J. Res. Technol. Eng. 6 (3), 2025, 81-92

Journal Of Research Technology & Engineering
www.jrte.org

Exploring the Interplay Between Academic Background, Digital Literacy, and Career Aspirations Among Pre-University Students in Sri Lanka

*Akarshani Amarasinghe[1], NDDA Nanayakkara[2], K.G.H.H.R. Kriella[3], B. H. C. Shiromali[4], D. Senthilnathan[5], Hiruni Liyanage[6], Dilshan Hettiarachchi[7], Ronalee Newandi[8], *Dian Baduge[9]

[1, 2]Department of Computer Engineering, Faculty of Engineering, University of Sri Jayewardenepura
[3, 4]Department of Manpower and Employment
[5]National Humans Resources Development Council of Sri Lanka
[6, 7, 8, 9]Future Lanka Research and Development Foundation
*akarshani.amarasinghe@sjp.ac.lk, *dian@futurelanka.lk

Received:02 July 2025; Revised:05 July 2025; Accepted: 10 July 225; Available online: 10 July 2025

Abstract: This study investigates the influence of English language proficiency, academic stream, and digital skills on the higher education preferences and career aspirations of pre-university students in Sri Lanka. Using data mining techniques in WEKA 3.8.6, we applied the Apriori, RandomTree, and REPTree algorithms to explore patterns and predict educational outcomes. The Apriori algorithm identified a key association between beginner-level English proficiency, a preference for state universities, and the lack of coding skills. RandomTree models highlighted the significance of English proficiency as a determinant in academic paths, with fluent and native speakers displaying greater flexibility in their university choices. REPTree further confirmed that English language proficiency and coding ability were the most influential factors in shaping students' higher education trajectories, emphasizing the limitations faced by those with lower proficiency levels. Our findings suggest that enhancing digital literacy and language skills among students can promote more equitable access to diverse academic and career opportunities. The study offers actionable insights for educators and policymakers to design targeted interventions that support students in navigating their academic and professional futures. Future research could incorporate longitudinal data, psychosocial factors, and advanced machine learning techniques to further refine predictive models and foster more inclusive educational systems.

Index Terms: Coding ability, Digital literacy, English language proficiency, Higher education preferences

1 INTRODUCTION

In the rapidly evolving global labor market, the alignment of students' academic choices, digital competencies, and career preferences plays a critical role in shaping future employment landscapes. In countries like Sri Lanka, where traditional education systems are increasingly interfacing with global technological trends, understanding students' readiness for the modern workforce is essential. Pre-university students often face a crossroads that involves not only selecting a career path but also aligning that choice with their skills, interests, and socio-cultural expectations.

This study aims to explore how students' Advanced Level (A/L) academic streams, coding abilities, and English language proficiency influence their preferences for job sectors and higher education institutions. By examining these variables, we seek to identify potential gaps in digital literacy and higher education preparedness that may impact employability and career success. The findings are intended to inform educators, policymakers, and career guidance practitioners on how to better support students in making informed decisions in a technology-driven world.

2 Literature Review

The intersection of education, technology, and career planning has garnered growing scholarly interest over the past decade. Several studies have highlighted the role of academic streams in determining students' career paths. For instance, research by De Alwis and Wijesekara emphasized that Sri Lankan students in science streams are more likely to pursue professional or overseas education due to better perceived employability [1].

Digital literacy, particularly coding ability, has become a critical skill in the 21st century job market. Papert suggested that coding not only improves problem-solving and logical thinking but also enhances creativity and self-expression [2]. In more recent studies, the presence or absence of coding knowledge significantly influenced students' inclination toward technology-driven careers [3].

English language proficiency also remains a key determinant of academic and professional opportunities, particularly in non-native English-speaking countries. Research conducted by Rameez found that Sri Lankan undergraduates with advanced English skills were more likely to secure jobs in foreign or multinational companies [4].

Moreover, the decision-making process regarding higher education is influenced by a combination of personal aspirations, socio-economic background, and perceived value of institutions. Ranwala and the team suggest that students from privileged backgrounds tend to favor foreign or private universities, viewing them as gateways to global employment markets [5].

Despite these insights, there remains a dearth of integrated studies examining how these factors interact to influence the career trajectories of Sri Lankan students. This research seeks to fill that gap by analyzing a preprocessed dataset to draw meaningful correlations and recommendations.

3 Methodology and Data Analysis

This study employs a data-driven approach to investigate the interrelationships among students' academic backgrounds, language proficiency, digital skills, and educational and occupational preferences. The analysis was conducted using the WEKA 3.8.6 machine learning environment, a widely adopted tool for data mining and pattern recognition.

