Learning Career Knowledge: Can AI Simulation and Machine Learning Improve Career Plans and Educational Expectations?

I-Chien Chen, Lydia Bradford, and Barbara Schneider

Contents

1 Introduction

Innovative automation advancements are profoundly affecting markets and societies in a rapidly changing information world (Arntz et al. 2016). Additionally, for young adults and those who have lost their jobs, the employment landscape is characterized by ambiguity and insecurity (Blustein et al. 2020a, b). Knowing the demands and requirements of specific jobs can be helpful for those seeking employment. How to align individual career goals and specific employment opportunities requires sophisticated information, guidance, and navigation (Kim et al. 2019; Nunley et al. 2016; Pinto and Ramalheira 2017). This process can become less complicated with

I.-C. Chen (✉) · L. Bradford · B. Schneider
Michigan State University, East Lansing, MI, USA
e-mail: ichiench@msu.edu; bradf134@msu.edu; bschneid@msu.edu

H. Niemi et al. (eds.), *AI in Learning: Designing the Future*,
https://doi.org/10.1007/978-3-031-09687-7_9

machine learning applications. Relying on several studies using game simulations and machine predictions to assist young adults in their career selections (Nie et al. 2020; Schumacher et al. 2010), this chapter explores the unique features of gamification in learning, machine learning, and artificial intelligence (AI) technology. The logic of gamification is described showing how these applications have been implemented to understand players' capacity, skills, and interests in selecting future occupations. This process includes machine learning decision tree algorithms that map out possible job selections, built upon players' career choices and opportunities given their background characteristics to increase the prediction precisions. Results from data insights can be implemented into a series of games to enhance users' knowledge of possible college and career choices. Finally, there are advantages of connecting mobile application, machine learning, and data insights used for predictions which extend user career knowledge, especially in domains where information is often ambiguous and inaccessible.

Unquestionably, young people today play games, often on their phones. Game technology has become a mainstay of entertainment, and it is a prime avenue for games that are challenging, fun, and transmit information at the same time. One area that has not yet been successfully designed and gamified is learning the link between education and career choices. This is a particularly important situation today, given that careers have expanded so rapidly, and vital information is not codified into an easily accessible place. Combining career and educational requirements—a combination where they can learn about emerging jobs, their corresponding educational requirements, and prospects for potential hiring, salaries, security, and advancement—is critical for adolescents' planning for the future.

Init2Winit was developed to fill this gap within smartphone technology. The gamified architecture follows the front-end, back-end system which introduces the participants into career knowledge and gameplay that is transferred to a confidential and de-identified individual database. The overall goal of Init2Winit is to help students learn more about the college-to-career process, which in turn will inspire students to improve their college applications and widen their college and STEM major choice options. The Init2Winit design combines a personalized exploration of career goals to assess individual-level alignment knowledge of the pathways from education to employment.

2 Making More Informed Career Choices: A Theoretical Framework

Recognizing the problem that limited information can create for informed college and career planning led to the creation of the theory of aligned ambitions (Schneider and Stevenson 1999). Alignment theory refers to a status of "aligned ambitions" for young people who begin to develop an emerging understanding of the types of jobs they aspire to, how much education they need to attain these positions, and realistic projections on the annual salary. When young people are more aware of their abilities, strengths, and skills, they are more likely to develop a strategic plan that aligns education expectations and aspirations for their career goals.

Renbarger and Long (2019) find that a lack of access to information on financial aid and college programs has detrimental effects on college enrollment and completion. Cohodes and Goodman (2014) also find that students in disadvantaged schools have limited information on how to apply for college or meet important college-related deadlines. As a result, many students may not know how to make a smooth transition from education to employment nor how to navigate an educational system where choices have real consequences on postsecondary enrollment, degree completion, and employment (Castleman and Goodman 2018). Not having a realistic sense of aligned goals can keep students from being able to focus on the required courses, preparation, and skill development.

The consequence of misaligned knowledge has been shown to result in overestimating or underestimating requirements for college for a career pathway (Schmitt-Wilson and Faas 2016; Perry et al. 2016). Under-aligned high school students assume the pathway to specific jobs can be achieved without completing a postsecondary degree (Kim et al. 2019). The consequences of misaligned knowledge for low-income students can be costly, leading to financial debt or dropping out before obtaining a college degree (Morgan et al. 2013; Bettinger et al. 2012). A recent study has shown that nearly one-third of low-income students had under-aligned career expectations (Chen et al. 2020, 2021). Under-aligned students, while able to estimate a realistic salary range for a job, often were unaware of the educational requirements for a desired job. Students with misalignment knowledge in high school show significantly lower educational expectations, college preparation, and school GPA (Kena et al. 2016; Schneider 2009).

The prevalence of misalignment among low-income students also occurs for students outside the USA. PISA 2018 results show that one-third (30%) of young people from disadvantaged backgrounds are more likely to have misaligned career expectations than one-tenth of their advantaged peers across countries (Mann et al. 2020; Nedelkoska and Quintini 2018). The impact of misalignment has become a global issue due to uncertainty surrounding the job market and automation, and risks have risen in the digital era, particularly among lower-educated workers. Although past research has identified the gap between young people's desired jobs and employment realities, research is lacking on how these differences correspond to labor demands, college knowledge and eligibility, and an individual's needs on a case-by-case basis (Hoff et al. 2021; Schneider and Young 2019; Albion and Fogarty 2002). Career knowledge is critical to help guide an individual's efforts and decisions about college planning during high school.

2.1 Why AI About Career Knowledge?

The introduction of AI applications with megatrend data gathering and forecasting would benefit this decision-making process. Enrolling in a college or finding a job is not a simple cost/benefit question. As concerns rise over mismatched expectations, overqualified skills, or youth unemployment, making better decisions to optimize an individual's strengths, considerations, interests, and skillsets becomes imperative for young people. The decision-making process tends to rely on information and situational assessment to navigate a personalized college-to-work pathway, which is needed to warrant the success of college and work life (Reyna and Farley 2006; Clark et al. 2017; Bureau of Labor Statistics 2015).

2.2 An Example of Gamified Career Knowledge: Init2Winit, an Overview

Init2Winit integrates data-based analytics with occupational information algorithms that allow users to make choices with respect to their education planning and salary projection in visualizing themselves in a dream job. Init2Winit uses points as a feedback mechanism to encourage student participation and performance. Point feedback aims to motivate students to sustain their effort and continue their exploration across different jobs, even for those jobs or college majors that are beyond the students' current plans. To further motivate participation and build college knowledge, Init2Winit allows student performance to be translated into real-world rewards. For example, if a student remains a top five scorer for a week, he or she could earn a voucher for a college visit or an internship with a local company.

2.2.1 Game Design

The gamified architecture structure of the Init2Winit lays out front-end engagement features and a back-end database. The following is an example of the Init2Winit game, designed to motivate a personalized exploration of postsecondary planning and career goals.

