

First Pitch

Speech recognition is present to enable users to communicate via voice to computers. These systems use Machine Learning methodologies such as Deep Learning - based on Neural Networks (NN) - to learn and improve from experience automatically. In speech recognition domains, questions like: How do we represent individuals from all demographic groups? How do we identify biases in this system?, and How might we mitigate these biases? are being addressed at a rapid pace. Previously, I explored the minimal dataset needed to effectively personalize speech recognition software with a focus on Long Short-Term Memory (LSTM) NN. As there exist various NNs for speech recognition, it is important to know the optimal NN to classify and recognize users. To this end, my project analyzes Recurrent NN (RNN), Long Short Term Memory (LSTM) NN, and hybrid Hidden Markov Model (HMM)-LSTM NN to determine the optimal NN for speech recognition.

Second Pitch

Matching with text data is about finding specific patterns, substrings, or structures with a given text. The quality of the match is dependent on the distance metric. Matching for causal inference is a well-studied problem, but standard methods fail when the units to match are text documents. That is, the high-dimensional and rich nature of the data renders exact matching infeasible, causes propensity scores to produce incomparable matches, and makes assessing match quality difficult. There, however, exist frameworks for matching methods that make it possible to match text data and evaluate the quality of the resulting matched samples. Matching with text data can be applied to areas where decision-making is based on selection criteria. Matching students and educational institutions is one area. Based on a fixed set of institution-related text documents, and a collection of potential covariates for student qualifications, I seek to build a program that combines different distance metrics to effectively match student-athletes to institutions. The program will be evaluated relative to human judgment about match quality.

Third Pitch

Students gain admission to universities with the hope of gaining knowledge and achieving high academic performance. Similarly, institutions are continuously seeking to enhance the learning experience and success outcomes of students. However, the success of a student is dependent on numerous factors such as demographics and academic environment. People often gain admission to institutions based on the reputation of the institution, financial aid, location, and many more. However, none of these are good predictors of the student's academic success. Understanding the factors contributing to academic success is vital for tailored interventions. To this end, my project aims to predict and understand the factors that contribute to student academic success in a university setting. By using available data on past students, I will perform educational analytics based on logistic regression and decision trees to inform on a student's success.