Capstone Two: Project Proposal

For my second capstone project, I will be analyzing a dataset of tips collected by a server over the course of a few months to determine if we can predict the amount of tips a server will receive at the end of their guests’ meals. Throughout the course of my project, I may detect other features that are worth examining in further detail. The results of this project could be important, not just to me due to personal curiosity, but to different levels of service industry staff. This could also potentially change the service industry landscape, as this sort of information has never been public knowledge. It could play an important role for workers when deciding where they want to work.

Every day, potential interviewees apply to jobs at restaurants and bars, and just like in the corporate world, there is usually one big question on their mind: How much does this job pay? However, unlike the corporate world, servers and bartenders do not receive a defined annual salary, such that most hiring managers cannot answer this question accurately during the hiring process. Most of the time, hourly pay is the only concrete number that can be given, yet this is often a small portion of a worker’s total wages and not a very useful metric. For example, most servers and bartenders are paid minimum wage, yet some can make $5/hour on tips while others can make $150/hour on tips. Without knowing this beforehand, it would take some educated guessing on the applicants’ part to figure this out. Being able to put this centuries-old question to rest would help managers create more informative job postings and interviewees to make more informed application decisions. The criteria of success for this project would be fitting an accurate predictive model for gratuity amount based on the features given. The scope of solution space would be any variable that is relevant to tip amount. The constraint within solution space pertains to the data collected. Since the data was only collected from one restaurant by one server within the span of a few months, it will be difficult to generalize our results to other seasons at the same restaurant, let alone other restaurants. Tip amount can obviously depend on quality of service provided, yet customer service quality is not one of the variables provided. This is beneficial, however, because such a variable is subjective, so our model will only be based on objective features. Thus, being able to predict tip amounts at one restaurant with easily collected variables could be enough to inspire other restaurants to come up with their own predictive models. Stakeholders to provide key insight will be my mentor, Eleanor Thomas, as well as anyone who I may share this project with. The key data source will come from Kaggle.

The Kaggle dataset I will be using has been scraped from https://www.kaggle.com/jsphyg/tipping. After making sure the data is clean and workable, I will analyze it for correlations between tips and customer features. This will include variables like gender, whether there is a smoker in the party, and total bill amount. I will then compare the strength of correlations between customer features and age to see whether some factors have a heavier influence than others. My deliverable will be a GitHub repo containing the work for each step of my project, including a slide deck and project report.