3.1 Dataset and Preprocessing

The dataset comprises records from pre-university students, capturing attributes such as A/L Stream, English Language Proficiency, Coding Ability (binary: Yes/No), Job Preference, and Preference for Higher

Education. Prior to analysis, the data was preprocessed to ensure consistency, remove redundancies, and convert categorical values into nominal formats compatible with WEKA's algorithms.

3.2 Applied Algorithms

Three algorithms were applied to analyze the dataset:

- **Apriori Algorithm:** A popular rule-based association learning technique used to uncover interesting relationships between categorical variables [6].
- **RandomTree:** A decision tree classifier that builds a tree using a random subset of features at each node, offering robustness in dealing with high-dimensional categorical data [7].
- **REPTree:** A fast decision tree learner that builds a regression/decision tree using information gain or variance reduction and prunes it using reduced-error pruning [8].

These algorithms were selected due to their interpretability and ability to reveal both global patterns (association rules) and predictive structures (classification trees) in educational datasets.

3.3 Association Rule Mining with Apriori

Apriori analysis yielded several rules, with the most significant rule being (Equation (1)):

$$\begin{aligned} &\textit{English Language Proficiency} = \textit{Beginner} \\ &\textit{Preference for Higher Education} = \textit{State University} \\ &\Rightarrow \textit{Can You Code?} = \textit{No} \end{aligned} \tag{1}$$

This rule implies a strong association between limited English proficiency, preference for state university education, and lack of coding ability. The result underscores a key insight: students with beginner-level English skills who favor local public institutions are statistically more likely to lack programming experience.

This finding may reflect broader educational inequities; students from under-resourced schools or non-English-medium backgrounds may have limited access to extracurricular ICT training or coding programs, thereby influencing their career readiness in tech-driven fields.

3. 4 Decision Tree Analysis Using RandomTree

To gain deeper insights into how English language proficiency influences students' educational and occupational preferences, RandomTree classification models were generated for each proficiency level separately. This approach enables the isolation of linguistic factors and their intersection with other variables such as coding skills, A/L stream, and job preferences. English language proficiency becomes the root of all decision trees.

- **Advanced Proficiency:** As illustrated in Fig. 1, students with Advanced English proficiency tend to exhibit a structured path toward professional and academic advancement. The tree reveals that:
    - Job Preference plays a critical role, with those opting for Private Jobs further split based on coding skills.
    - Coding ability is highly predictive: those who can code are channeled primarily toward State Universities, while those who cannot tend to be sorted based on A/L stream.

- For Foreign Job seekers, A/L stream and coding again influence the choice between Private Universities and Professional Institutions.
- Government Job seekers converge uniformly toward State Universities, regardless of coding ability.

This suggests that students with advanced English skills have diversified but structurally predictable academic pathways, especially influenced by their technological proficiency.

- **Beginner Proficiency:** Students in the Beginner category demonstrate the most rigid and narrow academic outcomes (Fig. 2):
  - Foreign Job seekers who cannot code are routed toward State Universities, across all A/L streams.
  - Those who can code have slightly more diverse options but still show heavy concentration in State and Professional Institutions.
  - In the Private Job track, regardless of A/L stream, coding ability, or preferences, students are typically funneled into State or Professional Institutions.
  - The model reveals minimal flexibility, indicating that limited language proficiency severely restricts career and educational diversification.

  This reflects a systemic constraint wherein low English proficiency may be a barrier to accessing private or foreign educational institutions and technology-driven careers.

- **Fluent Proficiency:** Students classified as Fluent show more individualized educational pathways:
  - For the Bio and Physical Science streams, coding ability strongly determines whether students attend State or Professional Institutions.
  - Those in Commerce show the widest diversity in job preference and educational pathways, with Foreign, Private, and Government job preferences each leading to different outcomes depending on coding ability.
  - ICT and Tech streams are consistently associated with Professional Institutions, aligning with vocational or hands-on skill development trajectories.

  As shown in Fig. 3, the tree structure indicates that while fluency opens up multiple routes, coding remains a key determinant in navigating higher education options effectively.

- **Intermediate Proficiency:** As represented in Fig. 4, with Intermediate proficiency, student decisions begin to stratify based on coding ability but within a narrower band compared to fluent users:
  - Foreign Job seekers are split sharply based on coding skills, with coders accessing Private Universities and non-coders often assigned to State or Vocational Institutions.
  - Private Job seekers in nearly all A/L streams are streamed toward State Universities, with some limited access to Professional Institutions.
  - The most consistent assignment occurs in Government Job preferences, where almost all routes lead to State Universities, regardless of coding.

  This group is at a pivotal stage; language proficiency enables some diversification, but other skills (especially coding) significantly shape outcomes.
- **Native Proficiency:** For Native English speakers, the decision tree is streamlined (Fig. 5):
  - Those who can code are sorted into job preferences, which then determine educational

placement in Foreign, Professional, or Government Institutions.

- Interestingly, those who cannot code are funneled directly to Foreign Universities, suggesting that language fluency alone can act as a gateway to international education, even without technical skills.