2.2.2 Engagement and the Front-End Design

The front-end development focuses on those components of the game that the user sees and interacts with, such as the graphics, interactive user functions, and audio components. The importance of the Init2Winit user experience (UX) design is to keep students' attention on the college-to-career information that students may not know. The game mechanics are a set of rules that dictate the outcome

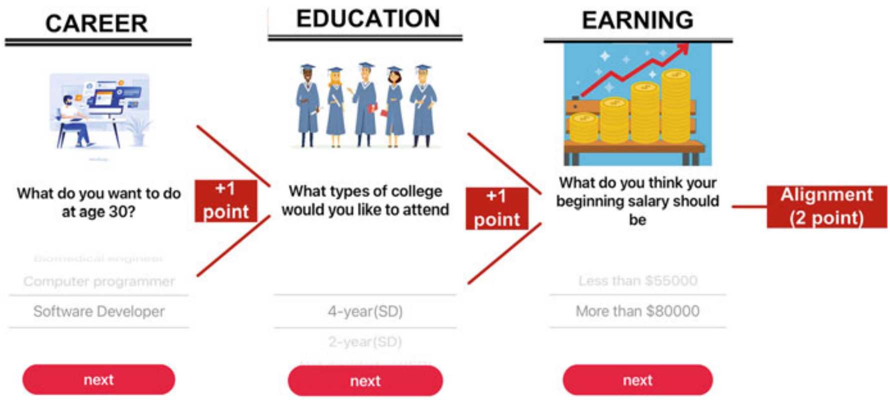


Fig. 1 The full-alignment scenario in the point reward process

of interactions within the system. The data collected are the users' responses to those mechanics. These coupled with an algorithm based on student responses was operated through an interactive interface – using points as real-time feedback on their level of alignment knowledge.

Alignment knowledge indicates that a student can visualize himself/ herself in a career pathway with aligned educational expectations and realistic salary projections. Figure 1 shows an example of how to earn full score points in one play. If a student chooses software developer as a career, he or she needs to know what the educational requirement is for this job and the yearly salary range. When the three informational pieces line up, the user earns the full-alignment score of 2 points. With this knowledge and preparation beforehand, the students are likely of knowing more about employment opportunities in the future.

A student with misaligned knowledge typically chooses either unaligned educational expectations or an unrealistic salary projection. These two types of misalignments cause different consequences to the student. Students with under-aligned knowledge are unaware of the requirements for a job or chose a lower yearly salary than reality. For example, Fig. 2 shows that a student who wants to be a "registered nurse," selects a 4-year college degree, but incorrectly predicts earning less than a $20K yearly salary, indicating a misunderstanding on the salary in the workforce for life science and health-related professionals.

Students with over-aligned knowledge expect to obtain more degrees than required or overestimates the potential annual salaries for their desired career choices. For example, Fig. 3 shows a student who wants to be a "police officer" chooses a 4-year university degree, and expects to earn more than $100K. These choices indicate a misunderstanding of the required education or profession for being in the law enforcement institute (Schmitt-Wilson and Faas 2016). Students earned 0 points if their alignment between career and college planning and career and salary projections are over-aligned.

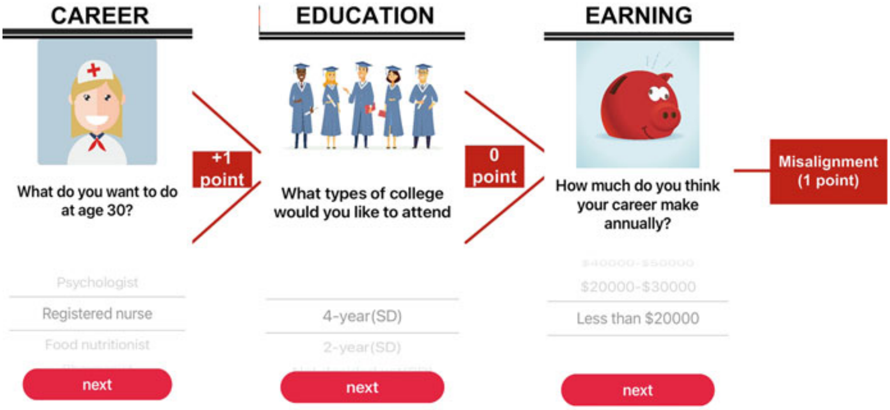


Fig. 2 The misalignment knowledge scenario in the point reward process

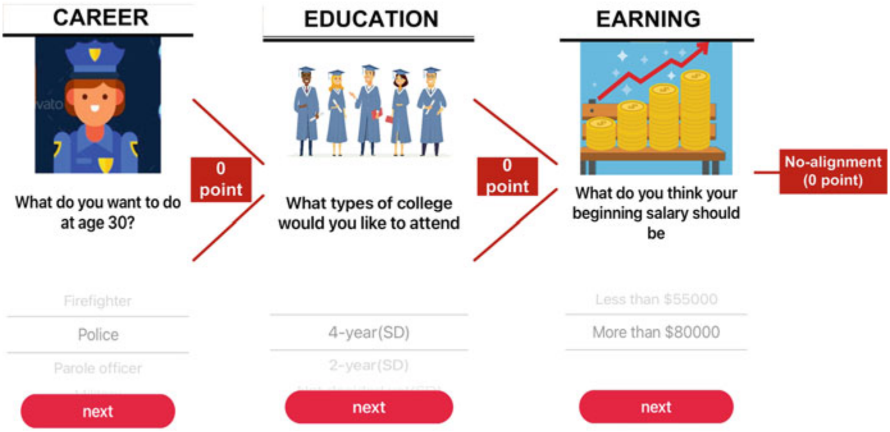


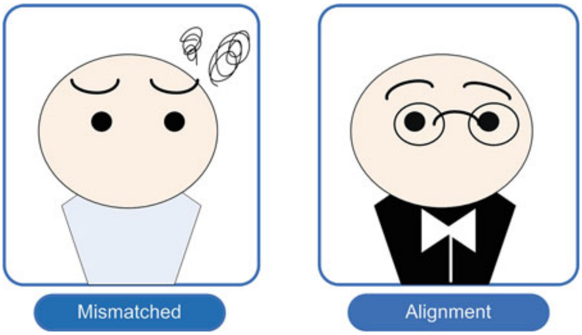
Fig. 3 No-alignment knowledge scenario in the point reward process

Computer-generated images (CGI) help to engage users during gameplay through augmented realities. Users can use forms, images, video, or visualized graphics to depict their stories, profiles, and imaginary selves. Every user can design his/her artwork to represent his/herself. All of this is under computer control and interactive with the servers (Fig. 4).

