Kathmandu University
Journal of Science, Engineering and Technology

Factors Influencing Nepali Students' Study Abroad Decisions: An Integration of Discrete Choice and Predictive Models

Mala Deep Upadhaya*[ab] and Barsha Thakuri[a]

[a]Faculty of Engineering, Environment and Computing, Coventry University, United Kingdom
[b]Sunway College, Kathmandu, Nepal, 0000-0002-6103-2669

Abstract

Global migration for study purposes has been a growing trend, with Nepali students aged 16 and above moving abroad for study. A study examining the socio-economic factors that motivate Nepali students (n=101) revealed that factors such as the target study country, the presence of family abroad, and part-time employment opportunities were significant. In contrast, sex, age, and educational level showed minimal impact. Lower living costs in environmentally diverse regions and university prestige were also prime factors. The research developed a study prediction abroad model using six machine algorithms and various sampling techniques, such as random undersampling, random oversampling, SMOTE, and ADASYN, to address class imbalances and prevent data leakage. The study found that logistic regression with RandomOverSample followed by SMOTE performed better than other methods, with the best mean accuracy of 0.83 and 0.82, respectively. The study also uses the ethical approach of the right to delete data promptly and a clean slate for sociodemographic analysis. However, the study acknowledges limitations, including a small sample size, possible influences from psychosocial factors, and the cross-sectional nature of the data. Future research with larger samples and longitudinal studies could offer additional information about the complex decision-making processes of students.

Keywords: discrete choice experiment with ranking, socio- demographic analysis, machine learning, ethical biases

1. Introduction

Moving from one's home country to a neighboring country or another developed country has always been a part of human civilization. The simple term for such movement is migration. It encompasses moving within a country, called internal migration, or across international borders, called external migration. Both internal and external have been occurring throughout human history; the modern patterns of large-scale, especially external, began to emerge in the late 19th to early 20th centuries [1]. The credit for this level of migration was noted due to the development of passport and visa systems during the same period. A study from [2] shows that before the 2020 pandemic, tens of millions of people crossed borders on a daily basis, which gave rise to the big number called three billion border crossings per year. The reason for migration is a tapestry of various factors and natures, ranging from internal, external, forced, voluntary, seasonal, and transnational. For easy categorization, all these types are placed under the "push" and "pull" sections. A push factor is the condition that drives people away from their home country, while "pull" means the attraction of the target (abroad) country. The push and pull factors are composed of various migration types. However, the research focuses more on the abroad study migration among the Nepali students of age above 16.

1.1. Background

In various periods of time, the local newspaper has been covering the number of cases in which Nepali citizens leave their home country for education or work reasons [3].The newspaper also highlighted that over 100,000 have left the country. Similarly, the data obtained from the Ministry of Education, Science, and Technology (MOEST) shows that from July 17 to January 14, 2023, 51585 people obtained No Objection Certificates (NOCs). The application for the NOCs shows that the student is interested in studying at a desired university in a foreign country. The letter confirms that students meet the necessary requirements and serves as governmental approval for studying abroad. With this, we can say that the increase in the number of NOC applications has a direct relation with the tendency of students to be ready to leave their home country. Furthermore, from July 12, 2022, to July 16, 2023, 104,155 NOCs were issued, which is an increase of 7,903 compared to July 16, 2021, to July 14, 2022. On average, about 500 people apply for NOCs online.

The paper [4] shows that most Nepali students do not plan to return after completing their education. The article's author shares his personal conversation findings that the Nepali from Australia and the United States of America (USA) do not want to return to Nepal, as they treat the student visa as a "one-way-ticket out of Nepal" [5]. This trend is evident in personal observations, where approximately 83% of a recent bachelor's cohort have relocated to countries including the USA, Australia, and Canada, with only a handful remaining in Nepal. The trend of Nepali students relocating abroad frequently feels like a single-journey departure from Nepal. This student migration brings significant challenges to our home country; some of the most prominent issues are as follows:

- Loss of Intellectual Capital: Migration of talented and potential young individuals leads to a drain of skilled manpower that could otherwise contribute to the country's economy [6].
- Demographic Imbalance: The departure leaves behind a population that includes either young children waiting to leave

*Corresponding author. Email: maladeep.upadhaya@gmail.com

once they are of age or elderly parents who are left to manage without the support of their children.
- Grassroots-Level Economic Impact: The absence of these young individuals can slow down local economic growth and innovation. Their skills and education are often not available to contribute to the local economy in their hometown [7].
- Social Challenges: Families left behind may struggle with increased responsibilities and emotional strain, particularly if they are caring for elderly parents or managing without the support of their children.
- Community Decline: The lack of returning youth can contribute to a decline in community vitality and social cohesion, as the community loses its younger generation, who might have contributed to cultural preservation.

In local news and interviews, I have noticed that even the student of primary grade is proudly saying that after class 12 they are more likely to go to a foreign country like Australia, the US, or Canada, as they say that *Nepal ma basey ra k cha ra* (what is there living in Nepal?). The prevalent attitude among adolescents regarding studying abroad raises concerns about the long-term development of the country. This project aims to identify the key factors influencing students aged 16 and above in their decision to pursue higher education abroad. Additionally, it seeks to provide valuable insights and guidance to help students select the destination country that best aligns with their academic and social goals. The ultimate objective is to support students in making informed and strategic decisions about their higher education opportunities.

1.2. Research Questions

This project is guided by primary research questions.

- RQ1: What socio-economic, academic, and personal factors significantly influence Nepali students aged 16 and above in their decision to pursue higher education abroad, and what is the relative importance of each factor as determined through Discrete Choice Experiments?
- RQ2: How effective are different machine learning algorithms in predicting students' likelihood of studying abroad based on socio-demographic factors?
- RQ3: What ethical considerations and bias mitigation strategies should be implemented when analyzing socio-demographic data for educational decision prediction using machine learning approaches?

1.3. Research Objectives

This study has three specific objectives that directly correspond to the research questions:

1. Objective 1 (addresses RQ1): To identify, quantify, and rank the socio-economic, academic, and personal factors that influence Nepali students' decisions to pursue higher education abroad using Discrete Choice Experiments within the Random Utility Theory with ranking methodology.

2. Objective 2 (addresses RQ2): To develop, implement, and evaluate predictive models using six machine learning algorithms with various sampling techniques (SMOTE, ADASYN, RandomOverSampler, and Random Undersampling) to assess students' likelihood of studying abroad and determine the most effective approach for this prediction task.

3. Objective 3 (addresses RQ3) aims to examine ethical considerations in the analysis of socio-demographic data, specifically focusing on strategies for mitigating bias, measures for protecting privacy (including the Right to Delete in Due Time and the Right to a Clean Slate), and the responsible implementation of predictive modeling in educational decision-making.

4. The objective is to identify and quantify the socio-economic, academic, and personal factors that influence Nepali students aged 16 and above in their decision to pursue higher education abroad. This goal will be achieved by utilizing a discrete choice experiment within the random utility framework to assess the relative importance of these factors.