This cohort shows the least restriction in educational mobility, illustrating that native fluency provides a competitive edge in accessing global academic and career opportunities.

3.5 Decision Tree Analysis Using REPTree

To enhance interpretability and evaluate the influence of specific educational and personal factors on students' higher education preferences, the REPTree algorithm was employed. REPTree, a fast and efficient decision tree learner in WEKA, uses information gain for node splitting and applies reduced-error pruning to improve generalization. The resulting model is depicted in Fig. 6.

The decision tree begins with English Language Proficiency as the root attribute, highlighting its primary influence on students' academic direction. Each branch corresponds to one of the predefined proficiency levels: Advanced, Beginner, Fluent, Intermediate, and Native.

- **Beginner Proficiency:** Students with beginner-level English proficiency are further classified based on their A/L Stream. Regardless of specialization (Bio, Physical Science, Commerce, Art, Tech, ICT, or Law), these students are overwhelmingly directed toward State Universities. For students in the Bio stream, a secondary analysis is conducted based on Job Preference and Coding Ability:
  - Bio students preferring Foreign Jobs and possessing coding skills are eligible for Private or Non-State Universities, while those without coding skills are recommended State Universities.
  - Those aiming for Private Jobs are steered toward Private or Non-State Universities, and Government Job aspirants return to State University options.
- **Intermediate and Advanced Proficiency:** These groups are assigned directly to State Universities, indicating that while language proficiency may enhance academic pathways, moderate fluency still largely confines students to public institutions.
- **Fluent Proficiency:** Students with fluent English are channeled toward Private or Non-State Universities, reflecting an elevated likelihood of selecting alternate or international educational tracks.
- **Native Proficiency:** Students with native-level English proficiency are categorized toward Professional Institutions, suggesting a preference for skill-based or applied learning environments over traditional academia.

The REPTree output reinforces the observation that English proficiency is a dominant determinant of educational trajectory. Moreover, it illustrates nuanced distinctions based on coding ability and career preference, particularly for students with lower proficiency. The tree's compactness compared to RandomTree also enhances interpretability while preserving predictive insight.

This decision logic can inform educational counseling strategies by highlighting key intervention points; such as promoting coding literacy among beginner-level English speakers or guiding fluent students toward broader institutional choices.