2.2.3 Design Component and Back-End System

The back-end development focuses on the "server side" of programming, where the connections between the server and the database are constructed. The Init2Winit

Fig. 4 Players' computer-generated characters during the gameplay



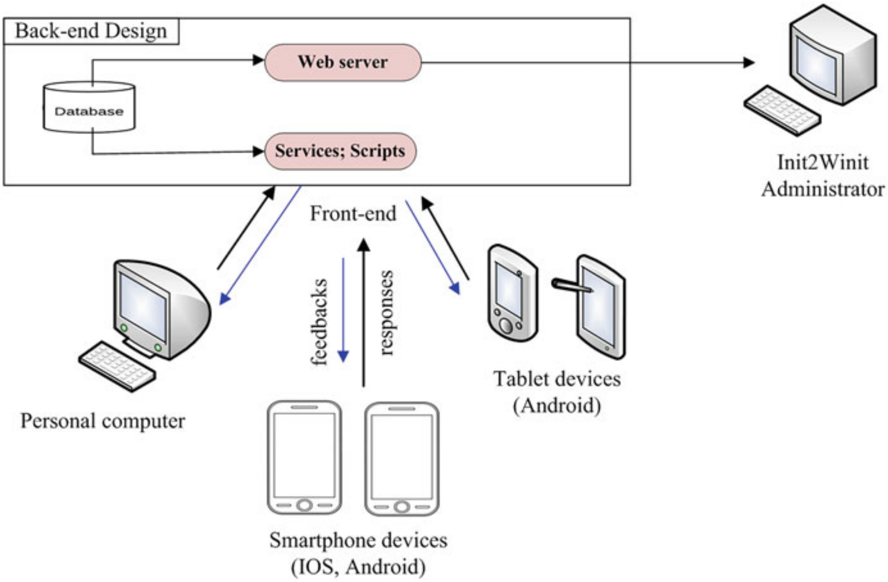


Fig. 5 Init2Winit system architecture

system architecture consists of the following components (Fig. 5): server-side computer system, web application, and mobile device users (including Android/ IOS application for smartphone and tablet). The operating system is a centralized data model which acts as a data hub that interacts with users and conducts data processing between the database and game mechanics as a set of rules and algorithms that guide the outcome of the user's interface interactions. The server-side computer system includes a relational database, user profile, web application, and services for communication with users or for retrieving users' previous records. Those four parts work together to allow for mega data storage and administration for both users and app administrators.

3 Opportunity for AI and Machine Learning (ML)

A broad definition of AI describes a computerized system which "... performs cognitive tasks, usually associated with human minds, particularly learning and problem-solving (Baker et al. 2019: p. 10)." AI and machine learning often refer to similar function as machine learning is a subset of AI, but they are not the same. Modern machine learning models have three types: (1) Supervised machine learning (ML) algorithms based upon existing labeled data or collected information to form a decision, recognizing a pattern, or predicting an outcome. For example, supervised ML can be used to predict dropping out from high school or a high rating score on a writing assignment. (2) Unsupervised classification and profiling are used to sort, identify, and filter unlabeled data based on structures, attributes, features, and densities of resolution. For example, unsupervised ML can be used for customer segmentation or to give recommendations on merchandise. (3) Semi-supervised ML classifies some of the unlabeled/ unidentified information along with labeled and categorized data. For example, semi-supervised ML can be used to classify and organize data, such as sorting writing assignments or job applications into a certain order.

In our case, Init2Winit app could design a function that can be easily integrated with artificial intelligence (AI) which has a broad multifaceted influence running from machine learning to data-based analytic algorithms. The algorithms can create a data feedback system and information loops that allow users to make choices and receive points for identifying correct answers, responses, and task values. The information that In2Winit feeds into the computational game program is based on several national databases. For example, students are asked to select an occupation to pursue, and then, the type of college and majors that they would have to attend to align with this goal in the "career tunnel." The information on what types of degrees or certificates are needed for various occupations is derived from the Occupational Information Network (National Center for O*NET Department 2019), an occupational and STEM knowledge database that contains 974 occupation descriptions and a mix of required knowledge, education, skills, and abilities for each "person–occupation fit" choice.

The Init2Winit app with AI-enabled function could collect real-time information and misalignment patterns of students' knowledge. This misinformation could incite a tool similar to an alarm system which alerts additional assistance and guidance by school counselors or the students' own profiling. An AI-enabled function could also adopt adaptive job-specific or major-specific assessment by adjusting level of difficulty, number of questions, and crucial steps of reaching college-going eligibility and requirements. The Init2Winit app with AI features can also identify student usage behavior, knowledge profiles, and patterns, which can be used to train the machine to adjust the database, and further improve users' personalized decision-making process (Sarker et al. 2019; Bashier et al. 2016).

3.1 Machine Learning and Decision Trees

The following explains how our small-scale pilot study on the Init2Winit prototype was used to understand students' college and career alignment. A small sample of 157, 10th to 12th graders volunteered to participate in the College Ambition Program (CAP). Two schools designed to assist upper secondary students find less costly, prestigious colleges that fit their academic and career interests. During the CAP program, the students completed a pre- and post-survey with valid app user records. Most users are 11th graders, minority, male with GPA ranging between 2.5 and 3.0, and have parents with less than a college education. The Daily Active Users (DAU) shows the frequency of records per user account of those who had at least one play of Init2Winit during 3 weeks of the prototype testing in 2019 (See Appendix A).

There are several algorithms that can be embedded in the operating system with regards to ML, such as linear regression, neural networks, logistic regression, random forest, decision trees, and support vector machines (SVMs). Decision trees are a type of supervised machine learning and can be divided into two major elements, decision nodes, and leaves. The leaves indicate the outcomes of a decision, and the nodes indicate a branch where the data is split. A simple example of a decision tree is to show how a tree grows in a binary regression. The decision nodes are a series of questions like "What major would you like to attend?", "What type of college would you like to attend?" "What do you think your beginning salary should be?". The leaves show the outcomes like "matched" or "mismatched." In the Init2Winit example, we can consider "matched" as a simple binary yes/no classification answer or a continuous classification answer that indicates the distance between desired goal and predictively matched goal.

