



Code for Good 2018

Halogen Foundation Singapore

Team 6

Problems with Existing Solutions



Halogen Foundation Singapore aims to provide quality leadership and entrepreneurship education to all youths regardless of their background.

However:

- Questionnaires are too long and cause inaccurate test results
- Some questions are redundant and can be removed
- Tabulation of questionnaire scores is manual and time-consuming
- Currently lacks automatically generated statistical reports and data visualisations
- Complications with collecting user data due to PDPA

Used Technologies



PostgreSQL

Database



GitHub

Repository

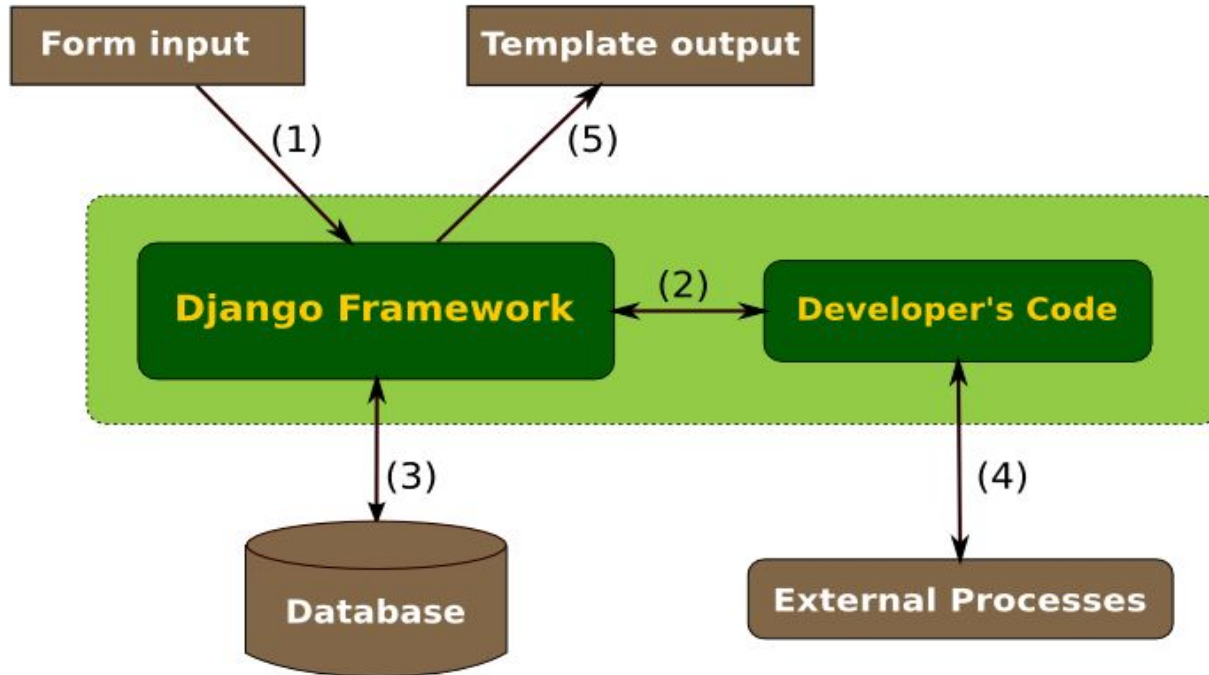


Development



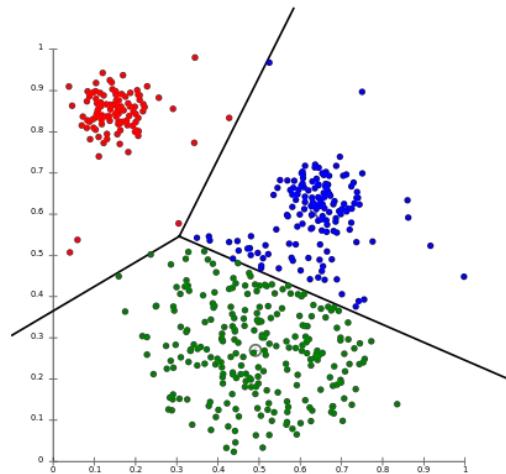
Machine
Learning

Django Python Framework



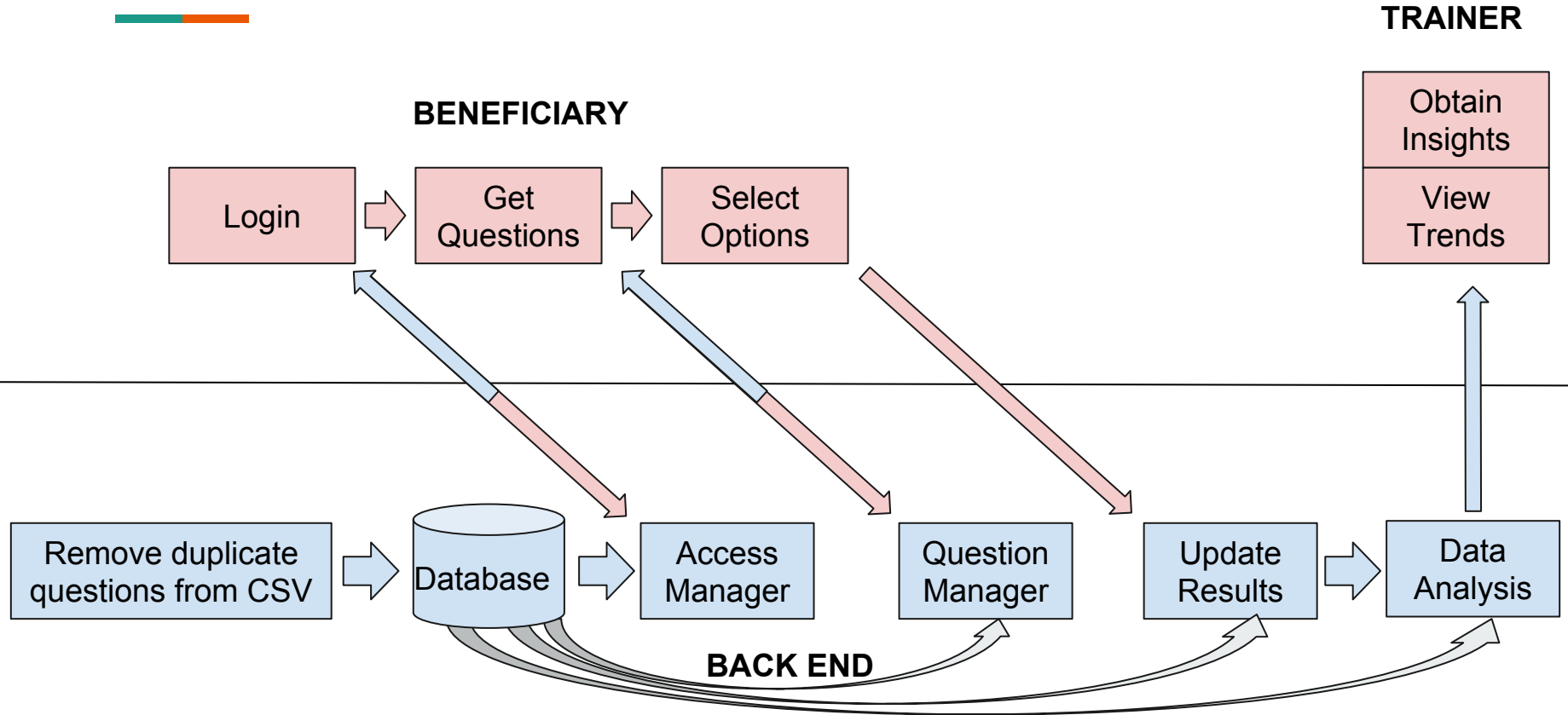
Machine Learning

- Clustering to group similar questions together



- Pick 1 question from each cluster to ensure that responses are coherent
- Regression models to predict scores based on remaining questions after clustering

Proposed Workflow



Progress / Accomplishments



- Clustered previous 100 behavioral questions to 60 questions. Redundant questions are archived from questions set.
- Did separate regression models for predicting each keystone's scores
- Python Flask setup on server
- PostgreSQL database set up and populated
- Attractive user-facing web form design to provide ease of usage



Data Analysis

Regression Model:

Prediction:

Further Improvements



- PostgreSQL Database to be hosted online
- Ability to add and modify questions
- Potential integration of National Digital Integrity (NDI) API to securely obtain users' data and avoid PDPA issues