5. Investigate the effectiveness of various predictive modeling techniques, including logistic regression, K-nearest neighbors, decision tree, random forest, naïve Bayes, and support vector machine, for assessing students' likelihood of studying abroad based on socio-economic factors.

6. Conduct a broad statistical analysis to identify and understand the key factors influencing Nepali students' decisions to study abroad. This includes using descriptive statistics, correlation analysis, and regression analysis.

7. Examine the ethical considerations in socio-economic data analysis, focusing on potential biases, privacy concerns, and the responsible use of machine learning methodologies for analyzing and predicting students' study-abroad decisions.

1.4. Significance/Rationale of the study

The research is highly significant in comprehending the determinants that motivate Nepali students aged 16 and above to seek higher education overseas. As global education patterns are changing, it is important to identify and study the socio-economic and personal elements that influence this decision. By analyzing socio-economic factors and employing advanced data science techniques, the study aims to provide valuable insights that benefit various stakeholders:

1. Students: Understand the specific drivers of study abroad decisions that can be utilized by the students to navigate their options more effectively and identify recommended courses tailored to individual profiles.

2. Educational Institutions: For overseas institutions, understanding the influencing factors allows for the development of targeted programs and strategies to attract and support international students. For domestic institutions, the research offers insights into creating initiatives that encourage students to pursue education locally.

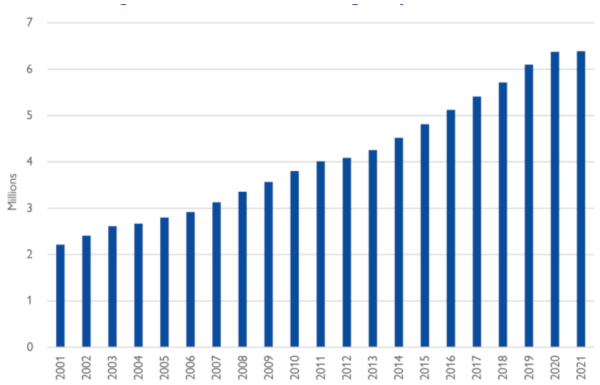
3. Policy Makers: The findings will help policymakers comprehend the changing factors behind students' migration, guiding them in formulating effective strategies to retain students and support domestic institutes to align new programs and offerings to student preferences.

4. Data Researcher: The findings and methodologies of this study will serve as a benchmark for similar research in other contexts or regions, thereby broadening the understanding of global education trends.

This section should articulate the significance of the proposed research, highlighting the potential beneficiaries and their potential benefits. The section should focus on justifying the topic of the proposed research.

2. Literature Review

The literature review will explore migration patterns and abroad study, socio-economic factors influencing higher education choices, decision-making theory, data science and machine learning in socio-demographic analysis, and ethical considerations in socio-economic data analysis.



Note: Exponential increase from 2 million to 6.4 million students worldwide, demonstrating the rising trend in educational migration that contextualizes Nepal's student outflow

Figure 1: Global International Student Mobility Growth (2001-2021). Source: UNESCO Institute for Statistics [15].

2.1. Migration Patterns and Abroad Study

Global migration patterns have evolved significantly. In migration we can see two distinct migration patterns termed as "International Migratory Highways," with asymmetric flows towards major immigration centers, and "Migratory Clusters," with symmetric flows within specific geographical areas [8]. The first pattern is explained as a large number of people moving in one direction (towards developed countries like the USA, Canada, Australia, or most of the Western European nations). Data from [9] shows that in 2022, 46.2 million immigrants are living in the USA, which is seen as the highest number in US history. This pattern is driven by a higher amount of job opportunity, higher wages, the presence of a prestigious university, and safety [10, 11]. The cluster seems to have more long-term migration. The second pattern resembles a cyclical movement between adjacent countries or specific regions. This migration is driven by regional or cultural ties. Movement between intra-European Union migration falls under this pattern [12]. Migration for the study purpose has risen exponentially. Students often seem to migrate for study purposes, but it is considered a broader strategy for future labor migration [13]. The US is regarded as the top destination for a pull country [14]. The data from [15] shares that from 2 million in 2000, it has reached 6.4 million in 2021. Among Nepali students, Canada, Australia, the USA, Japan, and the UK have been the top countries for study choices.

2.2. Socio-economic Factors Influencing Higher Education Choices

Race and socioeconomic status create disparities in the decision to study abroad. The study found that city-living students are more likely to access higher education abroad compared to their rural counterparts [16, 17]. In a global study, socio-economic factors like parental education levels and subject prestige [16], along with financial status, language of instruction, peer group, college environment, tuition fees, and even expected post-graduation salary [18, 19, 20, 21], also positively influence decisions to study abroad. In addition to this, students with international exposure or foreign-born parents are more influenced to study abroad [22]. Household income, abroad living expenses, political racism and discrimination, and cultural strictness decrease the likelihood of intent to study abroad [16, 17]. The study conducted in Nepal identifies broader career options, a desire for practical skills, expectations of higher earnings, and social interaction as the main pull factors for studying abroad [23, 24, 25]. Political instability, social injustice, insecurity, and uncertainty within Nepal are major push factors [25, 26].

2.3. Decision-making Theory

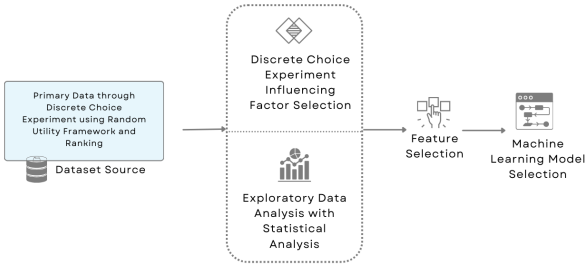
Decision-making is a cognitive process involving the evaluation and selection of a belief or concern over other possible alternatives [27]. Theories like the Rational Choice Model, Attribution Theory, Prospect Theory, Random Utility Theory (RUT), and Game Theory have been utilized for making decisions [27, 28, 29]. For addressing the student abroad study decision, push-pull theory and the rational choice model are used, but a recent study emphasizes the importance of individual motivation [30]. The same paper suggested that the decision to study abroad is a highly subjective and intrinsically driven behavior in which realizing one's self-worth or fulfilling one's purpose of life plays the most significant role. RUT with Discrete Choice experiments (DCE) is also widely used to analyze the preference and decision-making of individuals. The research [31, 32] utilized DCE and suggested a lot of potential in education research.

2.4. Data Science and Machine Learning in Socio-demographic Analysis

With the rise of computational power, machine learning (ML) has been highly utilized in the analysis of socio-demographic data [33]. Popular algorithms like K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), and Gaussian Naïve Bayes (GNB) were utilized for the prediction analysis [34]. For feature selection of the Socio-Demographic Characteristic (SDC), the Extra-Trees classifier and voting classifiers (ensemble-based algorithm) had outperformed logit and probit models [35]. ML has been accurately predicting COVID-19 infection based on socio-demographic and behavior factors [36] compared to traditional statistical methods. Socio-demographic data are often missing; thus, ML is utilized for handling data quality issues and bias. In the study, ML predicted employee salaries based on demographic information and work experiences, even in minimal demographic data [37,38]; however, it has shown biases related to race, age, gender, and socioeconomic status [39].

