22fraud%22+detection+%22bayesian%22+machine+learning>

<sup>2</sup> <https://www.inverse.com/gaming/dungeons-and-dragons-online-coronavirus-zoom>

<sup>3</sup> <https://maximumfun.org/podcasts/adventure-zone/>

<sup>4</sup> <https://twitter.com/justinmcelroy/status/710638075689340928?lang=en>

<sup>5</sup> [https://www.reddit.com/r/TheAdventureZone/comments/6pxwdr/im\\_back\\_with\\_more\\_stats\\_on\\_travis\\_rolls\\_its\\_the/](https://www.reddit.com/r/TheAdventureZone/comments/6pxwdr/im_back_with_more_stats_on_travis_rolls_its_the/)

<sup>6</sup>