3.2 Empirical Example: Decision Trees Algorithm in Init2Winit

Using our Init2Winit users as an example, Fig. 6 lays out the decision tree for predicting whether a student's career goal matched their college planning process given the information they obtained and whether their gameplay indicates a matched college degree and/or annual salary projection for the career they plan to pursue. The first decision test was based on the types of college students expect to attend. The sample included the 157 student users in the first job play as an example, 66% had matched college-going planning and 34% were mismatched. The second decision test identified accurate career knowledge of the annual salary in the targeted job. Here we tested the limited node (e.g., focusing on the nodes in the second decision test only) of the aligned college planners, 67% had a matched salary projection

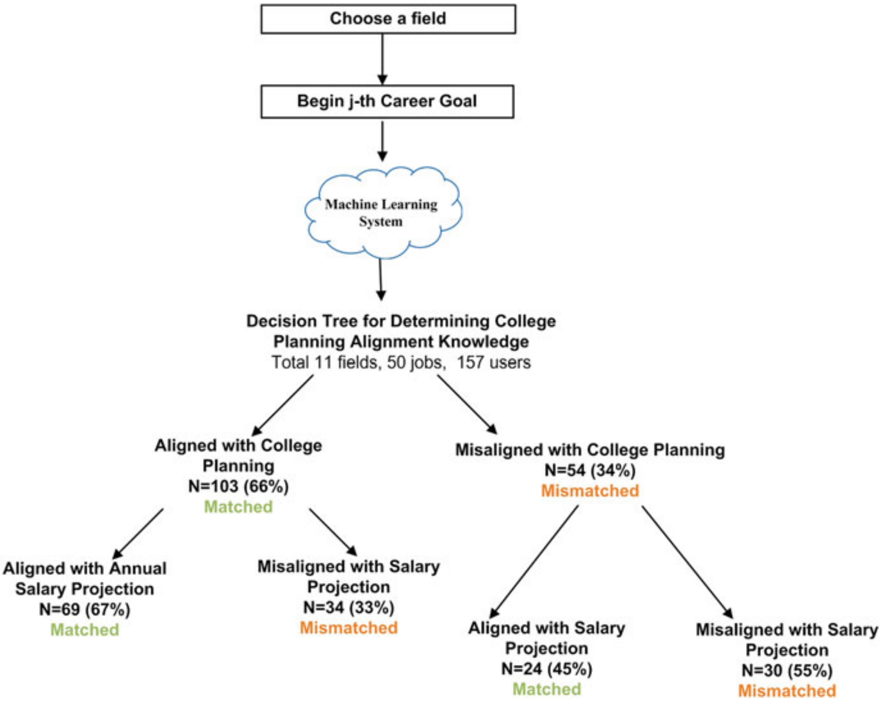


Fig. 6 Hypothetical decision trees using Init2Winit user data in the first job

and 33% were mismatched. On the contrary, when testing the limited node of the misaligned college planners, only 45% had a matched salary projection and 55% were mismatched. This result reflected the fact that users with misalignment knowledge in their college planning had a higher likelihood of having a wrong salary projection as well (55% versus 32%, $Z = 2.67$, $p = 0.0078$).

The decision tree method provides a predictive model in data exploration and training set for machine learning. Our goal is to create a system that models the value of target variables at the leaf of the tree based upon several input variables, including individual users' attributes, at the nodes of the tree. The decision trees in this study aim to identify the probability of a certain alignment results given a desired career choice. This method can also be used for classification and regression. There are several algorithms for decision trees, such as C4.5 (Quinlan 1993), CART (Breiman 2017), BehavDT (Sarker et al. 2019), and IntrudTree (Sarker et al. 2020a). In our example and in our prototype design, we use Iterative Dichotomiser 3 (ID3) algorithm and classification (James et al. 2013; details see Appendix B).

4 Result

4.1 Init2Winit Users' Profiles

Before we used the decision trees to predict users' attributes, we first explored user behavior to obtain prior known classified groups (Sarker 2019). We trained our ML model to be close to the reality of the users' behavior and their intention of exploring career-college planning pathways (Sarker et al. 2019, 2020b). To obtain some prior known classified group, we first looked at behavioral patterns of users' career goal-oriented responses in 3 weeks of playing. We restructured the activity record data into a user-specific data by generating indicators to represent the percent of play frequency in each career field (total 11 fields).

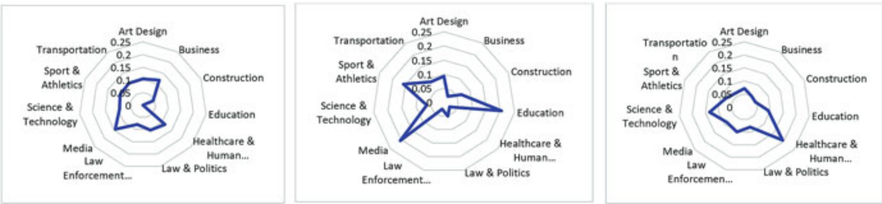
Our data shows that there are three patterns of behavioral career explorations. We named them as solo-goal explorers ($N = 67$), dual-goal explorers ($N = 46$), and multiple-goal explorers ($N = 44$). Solo-goal explorers only explored "one" career field and more than 80% of playing activities happened within one specific field. The top five career explorations for solo-goal users are 22% in Science and Technology careers, 20% in Health care careers, 20% in Business careers, 7% in Sport and Athletics, and another 7% in Media-related careers.

The dual-goal explorers choose only "two" career fields and nearly equal percentages of playing activities occurred between the two fields. For example, Kelly plays Init2Winit 12 times. Among those 12 times of plays, Kelly explores 50% (6 times) of career options in the Business field and another 50% (6 times) of career options in the Science and Technology. The top 3 of college planning and career exploration for dual-goal users are 8% in Business and Sport and Athletics careers, 8% in both Science and Technology and Transportation careers, and another 8% in both Science and Technology and Health care careers.

4.2 Init2Winit Users' Classification for Multiple Goals

The third pattern is the multiple-goal explorers, who explored "more than two" fields of career options. To allow multiple-goals users to explore nonexclusively career goals across 11 fields, we employ multi-label classification method to help classify their orientation in the training set of data. The multi-label classification can identify the association with several classes or labels, which could support mutually exclusive and nonexclusive classes or labels (Bashier et al. 2016; Hall et al. 2016).

Using 676 records in the data streams from 44 users, we built a classification model. After this multi-label classification, three classifications were identified and named: (1) Multiple field 1: Business, Media, and Healthcare ($n = 35$); (2) Multiple field 2: Education, Media, and Sports ($n = 3$); (3) Multiple field 3: Law, Healthcare, and Science Technology ($n = 6$) in Fig. 7. The models and model performance were examined for each classification.



Multiple field 1: Business, Media & Healthcare (n=35)

Multiple field 2: Edu, Media & Sport (n=3)

Multiple field 3: Law, Healthcare & Science Technology (n=6)

Fig. 7 User profiling results of the multi-label classification for multiple-goal explorers

Table 1 shows the descriptive statistics for the behavioral classifications. Parametric t-test and z-test are used to compare the means of two independent samples. In our case, we compare all subgroups with solo-goal explorers. Solo-goal explorers play the Init2Winit about 2 times with an average student GPA of 2.86. This group of explorers also has the highest percent of full alignment knowledge (56%) relative to other explorers (49% or 54%). Dual-goal explorers on average play the Init2Winit about 3 times with an average GPA of 2.81. As Table 1 shows, dual-goal and multiple-field explorers played Init2Winit more frequently than solo-goal users.