2.5. Ethical Consideration in Socio-economic Data Analysis in Data Science

Ethical considerations in data science, especially in regard to socio-economic data analysis, are composed of diverse parameters like privacy protection, responsible use, and comprehensive analysis of available data [40]. As the data science field is growing, feeding on a great number of data, the bibliometric analysis study from [41] shows that the ethical considerations are still underrepresented in scientific literature. The research from [42] focuses on the implementation of integrating algorithms with social awareness and suggests that if addressed properly, it directly affects fairness, transparency, and accountability in data-driven decisions. The study from [43] guided the way of analyzing the SDC dataset. On performing primary data collection, the proper consent should be obtained, and only after anonymizing or aggregating data should the analysis move forward. Regarding secondary SDC data analysis, the researcher must maintain a contextual understanding of the data to avoid misinterpretations. The study form [44] emphasizes that the SDC dataset researcher should give participants one of two options: Right to Delete in Due Time (RDDT) and Right to a Clean Slate (RCS). If RDDT is chosen, then participants can request the removal of their personal data from the analysis or when they withdraw consent. With RCS, the respondent can begin anew, free from the influence of previous datasets.



Note: Integration of Discrete Choice Experiments, Random Utility Theory, and Machine Learning for analyzing Nepali students' study abroad decisions (n=101)

Figure 2: Mixed-Methods Research Design Framework

3. Methodology

3.1. Research Design

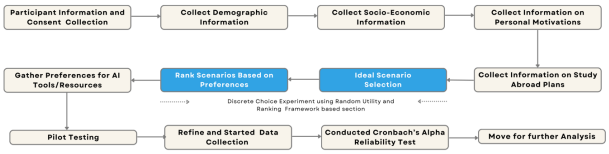
The research utilized the mixed-methods approach, integrating both quantitative and qualitative data collection and analysis to understand the factors influencing Nepali students' decisions to pursue higher education abroad. The research design is structured to capture a comprehensive view of socio-economic and personal elements while leveraging advanced data science techniques to develop a predictive model. The research workflow is shown in the figure below:

3.2. Survey Design and Data Collection

1. Survey Framework: The survey was designed using the Random Utility Theory (RUT) in combination with the Discrete Choice Experiment (DCE) methodology and the ranking method. This approach was chosen to model decision-making processes by capturing students' preferences across multiple attributes. The combination of RUT and DCE offers a robust framework for preference elicitation and economic valuation across various fields [47]

2. Sample Population: The survey targeted Nepali students aged 16 and above who are considering or have considered studying abroad. Participants were recruited through online platforms, educational institutions, and social media channels, ensuring a diverse and representative sample.

3. Sample Instrument: The survey included a range of questions covering demographic information (e.g., gender, age, education level), socioeconomic factors (e.g., family income, parents' education), and personal motivations (e.g., study subject, target country). The survey was developed using a survey tool called fillout.com. Only after the consent was recorded were the participants allowed to fill out the survey. The questionnaire was divided into five parts: demographics, planning to go abroad, ideal scenario, tool preferences, and scenario ranking. The instrument was piloted with a small group to refine questions and ensure clarity. Thereafter, over the period of a month, data was collected. Before proceeding with the analysis, to assess the reliability of the survey instrument, Cronbach's alpha reliability test was employed, as this test provided an estimate of the internal consistency of the survey items with 1.0 (perfect reliability). Much research suggests that values between 0.70 and 0.90 generally indicate good reliability, and the above is due to redundant [48, 49] items that align with this research, as the research consists of a ranking method. You can find the questionnaires and live link to the survey in the appendix.



Note: Random Utility Framework with ranking methodology for capturing student preferences across environmental, economic, social, and academic factors influencing study abroad decisions

Figure 3: Discrete Choice Experiment Survey Design

Table 1: Gender demographic information from the survey data.

Gender	Population
Female	42.6%
Male	53.5%
Non-Binary	4.0%

3.3. Data Collection

The initial dataset consisted of n = 157 responses. After cleaning, the dataset was refined to n = 101, with 70 features. The target variable was taken as plans to study abroad, and the rest were considered as independent variables for the predictive model.

3.4. Data Analysis Workflow

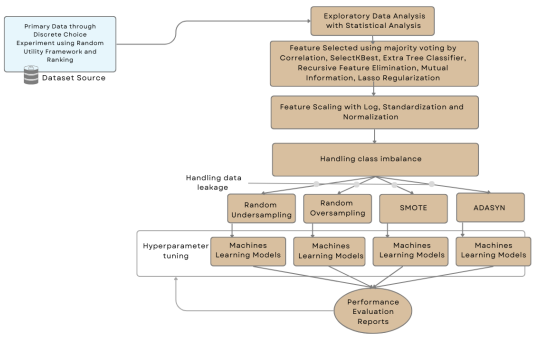
Once the dataset was obtained and exploratory data analysis (EDA) was performed on the survey as well as other datasets, feature selection was done with majority voting by Pearson correlation matrix, SelectKBest with chi-squared (Chi2) and mutual information as scoring functions, extremely randomized tree, Recursive Feature Elimination (RFE), and Lasso regularization with regard to DCEs. To ensure that features with lesser significance do not dominate the objective function, we followed feature scaling techniques with log, standardization, and normalization. Based on EDA, we used undersampling, random sampling, Synthetic Minority Oversampling Technique (SMOTE), and Adaptive Synthetic (ADASYN) for handling class imbalance and target, as well as target leakage, and preprocessing data leakage was tackled. Thereafter, we applied six different classification algorithms and generated the classification performance report for the predictive model. Additionally, the research utilized GridSearchCV for hyperparameter tuning to enhance the model's performance. Once the final evaluation was done, a predictive model was created. ML: The workflow is shown in the figure below.

3.5. Experimental Setup

To process data and train models, an Apple M1 chip with an 8-core CPU and an 8-core GPU was used with 8 GB of unified memory with 256 SSDs in macOS Ventura version 13.2.1. Python 3.10.10 was used as the programming language, with Jupyter Notebook version 6.4.8 as the development environment.

