Project Proposal
Predicting Mastery in Academic Assessments

Students taking advantage of practice exercises through various software applications seek an evaluation of their preparedness for formal academic assessments.

The goal of this project is to help test developers understand user progress by predicting whether a student answers a test question correctly. The project examines data from students preparing for three tests: the GMAT, SAT and ACT. Question fields in the dataset indicate outcome, group (test for which the student is preparing), track (test subject), sub-track, tags, question format, and times for when the question was started and answered or deactivated (timed-out). Relationships that could predict observations under the field labeled 'correct' are explored and modeled.

Approach

1. **Initial Data Exploration -** Import and inspect raw data. Isolate relevant variables, fill or calculate new variables, and organize the dataframe. Resolve missing, invalid, corrupted or duplicate values.

2. **Exploratory Data Analysis -** Create data visualizations to explain variables and analyze outliers. Possibly remove or mark outliers. Explore variable dependence and correlations.

3. **Machine Learning -** Build, fit, and validate a method to model the data. Evaluate the performance of each model tested. Methods to try: Logistic Regression, Random Forest, Linear Mixed Effects, Mixed Effects Random Forest, XGBoost.

4. **Final Reports & Presentations**

Deliverables

- All Code & Notebooks
- Project Reports
- Final Report
- Slide presentation