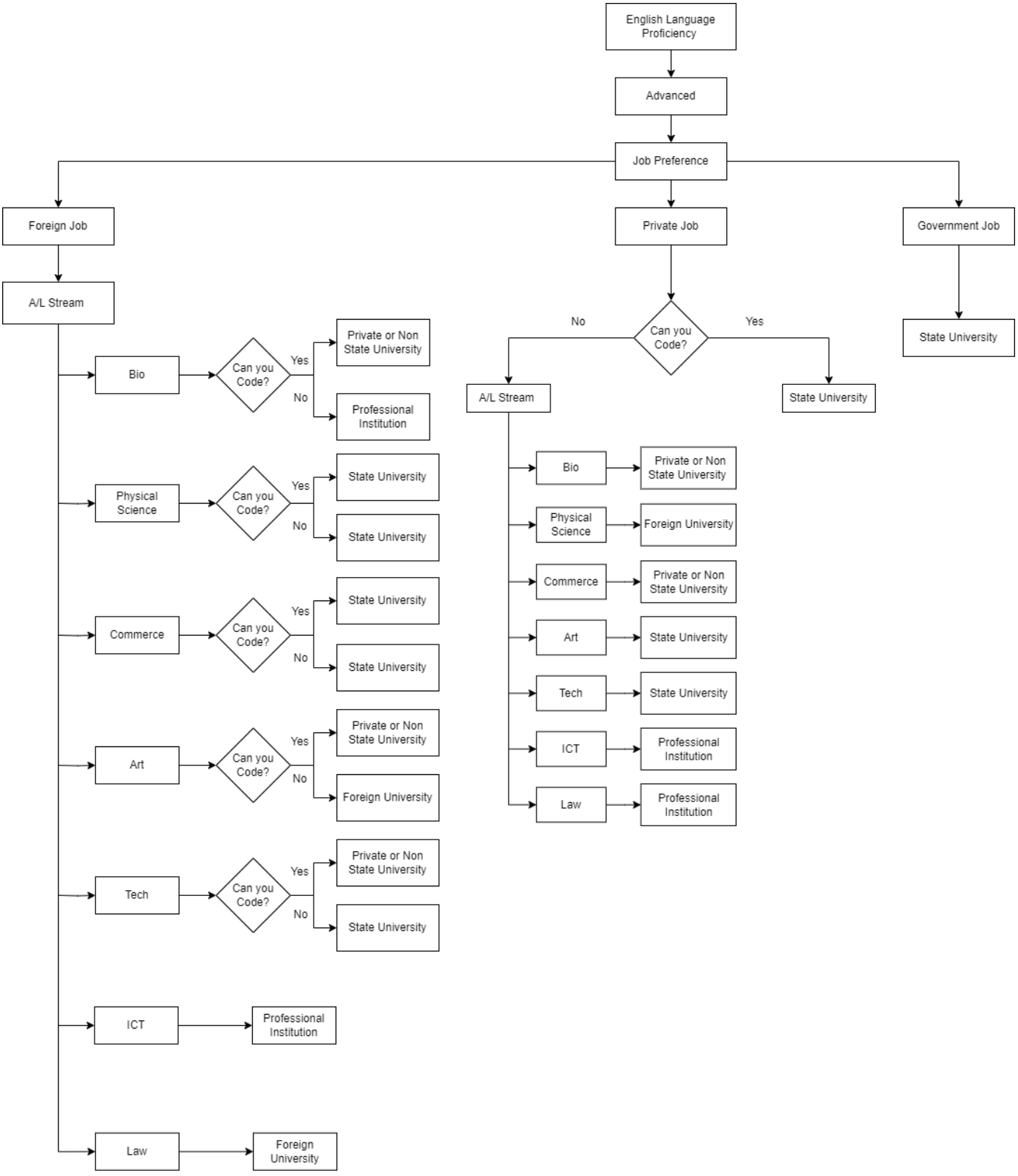

Fig. 1. RandomTree decision path for Advanced English language proficiency

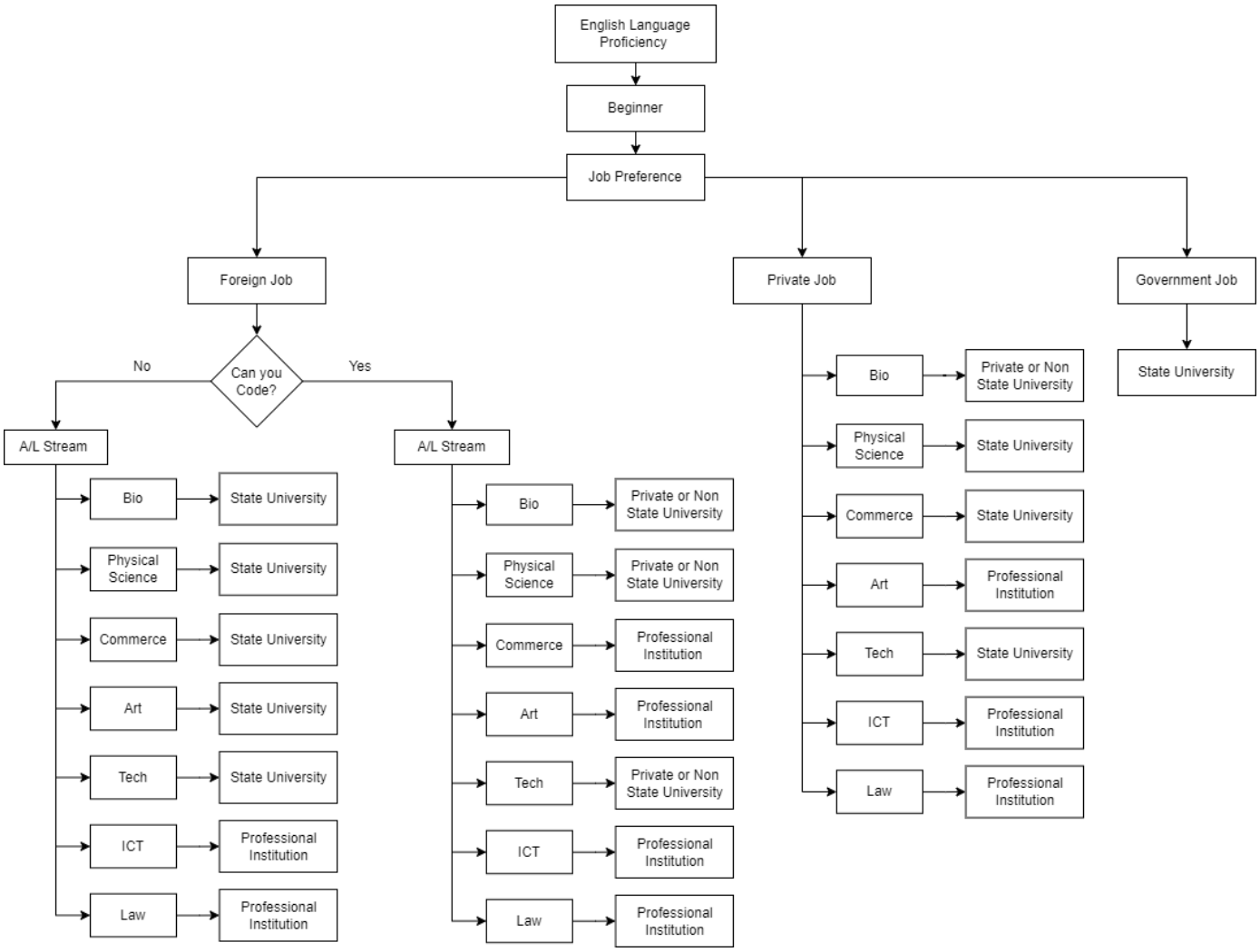

Fig. 2. RandomTree decision path for Beginner English language proficiency

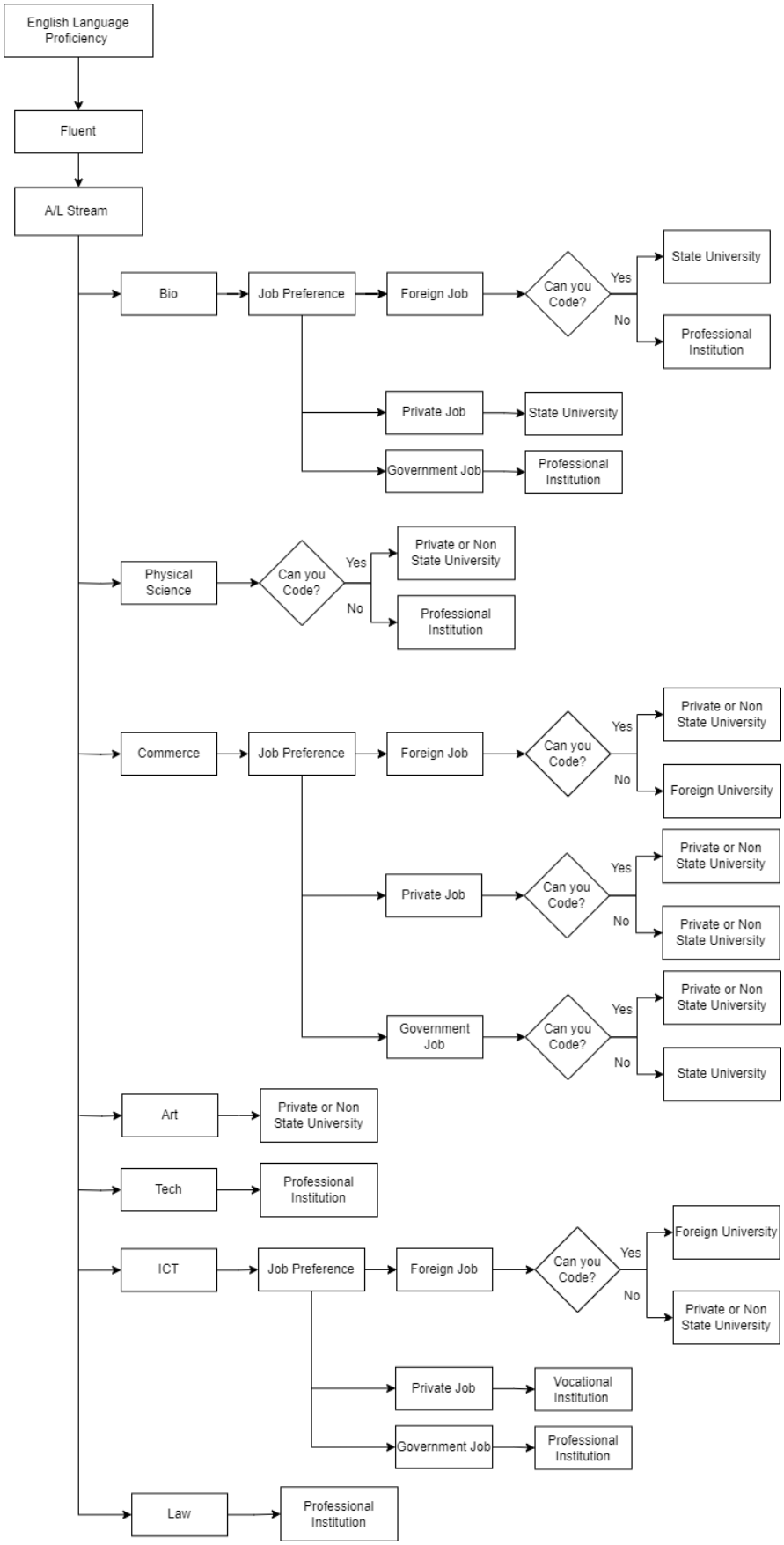