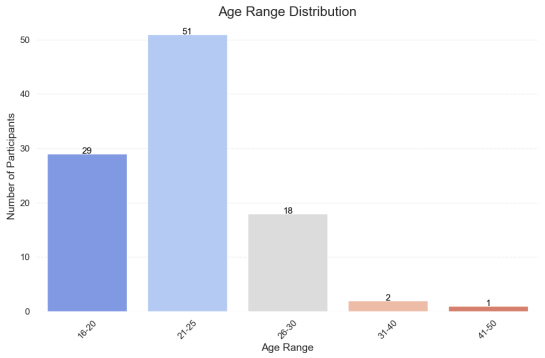
4. Results and Discussions

The results presented in this section are structured to address each research question and specific objective of the study. The findings are organized by the analysis methods used, beginning



Note: Six-algorithm approach with feature selection, class balancing (SMOTE, ADASYN, RandomOverSampler), and hyperparameter optimization for predicting study abroad likelihood

Figure 4: Machine Learning Pipeline Workflow



Note: Predominant representation of 21-25 age group (45.5%), followed by 16-20 (28.7%) and 26-30 (15.8%), reflecting the prime decision-making years for higher education choices

Figure 5: Age Distribution of Survey Respondents (n=101)

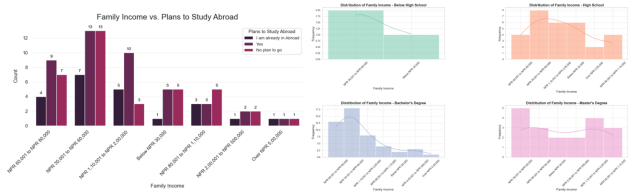
with exploratory data analysis, followed by feature selection for predictive modeling, feature selection for discrete choice experiments, model creation for likelihood of going abroad (predictive model), and ethical consideration in socio-economic data analysis. Each section will include a discussion of the results, a comparison with similar studies, and an exploration of causal factors.

4.1. Exploratory Data Analysis

As the sample consists of 101 respondents, the imbalance in the dataset was seen. The majority of the survey were in the age group 21-25, followed by 16-20, and so on.

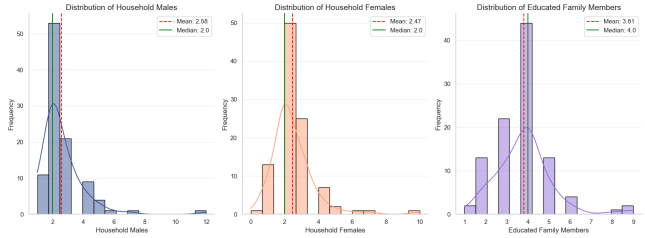
Similarly, the majority (n=23) has shown the US as the preferred country abroad, followed by Australia (10) and the UK and New Zealand in the same ratio, and so on. Looking at the family income vs. plan to study abroad, it was seen that the monthly income range of NPR 30000-60000 was more ready to plan for a study abroad. However, in this range, there was the same count of people who do not wish to go abroad too. Interestingly, the majority of people who are already abroad represent this income group. The monthly income range of 110,001 to 200,000 was more inclined to go abroad than any other income group, as shown in the figure below.

Looking at the household number distribution, it was seen that household males have 3.26 (highly skewed), which suggests that the majority of households have a low number of males, but a few households have a significantly higher number. When examining the household females, the moderately skewed value was 2.54, indicating that, similar to males, most households have a smaller number of females, while some have considerably more. On the contrary, the educated family member has a 1.15 slightly skewed distribution, being closer to normal, but there is still a mild tendency towards higher values. This conclusion suggests that while the majority of households have a moderate number of educated family members, there are some with a higher-than-average count.



Note: Students from NPR 30,000–60,000 monthly income families show highest migration propensity, with income groups NPR 110,001–200,000 demonstrating strongest abroad study commitment

Figure 6: Family Income and Study Abroad Intentions



Note: Male-dominated households (mean=3.26, highly skewed) and moderate female representation (mean=2.54), with educated family members averaging 1.15 per household

Figure 7: Gender Distribution in Households

On looking at the abroad study plan among the oldest children of the family, it was found that not all of the oldest wanted to go. Though the number is less compared to the non-oldest child of the family.

Using this data, we performed a statistical analysis, which is presented in the table:

Education level emerged as a significant factor influencing the decision to study abroad, suggesting that students with higher education levels are more likely to pursue studies overseas. This finding aligns with literature [20, 21] indicating that educational attainment influences international mobility decisions. In addition to this, family abroad is a critical factor strongly influencing students' decisions to study abroad, aligning with theories that family connections facilitate international education opportunities. Part-time work also showed significance, possibly indicating that students engaged in it are more inclined to study abroad due to increased financial independence or job-related opportunities. Environmental factors were significant, reflecting the importance of the study environment in influencing the decision, which is consistent with the literature on environmental [16, 17]and contextual factors affecting educational choices. The feature aligns with the finding of the study [18, 19]. Other demographic factors such as gender, age, and marital status did not show significant impact, which is consistent with some studies [18].

4.2. Feature Selection for Predictive Modeling

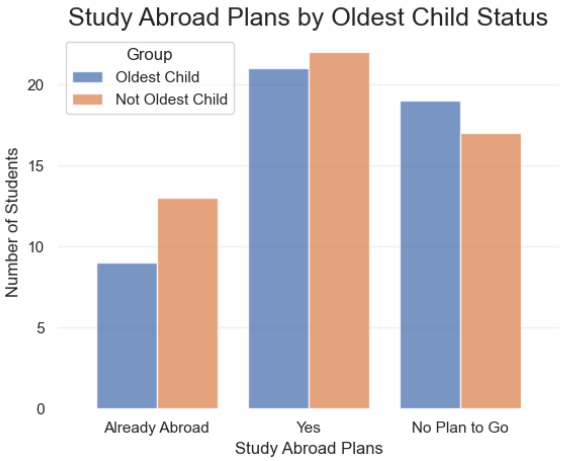
With the goal of improving the model's performance and interpretability by reducing the number of not-important attributes used in the model, this subsection is termed "feature selection." The study incorporated

1. Correlation by Pearson correlation for most correlated values among the dataset

Table 2: Chi-Square Test Results Across Features

Feature	χ^2	p-value	Result
Gender	7.7191	0.1024	NS
Age	8.5705	0.3798	NS
Education Level	14.6450	0.0232	S
Marital Status	3.9491	0.4129	NS
Household Males	9.8342	0.7742	NS
Household Females	13.5824	0.6298	NS
Educated Family Members	13.2746	0.5050	NS
Oldest Child?	0.7732	0.6794	NS
Younger Siblings	2.0137	0.9184	NS
Older Siblings	8.4901	0.3871	NS
Mother's Education	13.4125	0.6424	NS
Father's Education	7.9812	0.7866	NS
Parents' Jobs	9.1606	0.6892	NS
Family Income	5.5771	0.9359	NS
Part-Time Work?	9.5178	0.0086	S
District	92.3278	0.3009	NS
Family Abroad	15.6211	0.0004	S
Target Country	0.0000	1.0000	NS
Planned Year Abroad	0.0000	1.0000	NS
Study Subject Abroad	0.0000	1.0000	NS
Study Subject	0.0000	1.0000	NS
Education Quality	2.2363	0.3269	NS
Career Prospects	0.1387	0.9330	NS
Political Factors	1.5703	0.4561	NS
Social Factors	2.1298	0.3448	NS
Economic Factors	1.6106	0.4470	NS
Safety/Security	0.5457	0.7612	NS
University Preference	1.7516	0.4165	NS
Language Access	2.9357	0.2304	NS
Cultural Exposure	1.8158	0.4034	NS
Research Opportunities	1.7396	0.4190	NS
Environmental Factors	6.8171	0.0331	S
Quality Edu Rank A	13.9325	0.4548	NS
Quality Edu Rank B	6.7340	0.7503	NS
Career Rank A	18.9504	0.0897	NS
Career Rank B	6.8811	0.5495	NS
Political Rank A	3.2193	0.9198	NS
Political Rank B	17.5305	0.1307	NS
Social Rank A	9.8655	0.4524	NS
Social Rank B	8.8888	0.7124	NS
Economic Rank A	4.7039	0.7887	NS
Economic Rank B	16.9576	0.3883	NS
Safety Rank A	8.1769	0.4164	NS
Safety Rank B	7.5887	0.6689	NS
University Rank A	12.0551	0.4413	NS
University Rank B	13.2964	0.3479	NS
Language Rank A	7.0702	0.7188	NS
Language Rank B	9.4653	0.6628	NS
Cultural Rank A	5.1746	0.7388	NS
Cultural Rank B	13.7532	0.1846	NS
Research Rank A	10.6906	0.3821	NS
Research Rank B	15.5867	0.1121	NS
Environment Rank A	9.5458	0.4812	NS
Environment Rank B	9.1375	0.6911	NS