Multiple field 1 includes 35 users who mostly explored careers in the Business, Media, and Health fields, with the approximate proportion of playing in each field being 0.12, 0.12, and 0.14. This group of explorers also show interests in Art design, Law, and Sport and Athletics careers. Multiple field 2 includes only three users who mostly explored careers in Education, Media, and Sport and Athletics fields. The approximate proportion of playing in those field is 0.20, 0.20, and 0.16. On average, multiple field 1 explorers play Init2Winit 6 times and multiple field 2 explorers play Init2Winit 8 times. Multiple field 1 explorers play Init2Winit significantly more than sole-goal explorers. Multiple field 3 includes only 6 users who mostly explored careers in Law, Health care, and Science Technology. Students in this group of explorers have significantly higher GPA than solo-goal users ($M = 3.55$ versus $M = 2.86$, $p < 0.05$). Additionally, multiple field 3 explorers have relatively higher number of times played, percent of full alignment knowledge, and level of parents' education compared to other classifications.

4.3 Alignment Knowledge of Decision Trees and Partition

Before applying the tree-based prediction model, we explored the relationship between alignment knowledge and educational expectations after playing Init2Winit (using educational expectations in spring) by partitioning the three behavioral patterns and five career goal-oriented patterns. Due to the small sample size of multiple-field classification, we only report the partition results using the three behavioral patterns. In Fig. 8a, blue dots represent solo-goal explorers, pink dots represent dual-goal explorers, and green dots represent multiple-goal explorers.

Table 1 Descriptive statistics across user profiles in classification (five patterns)

	Solo-goal explorers ($n = 67$)	Dual-goal explorers ($n = 46$)	Multiple field: Business, Media and Healthcare explorers ($n = 35$)	Multiple field: Edu, Media and Sport explorers ($n = 3$)	Multiple field: Law, Healthcare and Science Technology explorers ($n = 6$)
	Mean/SD	**Mean/SD**	**Mean/SD**	**Mean/SD**	**Mean/SD**
Total N of paly[a]	1.97 (1.37)	2.89*** (1.08)	6.69*** (4.28)	8.00*** (3.46)	69.67*** (46.39)
Student GPA	2.86 (0.79)	2.81 (0.81)	2.96 (0.73)	2.79 (0.86)	3.55* (0.32)
Parent education	2.18 (1.28)	2.17 (1.52)	2.26 (1.26)	2.00 (1.00)	2.50 (1.00)
Educational expectation in spring[b]	5.25 (1.32)	4.92 (1.57)	5.35 (1.29)	5.33 (0.57)	5.50 (1.00)
	%	%	%	%	%
Female[c]	0.46	0.33	0.48	0.33	0.50
White	0.20	0.19	0.14	0.33	0.25
Black	0.07	0.10	0.07	0.33	0.25
Asian	0.22	0.36	0.31	0.00	0.25
Others	0.50	0.36	0.48	0.33	0.25*
Percent of no alignment	0.15	0.13	0.09	0.15	0.03
Percent of misalignment	0.29	0.38	0.37	0.29	0.24
Percent of full alignment	0.56	0.49	0.54	0.56	0.74
4-year college planning (yes/no)	0.65	0.60	0.66	0.67	0.50

Note: [a]For continuous variables, two-tailed t-tests compares each classification groups with solo-goal explorers. $*p < 0.05$ $**p < 0.01$ $***p < 0.001$
[b]Educational expectations were measured by students' response to the question, "How far in school do you think you'll get?" in a survey administered in the spring semester of the 2018–2019 after playing Init2Winit. The scale ranges from 1 (less than high school completion) to 7 (complete a Ph.D., M.D., law degree, or other high-level professional degree). A higher value indicated students' higher educational expectations
[c]For categorical variable, two proportion z-test compares each classification group with solo-goal explorers. $*p < 0.5$

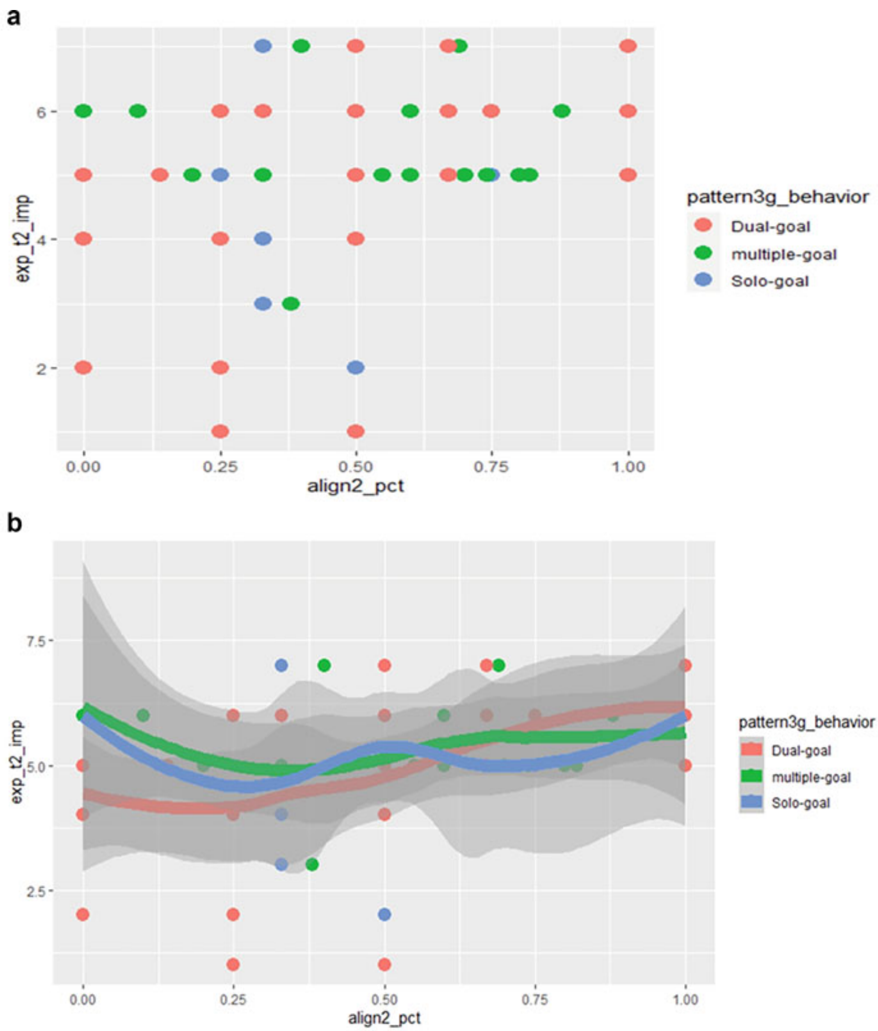


Fig. 8 (**a**) Partition results of three behavioral conditions: Percent of full alignment playing by educational expectations in spring. (**b**) Smoothing partition results of three behavioral conditions: Percent of full alignment playing by educational expectations in spring

The X-coordinate represents the percent of full alignment from the period of playing for 3 weeks, and the Y-coordinate represents users' level of educational expectations in spring. We assume Init2Winit users gain more alignment knowledge during the play, which in turn increases students' educational expectations in spring. We find that solo-goal explorers concentrate in the left middle of the partition space. The linear tendency is low and only happens in the middle level of alignment knowledge and expectations (expectation = 5, percent of alignment = 0.5). Most multiple-goal explorers have relatively higher educational expectations in spring, and the linear tendency is moderate in the upper-right panel of the partition space (expectation > 5, percent of alignment > 0.5). Dual-goal explorers show more variation on the partition space, and the linear tendency is more robust and more responsive to the percent of full alignment knowledge in Fig. 8b. After viewing the partition plot above, we conclude that a regression decision tree is the more appropriate method to estimate our current sample.

