

ML Team: Nathan Meiners, Thomas Miller, Alex Muzika, Angela Fleenor, Nathan Groshek

Find My Major

Ranjidha Rajan: Assistant Professor of Computer Science, MSU Denver
Mark Yoss: Lockheed Martin Endowed Director of Advanced Manufacturing Sciences Institute

UI Team: Bing Huang, Kayleen Gallegos, Owen Pick, RJ Sam, Sayuj Shrestha, Firehiwot Tiruneh

Introduction

Undecided students often receive broad or impersonal career guidance during short advising interactions, as traditional Holland Code (RIASEC) tools rely on lengthy, static assessments. This project explores a lightweight machine learning-based approach to inferring a student's RIASEC profile from a small set of adaptive questions and generating ranked career recommendations. A web-based interface allows users to refine results by removing undesirable options. This approach aims to improve the efficiency and relevance of career guidance in time-constrained settings such as college fairs.

Background



The Holland Code (RIASEC) framework categorizes career interests into six types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. Traditional systems estimate a user's profile through long questionnaires and map results to matching careers. While widely used, these approaches are often static and require significant user input, limiting their effectiveness in short advising scenarios.

Process

Baseline Synthetic Datasets

O*Net Career Library Data

Title	Realistic	Investigative	Artistic	Social	Enterprising	Conventional	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness
Chief Executives	0.043333	0.341666667	0.193333	0.423333	0.993333333	0.961666667	0.91628896	0.507352841	0.844861021	0.86199095	0.812217
General and Operations Manag	0.2	0.228333333	0.548333	0.396667	1	0.723333333	0.87566666	0.554205643	0.308239843	0.79966226	0.67420813
Legislators	0.09	0.391666667	0.283333	0.448333	0.753333333	0.436666667	0.9479638	0.599547511	0.802521306	0.817647059	0.7858220

Cleaned Kaggle Dataset

Normalized	Normalized	Normalized	Normalized	Normalized	Major	Broad Category
0.35	0.875	0.475	0.55	0.25	0.25	nursing
0.325	0.9	0.85	0.925	0.5	0.85	education
0.825	0.8	0.925	0.85	0.325	0.825	literature
0.45	0.35	0.325	0.5	0.4	0.325	math
0.4	0.8	0.85	0.775	0.625	0.5	mathematics and science
0.2	0.275	0.2	0.5	0.2	0.45	family resource management
0.475	0.425	0.65	0.75	0.6	0.725	counseling
0.3	0.35	0.475	0.725	0.7	0.25	business
						Business & Finance
						Psychology & Human Services
						Business & Finance

Models

Heuristic

- Rule-based baseline mapping responses directly to RIASEC scores

KNN (k-Nearest Neighbors)

- Predicts using the most similar past examples

Gradient Boosting

- Combines weak models sequentially to reduce errors

Random Forest

- Ensemble of decision trees using random subsets of data

Logistic Regression

- Linear model estimating probabilities for outcomes

Best Inference Model

- Gradient Boosting selected
- Highest ranking performance across evaluation metrics on synthetic data

Data Collection

- User rejects suggested careers until top 5 careers are found.
- RIASEC scores and 5 careers are stored to retrain and test models.
- All data is kept anonymous.

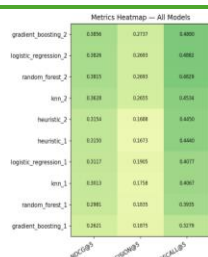
Evaluation

Precision@K

- Fraction of top K recommendations that are relevant

NDCG@K

- Rewards correct ranking order, prioritizing higher-ranked results

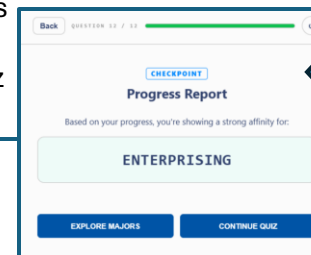


Export model

Implementation

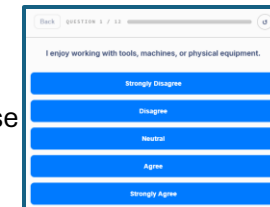
Checkpoint

- Displays top RIASEC score
- Explore Majors
- Can continue taking the quiz



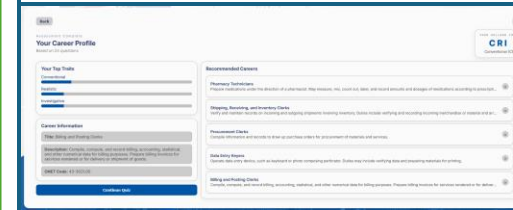
Questionnaire

- Built via vite, React, Vercel, Render
- Total of 48 questions.
- Checkpoint system at every 12 questions, for quick and easy results.
- Outputs response data for ML inference.



Results Page

- Displays a list of careers based on RIASEC scores
- Can exclude careers the student deems not accurate/not interested
- Save results via mailer if student likes their career options
- Mapping O*NET careers to MSU's major catalog



Conclusion

Preliminary results on synthetic data show moderate ranking performance across models, but these findings remain inconclusive without real-world validation and user interaction data.

Future Work

Future work will focus on real-world data collection and model validation, followed by iterative refinement. In addition, dynamic and purposeful question selection to get a more accurate outcome.

References, demos, etc.



© 2026 Find My Major Project Teams.
This work is licensed under CC BY-NC 4.0.
<https://creativecommons.org/licenses/by-nc/4.0/>