Note: S = Significant, NS = Not Significant



Note: Comparison of study abroad preference between eldest and non-eldest children, revealing that non-eldest children show slightly higher propensity for international education

Figure 8: Birth Order Analysis of Study Abroad Preferences

2. SelectKBest identifies features most correlated with the target.
3. RFE (Recursive Feature Elimination) selects features based on their importance within a model.
4. Extreme Tree Classifier provides insights into the features that have the greatest predictive power.
5. Mutual Information captures non-linear relationships between features and the target.
6. Lasso Regularization prioritizes features while enforcing simplicity by shrinking less important coefficients to zero.

Features were selected in two ways:

1. DCEs along with ranking
2. DCEs without ranking

For the predictive model, the top 10 features—'Target Study Abroad Country,' 'Family Abroad,' 'Study Subject Abroad,' 'Part-Time Work?', and 'Environmental Factors'—were chosen for the prediction of the likelihood of going abroad. These features correspond to various SECs mentioned in the literature above.

Since we have employed RUT with DCE and ranking, we need proper modeling of the data. On doing analysis, it was found that the main factor for choosing abroad was environmental factors, with the average rank of 1.99 (lower value is better). This finding is aligned with the study [18, 19]. Research opportunities, career prospects, and cultural exposure followed closely behind.

As the DCEs have a ranking method introduced, we revered to see (the higher the better) and found that under environment, most preferred to see a lower cost of living in environmentally diverse regions. From each main DCES, the sub-ranking option shows that the prestige and reputation of the university, abundant availability of part-time job opportunities during studies, rich cultural exposure and diversity on campus, access to diverse job opportunities with competitive salaries, high-quality education, different primary language spoken requiring language adaptation, political stability and security in the home country, inadequate emergency services, and abundant opportunities for social interaction were ranked higher. Surprisingly, people tend to view inadequate emergency services negatively, yet the rankings below reflect this.

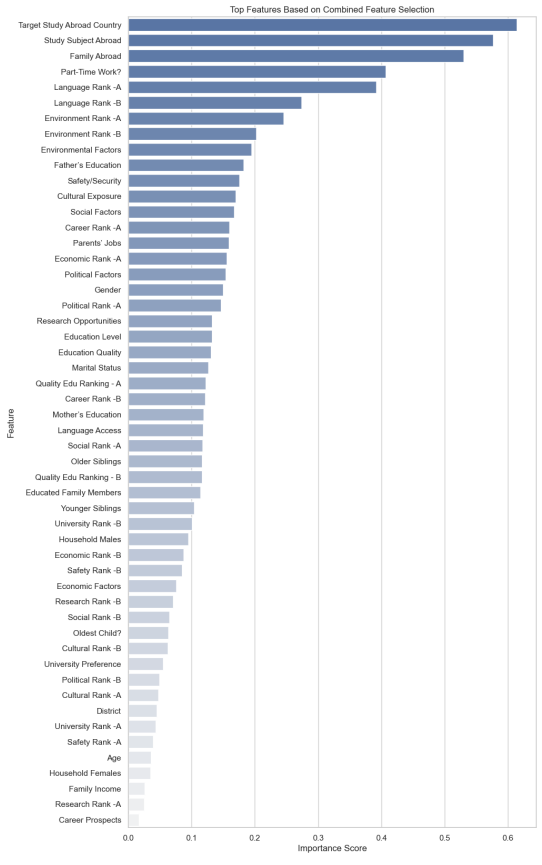


Figure 9: Top 10 Feature Importance Rankings (DCE Method): Target Study Country, Family Abroad, and Environmental Factors emerge as the most influential determinants of study abroad decisions.

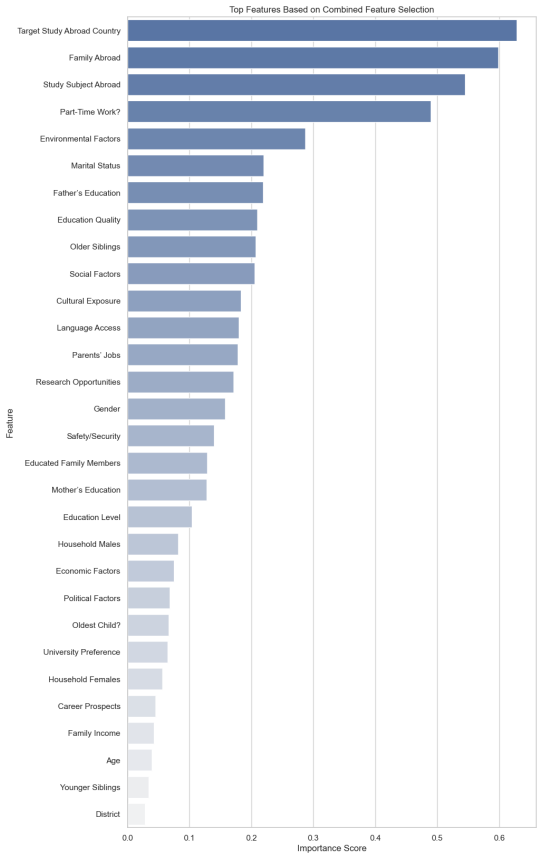


Figure 10: Top 10 Feature Importance Rankings (DCE Method): Reinforcing the previous analysis, Target Study Country, Family Abroad, and Environmental Factors remain primary predictors of study abroad intentions.