4.4 *Regression Decision Trees and Prediction of Educational Expectations*

We then build a regression decision tree using four college-planning and salary-prediction questions in the first two gameplays to predict educational expectations in spring. The results of the regression decision tree have seven terminal nodes as shown in Fig. 9. Each node shows the predicted educational expectations of Init2Winit player in the growing trees and the number of observations from the training dataset located at that node in Table 2.

At the top of Fig. 9, the predicted educational expectations of the overall sample is 5.1. We have 92 users with completed alignment knowledge records in both the first and second careers. The first node asks whether the college planning matched with the first job goal is equal to 0. If no, then the users go down to the right node. The second node asks whether the college planning matched with the second job goal. If no, then the users go down to another right node. If the users have alignment knowledge on those two nodes, then the predicted educational expectations are 5.5 (ranged between a 4-year college degree and a master's degree). In this tree-based

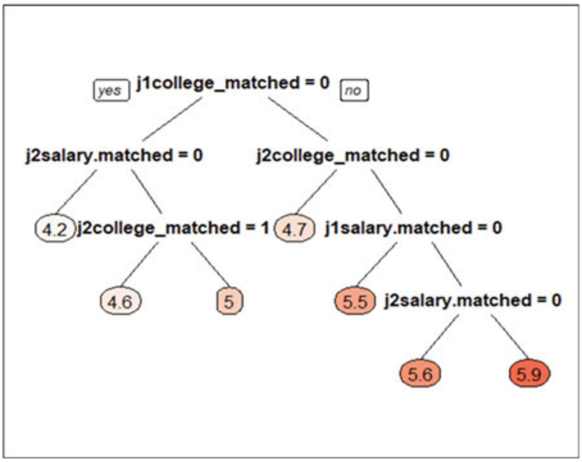


Fig. 9 Decision tree of the alignment knowledge predicted educational expectation in spring

Table 2 Decision tree predicted rules, predicted expectations, and percent of sample

Percent of sample	Predicted educational expectation in spring	Prediction rule
14% ($n = 13$)	4.2	j1college_matched = 0 and j2salary.matched = 0
15% ($n = 14$)	4.6	j1college_matched = 0 and j2college_matched = 1 and j2salary.matched = 1
12% ($n = 11$)	**4.7**	**j1college_matched = 1 and j2college_matched = 0**
8% ($n = 7$)	5.0	j1college_matched = 0 and j2college_matched = 0 and j2salary.matched = 1
21% ($n = 19$)	**5.5**	**j1college_matched = 1 and j2college_matched = 1 and j1salary.matched = 0**
9% ($n = 8$)	5.6	j1college_matched = 1 and j2college_matched = 1 and j2salary.matched = 0 and j1salary.matched = 1
22% ($n = 20$)	5.9	j1college_matched = 1 and j2college_matched = 1 and j2salary.matched = 1 and j1salary.matched = 1

Note: Bold indicates an example we described in the main text

model, 19 users belong to this pathway. If the users did not have a matched college planning knowledge in the second job goal, the predicted educational expectations are 4.7 (ranged between some college and a 4-year college degree). We have 11 users who belong to this pathway. Our tree could grow and help us understand which primary alignment knowledge (college planning or salary prediction) impacts educational expectations prediction more.

To evaluate the prediction performance of the tree-based model, we split the current sample randomly by an 8:2 ratio into the training and testing sets. Then, we train our model on the training set and tested it. We used the averaged F1-score to measure the overall performance of the algorithm (Lipton et al. 2014). The F1 score is a weighted average of the precision rate for recall. The range of an F1 score is 0–1. Our current model has a F1 score of 0.72 using four college planning to salary prediction questions in the first two jobs of gameplay. We can increase the prediction performance to 0.85 by including more variables and questions, such as GPA, parent education, and students' characteristics. We report the simplest results in current study because the inclusion of more variables in the tree-based model also increases the number of missing cases (other decision trees results are available upon request).

5 Strengths and Weakness of Current Design

One of the strengths of the current design is the simplicity of the design and the effectiveness. The simplicity is the increase in students' alignment by playing the career exploration tunnel in the Init2Winit. The effectiveness is in predicting how student alignment knowledge corresponds to their educational expectations after game playing through the use of a decision tree. Using this prediction, the importance of increasing students' alignment knowledge and leading to increasing educational expectations after game playing becomes clear. Importantly, this prediction does not require a lot of users' background information or covariates but can still provide valuable data insights with a high prediction level. This feature is very useful with data where background data is not available or where there is over 10% of missing data.

Additionally, embedding the machine learning and decision tree algorithm in a mobile application is also quite useful with respect to users becoming more informed by the optimization students' college planning or forecasting the success rates for various career goals. Users' behavior patterns and goal-oriented explorations can also profile the individual's motivation and preparedness based upon a predetermined classification analysis. However, this design also leaves several open questions surrounding the factors which drive students' misalignment in their career/college knowledge, how to distinguish higher scorers between playing within the same career options versus playing across multiple career options, and the genuine learners of alignment.

The decision tree, as one of the simplest ML models, could incorporate several different functions to account for complex data structure and conditions, such as boosting when there is high variance in the outcomes. However, this method also has some limitations. First, decision trees are less efficient in estimation compared to other supervised ML methods, especially in big trees where increasing efficiency results in poor prediction accuracy (James et al. 2013). Second, large decision tree models cause high complexity in processing the data, increasing computation time, and difficulties in converging. More advanced methods, such as random forest, neural network, and support vector machines (SVMs), can be more computationally effective and handle nonlinear patterns and large samples (Puterman 2014). Third, the prediction of decision trees generally does not have comparable accuracy rate to other approaches, especially in a small sample (Wu et al. 2016).