Fig. 3. RandomTree decision path for Fluent English language proficiency

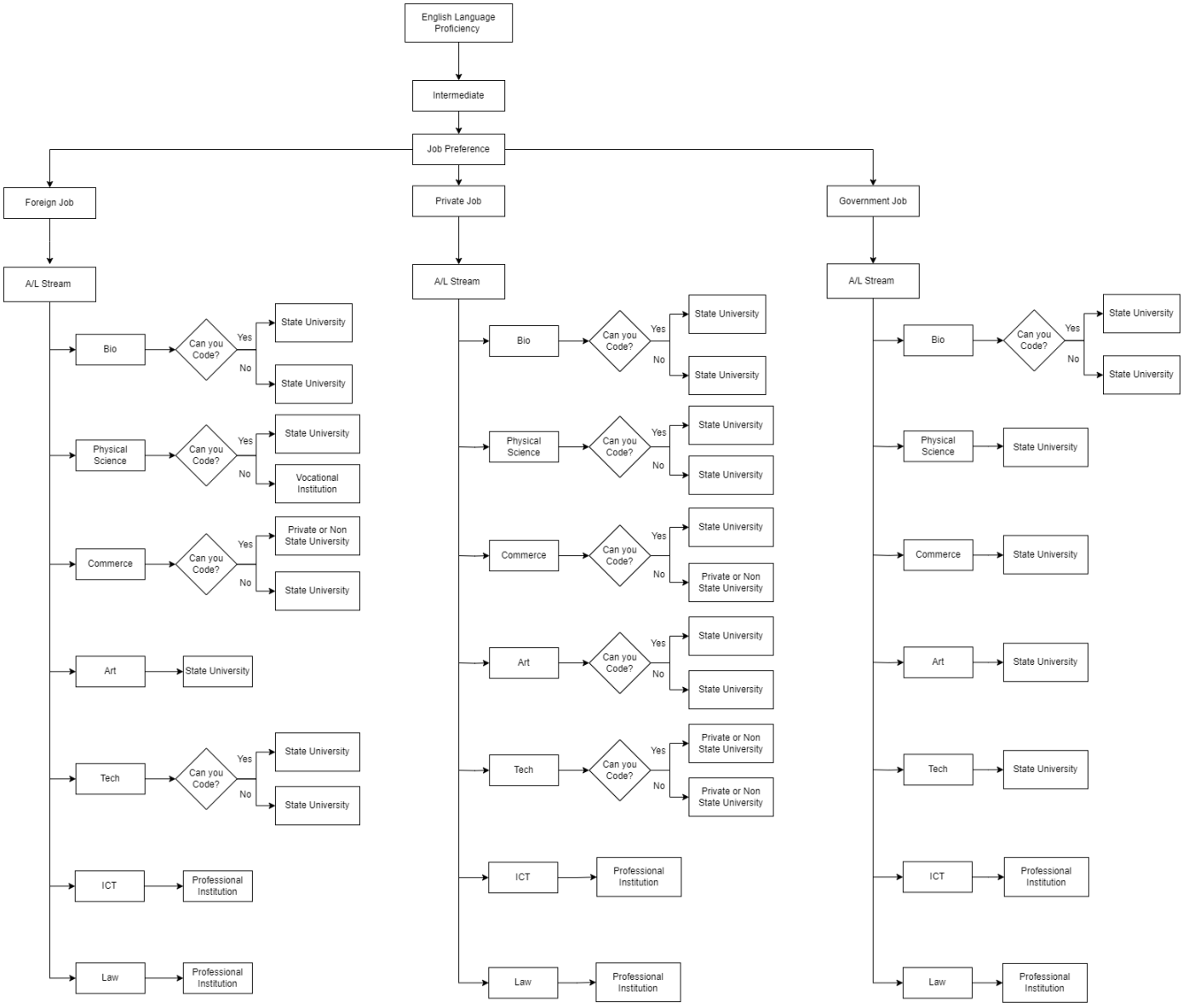

Fig. 4. RandomTree decision path for Intermediate English language proficiency

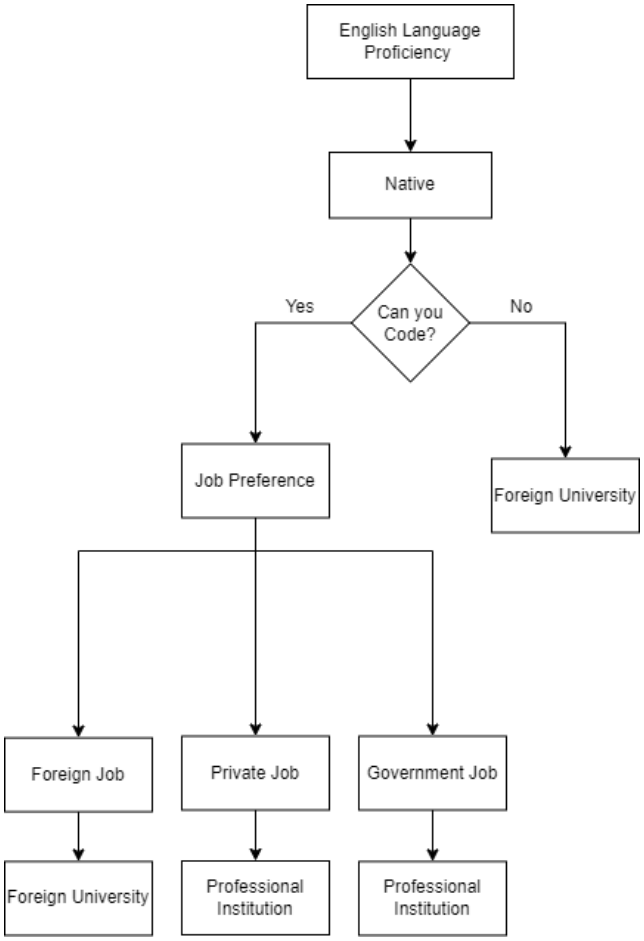

Fig. 5. RandomTree decision path for Native English language proficiency