4.3. Feature Selection in Discrete Choice Experiment

It was found that logistic regression performed the best overall, especially when combined with RandomOverSampler. However, the research took SMOTE for the model, as it prevents overfitting because of its nature of adding variability, while RandomOverSampler does not, and SMOTE has been found to outperform other oversampling methods in similar research [50, 51]. The optimal hyperparameters for logistic regression with SMOTE were C = 1 and solver = 'newton-cg'. This combination achieved the best score (hyperparameter tuning): 0.8889, with an F1 score of 0.8148. However, for cross-validation, it was logistic regression with SMOTE, with a mean accuracy of 0.83. SMOTE proved to be an effective resampling technique, balancing the dataset while maintaining performance across different models. The top-performing models based on overall accuracy, stability, and interpretability were logistic regression with SMOTE, K-Nearest Neighbors with SMOTE, Random Forest with SMOTE, and Decision Tree with SMOTE. The finding chimps with the study [50, 51, 52] that suggested logistic regression with SMOTE and random forest with SMOTE consistently outperforms other methods, achieving high accuracy, sensitivity, and specificity in classification tasks. However, this finding contradicts the study [54], which found that logistic regression with SMOTE tends to underperform compared to other algorithms.

4.4. Ethical Considerations in Socio-economic Data Analysis

Since the start of the research, survey and data collection processes ensured that personal and socio-economic information was anonymized to protect participants' identities. This aligns with ethical standards for safeguarding personal data [40]. Informed consent was given on the first page of the survey that made the participant the direct control of the Right to Delete in Due Time (RDDT)

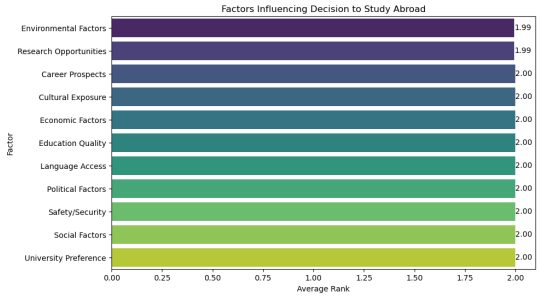


Figure 11: Decision Factor Priority Ranking: Environmental Factors are the top priority (average rank = 1.99), followed by Research Opportunities, Career Prospects, and Cultural Exposure as key motivators for study abroad.

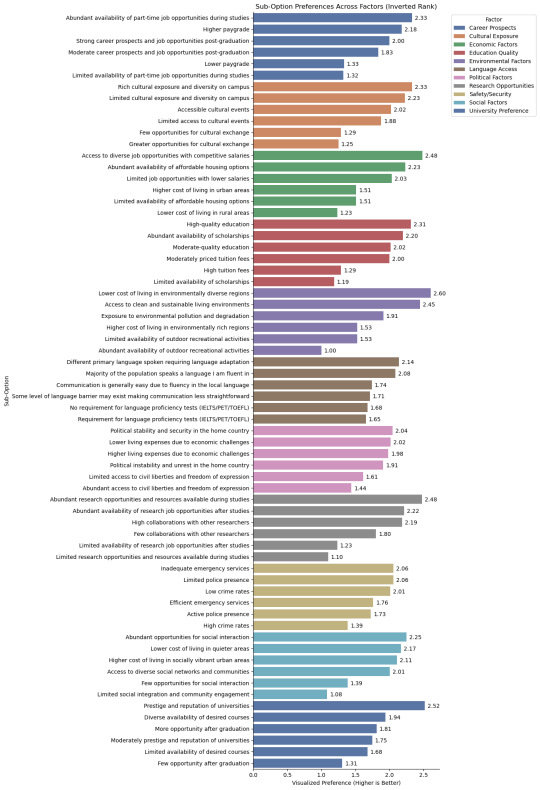


Figure 12: Discrete Choice Experiment Sub-factor Preferences: Highlights top considerations including University Prestige, Part-Time Job Availability, and Cultural Diversity across main categories.

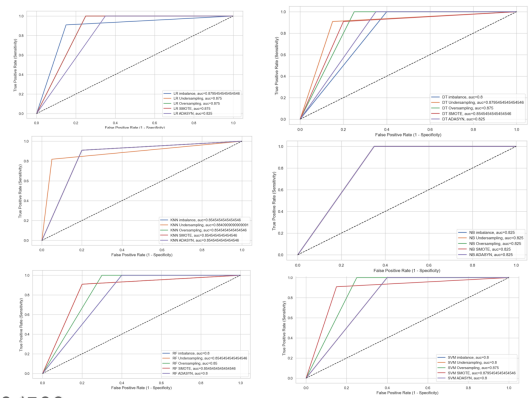


Figure 13: ROC Curve Analysis Across Models and Sampling Techniques: Logistic Regression with RandomOverSampler achieves the highest AUC, followed by SMOTE-enhanced models across six algorithms.

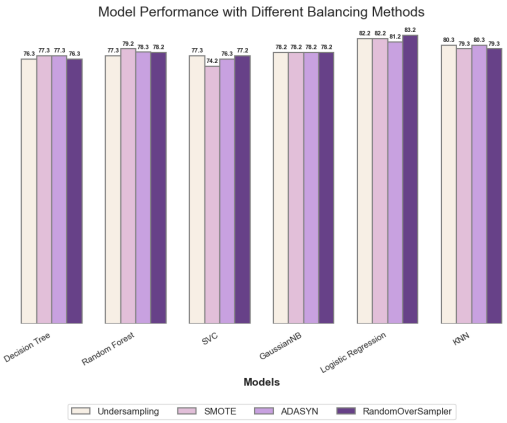


Figure 14: Sampling Technique Performance Evaluation: Cross-algorithm comparison shows SMOTE and RandomOverSampler outperform ADASYN and Random Undersampling for addressing class imbalance in study abroad predictions.

and the Right to a Clean Slate (RCS) [44, 43]. Use of a technique like SMOTE for balancing the dataset addresses the potential biases of underrepresenting the particular socio-demographic group as suggested in [55]. Furthermore, the questionnaire included a diverse range of socio-economic factors and demographic groups, ensuring that the analysis does not disproportionately favor or disadvantage any specific group. For addressing transparency and accountability, the entire research methodology was well-shared with the participants who requested it through the survey form.

5. CONCLUSION AND RECOMMENDATIONS

By examining socio-economic and personal determinants, the research sheds light on key drivers and barriers in the decision-making process. Utilizing the approach of a discrete choice experiment with a random utility framework and ranking the findings highlights the significant role of certain factors, such as the target country, family abroad, and part-time work, in shaping students' decisions, while other variables, including gender, age, and education level, showed no significant impact. The use of advanced data science techniques, including the combined feature selection of Correlation by Pearson, SelectKBest, Recursive Feature Elimination, Extreme Tree Classifier, Mutual Information, and Lasso Regularization, made the feature selection robust for the model. The high accuracy of the logistic regression model with SMOTE demonstrates the effectiveness of these techniques in understanding and predicting students' choices and removing the biases while analyzing socio-demographic data and creating data science products. The study's recommendations are as follows:

1. For Student: When evaluating potential study destinations, students should consider factors such as family abroad and part-time work opportunities. Furthermore, look out for the environment of the destination country and the prestige of the university, and check for abundant availability of part-time job opportunities during studies, rich cultural exposure and diversity on campus, access to diverse job opportunities with competitive salaries, high-quality education, and different primary languages spoken that require language adaptation.