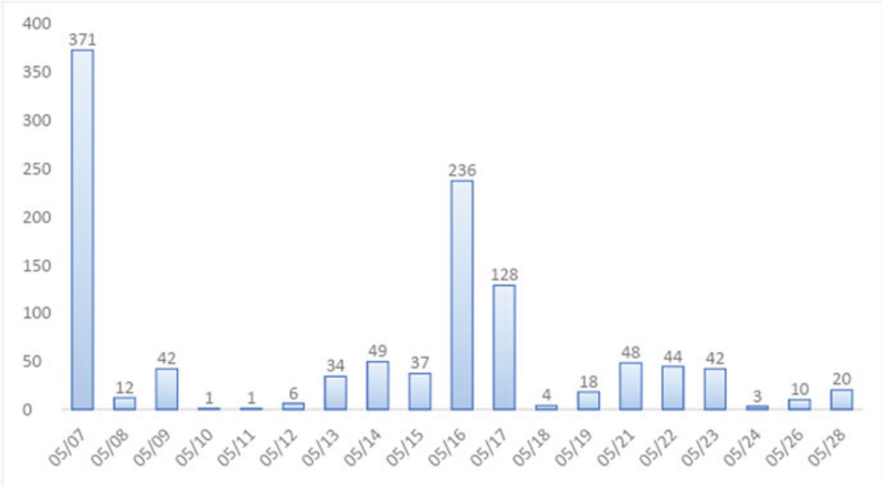
6 Conclusion and Recommendation

This study develops and tests the AI features of machine learning in Init2Winit, using the decision tree-based method, to identify users' usage behavior, goal-oriented patterns, and prediction of future educational expectations. Our results show promise in terms of the prediction accuracy of educational expectations and users' behavioral classifications. Beyond this, machine learning could incorporate a game designed to measure students' strengths and weaknesses to give career recommendations and pathways. Init2winit can be an informational channel for low-income students who lack informal networks or whose parents have not earned college degrees. It also serves as a supplementary network supporting career/ college planning knowledge for students to make better education and employment decisions. This study is just one example of how AI and machine learning can help students explore careers and increase their educational aspirations and college-going choices. It shows how a mobile application can be built upon previous theory (alignment theory) to increase students' knowledge and educational expectations and to further flag students who may be mismatched, misaligned, or disoriented in their planning and decision-making for college and career choice.

The study has three primary goals, each of which informs the alignment theory of career-to-college explorations and applies efforts to strengthen the pipeline of STEM careers during high schools. First, we develop a mobile application Init2Winit to test theoretical assumptions about alignment knowledge. Second, we compare students' goal exploration behavior, orientation, and profile, which are important in shaping career choices and college decisions. Third, we provide data insights for school counselors, parents, and students to optimize their choices and college plans. Altogether, our study evaluates and recommends an outlook of Init2Winit in the coming decades.

We propose a few steps that should be considered to ensure that all students are served and provided with the information and social capital needed for college readiness and planning. The first suggestion is to consider the ways in which school counselors and homeroom teachers serve as role models and informational hubs in the lives of many students through the use of mobile technology and its applications. Teachers' participation can facilitate parents and students' knowledge, using machine learning to improve users' personalized decision-making (Thompson and Subich 2006). Another suggestion is to provide students with a real-time intervention and guidance even in resources-restricted schools. Educational technology can provide unlimited access to information and data feedback based on student usage behavior, goal-oriented profiles, and response patterns. Fundamentally, our goal is to use AI technology to formulate more realistic engaging tasks and scoring procedures that can provide improved college knowledge and career aspiration for students, their parents, and school professionals. The goal here is efficiency but not at the expense of students' interests or in trying to force career choice too early in a young person's life.

Appendix A Record of e Daily Active Users



Note: The numbers in each bar represent the total number of individual users per day.

Appendix B Iterative Dichotomiser 3 (ID3) Algorithm

The ID3 algorithm uses the most significant information gain after splitting the measure to partition the outcome and make each branch belong to the same classification. The criteria to separate the node is the Gini impurity and "entropy" for the information gain. Entropy measures the discriminatory power of an attribute in the classification task. It defines the amount of randomness in the attribution of classification or regression. Gini and Gini impurity are used to decide the best split. Gini ranges from 0–1. The higher the Gini coefficient, the more different instances within the node.

$$\text{Entropy}: H(S) = -\sum\nolimits_{i=1}^{n} p\left(x_i\right)\ \log_2\ p\left(x_i\right) \tag{1}$$

$$\text{Gini}\ (E) = 1 - \sum\nolimits_{I=1}^{c} p_i^2 \tag{2}$$

Information gain defines as a set of S, which are effective changes in entropy after deciding on a particular attribute or goal. Information gain measures the relative changes in entropy conditional on the independent variables in the tree. A training set S could be a positive or a negative example. The indicates the probability of event *x*. Our goal is to use this method to train the machine to classify users' response patterns and provide predictive data insights for students and school counselors.

References

Arntzi M, Gregoryi T, Zierahni U. (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. Paris: OECD.

Albion MJ, Fogarty GJ. (2002). Factors Influencing Career Decision Making in Adolescents and Adults. Journal of Career Assessment, 10(1), 91–126. doi: https://doi.org/10.1177/1069072702010001006

Baker, T., Smith, L., & Anissa, N. (2019). Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. Retrieved from nesta.org.uk website: https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf

Bashier, E. B. M., Mohammed, M., & Khan, M. B. (2016). Machine learning: algorithms and applications. CRC Press.

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. The Quarterly Journal of Economics, 127(3), 1205–1242. doi: https://doi.org/10.1093/qje/qjs017

Blustein, D. L., Duffy, R., Ferreira, J. A., Cohen-Scali, V., Cinamon, R. G., & Allan, B. A. (2020a). Unemployment in the time of COVID-19: A research agenda. Journal of Vocational Behavior, 119, 103436. doi: https://doi.org/10.1016/j.jvb.2020.103436

Blustein, D. L., Erby, W., Meerkins, T., Soldz, I., & Ezema, G. N. (2020b). A Critical Exploration of Assumptions Underlying STEM Career Development. Journal of Career Development, 0894845320974449. doi: https://doi.org/10.1177/0894845320974449

Breiman, L. (2017). Classification and regression trees: CRC Press.

Bureau of Labor Statistics. Occupational employment projections 2014-2024. Monthly Labor Review. 2015. https://www.bls.gov/opub/mlr/2015/article/occupational-employment-projections-to-2024.htm

Castleman, B., & Goodman, J. (2018). Intensive College Counseling and the Enrollment and Persistence of Low-Income Students. Education Finance and Policy, 13(1), 19–41. doi: https://doi.org/10.1162/edfp_a_00204

Chen, I. C., Rocha-Beverly, C., & Schneider, B. (2020). Alignment of educational aspirations and career plans in high school with Mobile app technology. Education and Information Technologies, 1–19

Chen, I. C., Rocha-Beverly, C., & Schneider, B (2021) Learning by playing Init2Winit: How alignment knowledge increases educational aspirations and college plans in high school, Journal of Research on Technology in Education, doi: https://doi.org/10.1080/15391523.2021.1877225

Clark, B., Joubert, C., & Maurel, A. (2017). The career prospects of overeducated Americans. IZA Journal of Labor Economics, 6(1), 3. doi: https://doi.org/10.1186/s40172-017-0053-4

Cohodes, S. R., & Goodman, J. S. (2014). Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy. American Economic Journal: Applied Economics, 6(4), 251–285. doi: https://doi.org/10.1257/app.6.4.251

James, G M., Witten, D., Hastie, T., & Tibshirani, R. R. Tree-Based Methods. In: Casella G FS, Olkin I, ed. An introduction to statistical learning – with applications in R. New York: Springer; 2013:373–418.