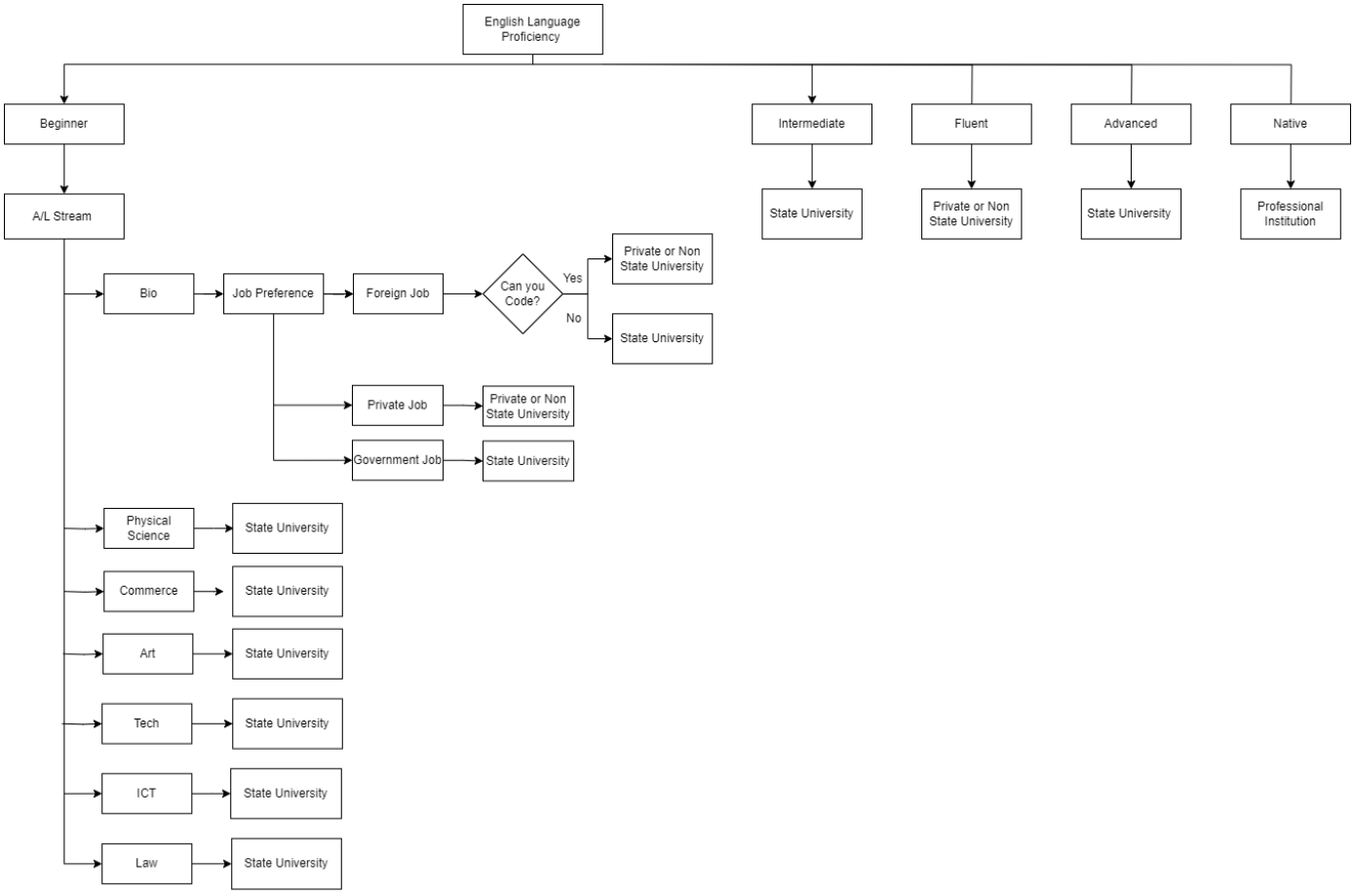

Fig. 6. REPTree decision model for higher education preference based on English language proficiency, A/L stream, job preference, and coding ability

4 Conclusion and Future Work

This study explored the interrelationship between Sri Lankan pre-university students' academic backgrounds, digital skills, English language proficiency, and their preferences for higher education institutions. Using WEKA 3.8.6, three machine learning algorithms: Apriori, RandomTree, and REPTree were employed to uncover associative patterns and decision rules within the dataset.