2. For Education Institutes: For international institutes, it is suggested to create a culturally friendly and diverse school environment and have the prestige of the university shared as the first factor among the targeted students. Additionally, it's crucial to offer employment opportunities to students

during their studies. Domestic institutes should develop career counseling guidance and collaborate with global institutes to prevent students from leaving with a brain dump approach, thereby guiding them toward informed career and study choices as global citizens.

3. For Policymaker: The research strongly recommends formulating a policy that fosters partnerships with international education or the government, thereby creating opportunities that are exclusively available within their home country. Additionally, the research suggests considering factors such as the number of family members before granting NOCs to ensure that the departure of the young does not negatively impact the lives of parents and dependent children.

4. For Researchers: When analyzing socio-demographic data, it is crucial to inform participants about the Right to Delete in Due Time (RDDT) and the Right to a Clean Slate and to anonymize the data prior to analysis. Also highly recommended.

However, we must interpret and acquire these results and recommendations with caution, keeping in mind a number of limitations. The study was based on a survey dataset with 101 responses, which may not fully represent the diverse population of Nepali students considering studying abroad. A larger sample size could provide more robust and generalizable results. The study did not take into account other factors such as psychosocial pressure and peer pressure, which could have influenced the outcome. The data collected is cross-sectional, and having a longitudinal study might supply more profound insight into the underlying factor. Since the decision-making process is highly complex, the logistic regression with SMOTE may not fully capture all aspects of the students' decision-making.

References

[1] Wilson Center. Migration, forced displacement, and human development (2024). URL https://www.wilsoncenter.org/article/migration-forced-displacement-and-human-development, accessed: Aug. 25, 2024.

[2] Recchi E, Deutschmann E & Vespe M, Estimating transnational human mobility on a global scale, *SSRN Journal*. https://doi.org/10.2139/ssrn.3384000.

[3] OnlineKhabar. Over 100,000 left nepal to study abroad in fy 2023/24 (2024). URL https://english.onlinekhabar.com/nepal-student-study-abroad.html, accessed: Aug. 25, 2024.

[4] Silwal A. Assessment of brain drain and its impact on the sending economy: A case study of nepal (2024). URL https://www.semanticscholar.org/paper/Assessment-of-Brain-Drain-and-its-Impact-on-the-A-Silwal/43c2a8abfb637b4098152ce03c4b63f689c49f43, accessed: Aug. 25, 2024.

[5] Dhital M. Reversing nepal's brain drain (2024). URL https://asianews.network/reversing-nepals-brain-drain/, accessed: Aug. 25, 2024.

[6] Transnational work migration of nepali youths: The changing phenomena and the context of achieving education, *Journal of Education and Research*. URL https://nepjol.info/index.php/JER/article/view/30464, accessed: Aug. 25, 2024.

[7] Remittance income in nepal: Need for economic development, *Journal of Nepalese Business Studies*. URL https://www.nepjol.info/index.php/JNBS/article/view/49, accessed: Aug. 25, 2024.

[8] Akbari H, Exploratory social-spatial network analysis of global migration structure, *Social Networks*, 64 (2021) 181–193. https://doi.org/10.1016/j.socnet.2020.09.007.

[9] Camarota S A & Zeigler K. Just-released data: Foreign-born population above 46 million in july 2022 (2022). URL https://cis.org/Camarota/JustReleased-Data-ForeignBorn-Population-Above-46-million-July-2022, accessed: Aug. 25, 2024.

[10] The form and evolution of international migration networks, 1990–2015 (2015). URL https://www.researchgate.net/publication/349415549_The_form_and_evolution_of_international_migration_networks_1990-2015, accessed: Aug. 25, 2024.

[11] Europe U N W. Migration to the eu: facts, not perceptions (2024). URL https://unric.org/en/migration-to-the-eu-facts-not-perceptions/, accessed: Aug. 25, 2024.

[12] European Commission. Statistics on migration to europe (2023). URL https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/promoting-our-european-way-life/statistics-migration-europe_en#migration-to-and-from-the-eu, accessed: Aug. 25, 2024.

[13] Li M, Migrating to learn and learning to migrate: International student migrants, *International Journal of Population Geography*. URL https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-1220(199603)2:1%3C51::AID-IJPG17%3E3.0.CO;2-B.

[14] Mehra A & Kaur T, Flight for greener pastures: International migration of indian students, *International Journal of Economic Policy in Emerging Economies*, 15. https://doi.org/10.1504/ijepee.2022.121342.

[15] International students (2024). URL https://www.migrationdataportal.org/themes/international-students, migration Data Portal, Accessed: Aug. 25, 2024.

[16] Ekinci C E, Bazı sosyoekonomik etmenlerin türkiye'de yükseköğretime katılım Üzerindeki etkileri, *Eğitim ve Bilim*, 36(160). URL https://egitimvebilim.ted.org.tr/index.php/EB/article/view/792.

[17] Xiong L, Nyland C, Fisher B S & Smyrnios K X, International students' fear of crime: an australian case study, *Australian and New Zealand Journal of Criminology*. URL https://journals.sagepub.com/doi/10.1177/0004865815608676, accessed: Aug. 25, 2024.

[18] Creating a psychological paradigm shift in students' choice for tertiary education in sri lanka: the influence of socioeconomic factors, *International Journal of Educational Administration and Policy Studies*. URL https://academicjournals.org/journal/IJEAPS/article-abstract/C155A5D68584, accessed: Aug. 25, 2024.

[19] Attendance at higher-cost colleges: Ascribed, socioeconomic, and academic influences on student enrollment patterns, *Research in Higher Education*. URL https://www.sciencedirect.com/science/article/abs/pii/0272775788900726?via%3Dihub, accessed: Aug. 25, 2024.

[20] Choosing the future: Economic preferences for higher education using discrete choice experiment method, *Research in Higher Education*. https://doi.org/10.1007/s11162-019-09572-w. URL https://link.springer.com/article/10.1007/s11162-019-09572-w.

[21] Gong X & Huybers T, Chinese students and higher education destinations: Findings from a choice experiment, *Australian Journal of Education*. URL https://journals.sagepub.com/doi/10.1177/0004944115584482, accessed: Aug. 25, 2024.

[22] Simon J R & Ainsworth J W. Race and socioeconomic status differences in study abroad participation (2000). URL https://scholarworks.gsu.edu/sociology_facpub/1/.

[23] Unveiling motivational factors driving nepali students to pursue higher education abroad, *OCEM Journal of Management, Technology & Social Sciences*. URL https://www.nepjol.info/index.php/ocemjmtss/article/view/62221.

[24] Tamang M K & Shrestha M, Let me fly abroad: Student migrations in the context of nepal, *Research in Educational Policy and Management*, 3(1). https://doi.org/10.46303/repam.2021.1.

[25] International educational consultancies and students migration from nepal, *Journal of Population and Development*. URL https://www.nepjol.info/index.php/jpd/article/view/64238.

[26] Upadhyay-Dhungel K & Lamichhane S, Cost and financing higher education by nepalese student in australia, student loans and role of bank in nepal, *Banking Journal*, 1(1). https://doi.org/10.3126/bj.v1i1.5143.