Hall, M. A., Witten, I. H., Frank, E., & Pal, C. (2016). Data Mining: Practical machine learning tools and techniques. San Francisco: Morgan Kaufmann.

Hoff, K., Van Egdom, D., Napolitano, C., Hanna, A., & Rounds, J. (2021). Dream Jobs and Employment Realities: How Adolescents' Career Aspirations Compare to Labor Demands and Automation Risks. Journal of Career Assessment. doi: https://doi.org/10.1177/10690727211026183

Kena G, Hussar W, McFarland J, Brey C, Musu-Gillette L. (2016). The condition of education 2016 (NCES 2016-144). Washington, DC: U.S. Department of Education, National Center for Education Statistics. 2016. Retrieved from https://nces.ed.gov/pubs2016/2016144.pdf

Kim, S., Klager, C., & Schneider, B. (2019). The Effects of Alignment of Educational Expectations and Occupational Aspirations on Labor Market Outcomes: Evidence from NLSY79. The Journal of Higher Education, 90(6), 992–1015. doi: https://doi.org/10.1080/00221546.2019.1615333

Lipton Z.C., Elkan C., Naryanaswamy B. Optimal Thresholding of Classifiers to Maximize F1 Measure. In: Calders T., Esposito F., Hüllermeier E., Meo R. (eds) Machine Learning and Knowledge Discovery in Databases. 2014. ECML PKDD. Lecture Notes in Computer Science, vol 8725. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44851-9_15

Mann, A., Denis, V., Schleicher, A., Ekhtiari, H., Forsyth, T., Liu, E., & Chambers, N. (2020). Teenagers'career aspirations and the future of work: OECD.

Morgan, S. L., Leenman, T. S., Todd, J. J., & Weeden, K. A. (2013). Occupational Plans, Beliefs about Educational Requirements, and Patterns of College Entry. Sociology of Education., 86(3), 197–217. doi: https://doi.org/10.1177/0038040712456559

National Center for O*NET Development. (2019). Job zone. O*NET OnLine. Retrieved from https://www.onetonline.org/find/zone

Nedelkoska, L., & Quintini, G. (2018). Automation, skills use and training OECD.

Nie, M., Xiong, Z., Zhong, R., Deng, W., & Yang, G. (2020). Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students. Applied Sciences, 10(8). doi: https://doi.org/10.3390/app10082841

Nunley, J. M., Pugh, A., Romero, N., & Seals, R. A., Jr. (2016). College major, internship experience, and employment opportunities: Estimates from a resume audit. LABOUR ECONOMICS, 38, 37–46. doi: https://doi.org/10.1016/j.labeco.2015.11.002

Perry, B. L., Martinez, E., Morris, E., Link, T. C., & Leukefeld, C. (2016). Misalignment of Career and Educational Aspirations in Middle School: Differences across Race, Ethnicity, and Socioeconomic Status. Social sciences (Basel, Switzerland), 5(3). doi: https://doi.org/10.3390/socsci5030035

Pinto, L. H., & Ramalheira, D. C. (2017). Perceived employability of business graduates: The effect of academic performance and extracurricular activities. Journal of Vocational Behavior, 99, 165–178. doi: https://doi.org/10.1016/j.jvb.2017.01.005

Puterman, M.L. (2014). Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons.

Quinlan, J.R. (1993). C4.5: programs for machine learning. Mach Learn.

Reyna, V. F., & Farley, F. (2006). Risk and Rationality in Adolescent Decision Making. Psychological Science in the Public Interest, 7, 1–44.

Renbarger, R., & Long, K. (2019). Interventions for Postsecondary Success for Low-Income and High-Potential Students: A Systematic Review. JOURNAL OF ADVANCED ACADEMICS, 30(2), 178–202. doi: https://doi.org/10.1177/1932202X19828744

Sarker, I. H. (2019). A machine learning based robust prediction model for real-life mobile phone data. Internet Of Things, 5, 180–193. doi: https://doi.org/10.1016/j.iot.2019.01.007

Sarker, I. H., Colman, A., Han, J., Khan, A. I., Abushark, Y. B., & Salah, K. (2020a). BehavDT: A Behavioral Decision Tree Learning to Build User-Centric Context-Aware Predictive Model. Mobile Networks & Applications, 25(3), 1151–1161. doi: https://doi.org/10.1007/s11036-019-01443-z

Sarker, I. H., Alqahtani, H., Alsolami, F., Khan, A. I., Abushark, Y. B., & Siddiqui, M. K. (2020b). Context pre-modeling: an empirical analysis for classification based user-centric context-aware predictive modeling. Journal Of Big Data, 7(1). doi: https://doi.org/10.1186/s40537-020-00328-3

Sarker, I. H., Kayes, A. S. M., & Watters, P. (2019). Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. Journal Of Big Data, 6(1). doi: https://doi.org/10.1186/s40537-019-0219-y

Schumacher, P., Olinsky, A., Quinn, J., & Smith, R. (2010). A Comparison of Logistic Regression, Neural Networks, and Classification Trees Predicting Success of Actuarial Students. Journal of Education for Business, 85(5), 258–263

Schmitt-Wilson, S., & Faas, C. (2016). Alignment of Educational and Occupational Expectations Influences on Young Adult Educational Attainment, Income, and Underemployment. SOCIAL SCIENCE QUARTERLY, 97(5), 1174–1188. doi: https://doi.org/10.1111/ssqu.12244

Schneider, B., & Stevenson, D. (1999). The Ambitious Generation: America's Teenagers, Motivated but Directionless. New York: Yale University Press.

Schneider, B. (2009). Challenges of transitioning into adulthood. In I. S. R. K. Silbereisen (Ed.), Transitions from school to work: Globalization, individualization, and patterns of diversity (pp. 265–290). Cambridge, England: Cambridge University Press.

Schneider, B., & Young, L. (2019). Advancing workforce readiness among low-income and minority high school students. In F. L. Oswald, Behrend, T. S., & Foster, L. L. (Ed.), Workforce readiness and the future of work. New York: Taylor & Francis.

Thompson, M. N., & Subich, L. M. (2006). The relation of social status to the career decision-making process. Journal of Vocational Behavior, 69(2), 289–301. doi: https://doi.org/10.1016/j.jvb.2006.04.008

Wu, C.-C., Chen, Y.-L., Liu, Y.-H., & Yang, X.-Y. (2016). Decision tree induction with a constrained number of leaf nodes. Applied Intelligence, 45(3), 673–685. doi: https://doi.org/10.1007/s10489-016-0785-z