The Apriori algorithm identified a strong association between students with beginner-level English proficiency who prefer state universities and their inability to code. This rule suggests a critical dependency between language and digital literacy, with potential implications for workforce readiness.

RandomTree models, customized by English language proficiency level, revealed how career and educational paths diverge dramatically based on both linguistic and technological skills. Students with fluent or native proficiency enjoyed broader and more flexible academic pathways, while those with beginner or intermediate skills were often routed into state university systems, regardless of A/L stream or job aspiration.

The REPTree model further distilled these insights by highlighting English proficiency as the dominant determinant in educational outcomes. It confirmed that even with variation in A/L stream and job preference, students' language and coding capabilities significantly shape institutional recommendations.

Across all models, English language proficiency and coding ability emerged as decisive factors, underscoring a systemic pattern: students with lower exposure to English and ICT education are more likely to be confined to limited academic and professional paths. These findings signal a pressing need for early intervention in language and digital education to ensure equity in higher education access.

Future research can build upon the findings of this study by incorporating additional dimensions and broader datasets to deepen the understanding of student decision-making in education. One promising direction is the use of longitudinal data to track students' actual academic and career outcomes over time, which would validate and potentially refine the predictive models developed here. Furthermore, integrating psychosocial variables such as socio-economic status, parental education, access to digital resources, and personal motivation could enhance the models' explanatory power. Geographic expansion of the dataset to include students from diverse districts or provinces would allow for regional comparisons and help uncover localized disparities in educational access. Intervention-based studies such as implementing targeted English language and coding training could also be explored to assess how these skills influence student aspirations and institutional preferences. Finally, experimenting with more sophisticated machine learning techniques, such as ensemble methods or deep learning models, may improve classification performance and reveal deeper, nonlinear patterns that simpler decision trees might miss.

REFERENCES

[1] S. Y. De Alwis and N. M. Wijesekara, "DETERMINANTS INFLUENCING MIGRATION DECISION AMONG SCHOOL LEAVERS IN SRI LANKA (WITH SPECIAL REFERENCE TO COLOMBO DISTRICT IN SRI LANKA)," *Sri Lanka J. Econ. Stat. Inf. Manag.*, vol. 3, no. 1, Nov. 2024, Accessed: Jul. 06, 2025. [Online]. Available: http://repo.lib.sab.ac.lk:8080/xmlui/handle/susl/4559

[2] S. A. Papert, *Mindstorms: Children, Computers, And Powerful Ideas: 9780465046744: Papert, Seymour A: Books*. Basic Books, 1993. Accessed: Jul. 06, 2025. [Online]. Available: https://www.amazon.com/Mindstorms-Children-Computers-Powerful-Ideas/dp/0465046746

[3] A. A. Ogegbo and A. Y. Aina, "Exploring young students' attitude towards coding and its relationship with STEM career interest," *Educ. Inf. Technol.*, vol. 29, no. 8, pp. 9041–9059, Jun. 2024, doi: 10.1007/S10639-023-12133-5/TABLES/4.

[4] A. Rameez, "English Language Proficiency and Employability of University Students: A Sociological Study of Undergraduates at the Faculty of Arts and Culture, South Eastern University of Sri Lanka (SEUSL)," *Int. J. English Linguist.*, vol. 9, no. 2, p. 11, 2019, doi: 10.5539/ijel.v9n2p199.

[5] L. Ranwala, S. Siriwardena, V. Kurukulaarachchi, and L. Edirisinghe, "Factors affecting on students' university choice in the tertiary education in Sri Lanka," *Int. J. Educ. Adm. Policy Stud.*, vol. 15, no. 2, pp. 97–109, 2023, doi: 10.5897/IJEAPS2023.0752.

[6] J. Du, X. Zhang, H. Zhang, and L. Chen, "Research and improvement of Apriori algorithm," *6th Int. Conf. Inf. Sci. Technol. ICIST 2016*, pp. 117–121, Jun. 2016, doi: 10.1109/ICIST.2016.7483396.

[7] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324/METRICS.

[8] K. Suneetha, K. Sutariya, and R. Garg, "Deployment and analysis of linear regression and REPtree for finding anomalies in network traffic," *Sustain. Digit. Transform. Era Driv. Innov. Growth*, pp. 183–190, Aug. 2024, doi: 10.1201/9781003534136-30.