[27] Edwards W, The theory of decision making, *Psychological Bulletin*, 51(4) (1954) 380–417. https://doi.org/10.1037/h0053870.

[28] Oliveira. A discussion of rational and psychological decision-making theories and models: The search for a cultural-ethical decision-making model (2024). URL https://www.semanticscholar.org/paper/A-Discussion-of-Rational-and-Psychological-Theories-Oliveira/2e22f06755ed5642edf740e7f2c1fad8ad9abbbd, accessed: Aug. 25, 2024.

[29] Thurston W E, Decision-making theory and the evaluator, *Canadian Journal of Program Evaluation*, 5(2) (1990) 29–45. https://doi.org/10.3138/cjpe.5.003.

[30] Yue Y & Lu J, International students' motivation to study abroad: An empirical study based on expectancy-value theory and self-determination theory, *Frontiers in Psychology*, 13. https://doi.org/10.3389/fpsyg.2022.841122.

[31] Kennelly B, Flannery D, Considine J, Doherty E & Hynes S, Modelling the preferences of students for alternative assignment designs using the discrete choice experiment methodology, *Practical Assessment, Research, and Evaluation*, 19(1). https://doi.org/10.7275/y9r2-nc06.

[32] Design of discrete choice experiments (2024). URL https://oxfordre.com/economics/display/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-91, accessed: Aug. 25, 2024.

[33] Ensari T, Ensari B & Dağtekin M, Violence detection with machine learning: A sociodemographic approach, *EJOSAT*, (44). https://doi.org/10.31590/ejosat.1225896.

[34] Islam T, Meade N, Carson R T, Louviere J J & Wang J, The usefulness of socio-demographic variables in predicting purchase decisions: Evidence from machine learning procedures, *Journal of Business Research*, 151 (2022) 324–338. https://doi.org/10.1016/j.jbusres.2022.07.004.

[35] Dash P B, Naik B, Nayak J & Vimal S, Socio-economic factor analysis for sustainable and smart precision agriculture: An ensemble learning approach, *Computer Communications*, 182 (2022) 72–87. https://doi.org/10.1016/j.comcom.2021.11.002.

[36] Machine learning-based screening solution for covid-19 cases investigation: Socio-demographic and behavioral factors analysis and covid-19 detection (2023). URL https://link.springer.com/article/10.1007/s44230-023-00049-9, accessed: Aug. 25, 2024.

[37] Satpute B S, Yadav R & Yadav P K. Machine learning approach for prediction of employee salary using demographic information with experience. In: *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)* (2023). https://doi.org/10.1109/GCAT59970.2023.10353537.

[38] Luo W et al., Is demography destiny? application of machine learning techniques to accurately predict population health outcomes from a minimal demographic dataset, *PLOS ONE*, 10(5) (2015) e0125602. https://doi.org/10.1371/journal.pone.0125602.

[39] Franklin G et al., The sociodemographic biases in machine learning algorithms: A biomedical informatics perspective, *Life*, 14(6). https://doi.org/10.3390/life14060652.

[40] Scime A & Murray G R. Social science data analysis: The ethical imperative. In: *IGI Global Handbook*. IGI Global (2014). URL https://www.igi-global.com/gateway/chapter/76260.

[41] Kuc-Czarnecka M & Olczyk M, How ethics combine with big a bibliometric analysis, *Humanities and Social Sciences Communications*, 7 (2020) 1–9. https://doi.org/10.1057/s41599-020-00638-0.

[42] Kearns M & Roth A. *The Ethical Algorithm*. Oxford University Press (2019). URL https://global.oup.com/academic/product/the-ethical-algorithm-9780190948207, accessed: Aug. 25, 2024.

[43] The ethics of secondary data analysis: Learning from the experience of sharing qualitative data from young people and their families in an international study of childhood poverty (2024). URL https://eprints.ncrm.ac.uk/id/eprint/3301/, accessed: Aug. 25, 2024.

[44] Koops B J, Forgetting footprints, shunning shadows: A critical analysis of the 'right to be forgotten' in big data practice, *SSRN Electronic Journal*. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1986719.

[45] Cascetta E. Random utility theory. In: *Springer Optimization and Its Applications* (2009), pp. 89–167. https://doi.org/10.1007/978-0-387-75857-2_3.

[46] Wuttaphan N, Human capital theory: The theory of human resource development, implications, and future, *Life Sciences and Environment Journal*, 18(2) (2017) 240–253. URL https://ph01.tci-thaijo.org/index.php/psru/article/view/76477.

[47] Robinson A, Spencer A & Moffatt P, A framework for estimating health state utility values within a discrete choice experiment: Modeling risky choices, *Medical Decision Making*, 35(3) (2015) 341–350. https://doi.org/10.1177/0272989X14554715.

[48] Tavakol M & Dennick R, Making sense of cronbach's alpha, *International Journal of Medical Education*, 2 (2011) 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd.

[49] Kalkbrenner M T, Choosing between cronbach's coefficient alpha, mcdonald's coefficient omega, and coefficient h: Confidence intervals and the advantages and drawbacks of interpretive guidelines, *Measurement and Evaluation in Counseling and Development*. URL https://www.tandfonline.com/doi/abs/10.1080/07481756.2023.2283637, accessed: Aug. 27, 2024.

[50] Ghorbani R & Ghousi R, Comparing different resampling methods in predicting students' performance using machine learning techniques, *IEEE Access*, 8 (2020) 67899–67911. https://doi.org/10.1109/ACCESS.2020.2986809.

[51] Penerapan synthetic minority oversampling technique (smote) terhadap data tidak seimbang pada pembuatan model komposisi jamu, *Semantic Scholar*. URL https://www.semanticscholar.org/paper/Penerapan-Synthetic-Minority-Oversampling-Technique-Barro-Sulvianti/0c193ba1485333a6d6c53b49e585d443aac70247, accessed: Aug. 28, 2024.

[52] Comparison of smote random forest and smote k-nearest neighbors classification analysis on imbalanced data, *ResearchGate*. URL https://www.researchgate.net/publication/369828798_COMPARISON_OF_SMOTE_RANDOM_FOREST_AND_SMOTE_K-NEAREST_NEIGHBORS_CLASSIFICATION_ANALYSIS_ON_IMBALANCED_DATA, accessed: Aug. 25, 2024.

[53] Stroke prediction based on random forest with smote, *ResearchGate*. URL https://www.researchgate.net/publication/374048551_Stroke_Prediction_Based_on_Random_Forest_with_SMOTE, accessed: Aug. 25, 2024.

[54] Rawat S S & Mishra A K. The best ml classifier(s): An empirical study on the learning of imbalanced and resampled credit card data. In: *2023 Second International Conference on Informatics (ICI)* (2023), pp. 1–6. https://doi.org/10.1109/ICI60088.2023.10421691.

[55] Bias in machine learning software: why? how? what to do? In: *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (2021). URL https://dl.acm.org/doi/10.1145/3468264.3468537, accessed: Aug. 25, 2024.