

Final Research Approval

Research Title: Predicting CSE Undergraduate Students' Careers Using Various Machine Learning And Ensemble Learning Algorithms in the Context of Bangladesh.

Student Information:

| | |
|---------|--------------------|
| Name | Israt Jahan Reshma |
| Id No. | 18CSE241 |
| Session | 2018-19 |

Supervisor Information:

| | |
|-------------|---------------------|
| Name | Dr. Saleh Ahmed |
| Designation | Associate Professor |
| E-mail | sumon.edu@gmail.com |

ABSTRACT/SUMMARY:

Background: Career selection is a critical decision for undergraduate students, particularly in the field of Computer Science and Engineering (CSE), where numerous career paths, such as software development, data analysis, cybersecurity, and government jobs, are available. However, many students face challenges in identifying a suitable career due to a lack of proper guidance and insight into their skills and preferences.

Methods: This study utilized data collected from 468 CSE undergraduate students in Bangladesh through a structured questionnaire. The dataset includes 24 features such as CGPA, soft and academic skills, research abilities, publication records, project experience, and preferred career options. Ten machine learning algorithms were applied to analyze the data, with models evaluated on accuracy, precision, recall, F1 scores (micro and macro), hamming loss, and execution time.

Results: Among the models, the XGBoost Classifier (XGBC) is expected to perform the best, with the highest accuracy (61.7%), precision (0.85), recall (0.73), and F1 Macro (0.69). The study anticipates identifying software development as the most preferred career path, with over 53.5% of students opting for it as their primary choice.

Implications: This research provides a framework for career counseling tailored to the strengths, skills, and interests of CSE students. By offering personalized career guidance, it enhances students' employability and aligns their career choices with their goals. Additionally, the study contributes to developing data-driven career counseling tools, empowering students to make informed decisions about their future paths.

INTRODUCTION:

Background and Motivation: Each year, around 20,000 CSE graduates in Bangladesh compete for 7,500–8,000 technical roles, with over 80% failing basic skill assessments due to insufficient guidance. Many students struggle to align their skills and interests with career opportunities, leading to frustration and mismatched roles. Machine learning provides a data-driven solution to predict suitable careers and guide students effectively.

Objectives

1. Develop a machine learning-based career prediction model for CSE students.
2. Compare and evaluate algorithms on accuracy, precision, recall, and F1 scores.
3. Address gaps in previous studies by including multi-label classification and broader factors like job security.

Research Questions

1. Which machine learning model offers the best career predictions for CSE students?
2. How can diverse attributes be integrated into a prediction model?
3. What factors most influence CSE students' career decisions?
4. How does multi-label classification improve prediction?

RELATED LITERATURE REVIEW:

1. **Panday (2021)¹:** Predicted engineering students' careers using seven machine learning algorithms, achieving 53.03% accuracy with a Soft Voting Classifier.
 - **Gap:** Lacked focus on specific fields and skill-based career analysis.
2. **Madhan & Reddy (2021)²:** Developed a career prediction system using only two to three algorithms.
 - **Gap:** Limited algorithms hinder comprehensive comparisons and robustness.

¹ [Prediction Of Undergraduate Students' Career Using Various Machine Learning And Ensemble Learning Algorithms](#)

² [Career Prediction System](#)

3. **Ade & Deshmukh (2015)³**: Used a pair of classifiers for career choice prediction and donot mention the accuracy.
 - **Gap**: Insufficient algorithm and unavailable of accuracy.
4. **Cherry (2020)⁴**: Predicted employability with SVM, using mock interview scores and academic results.
 - **Gap**: Did not analyze broader attributes like skills, interests, and job preferences. Only predict on mock interview score.

Current Study: This research addresses these gaps by focusing on CSE-specific career paths, analyzing 24 attributes, and applying 10 machine learning algorithms to ensure accurate and comprehensive predictions tailored to students' skills and preferences.

RESEARCH METHODOLOGY:

Research Type: Quantitative study using machine learning to predict careers for CSE students in Bangladesh.

Data Description:

- **Sample:** 468 CSE students.
- **Features:** CGPA, technical and soft skills, publications, projects, and job preferences.

Methods:

1. **Data Preprocessing:** Cleaning, encoding, and normalization.
2. **Algorithms:** 10 machine learning models, including XGBoost, Random Forest, and SVM.
3. **Evaluation:** Metrics include accuracy, precision, recall, F1 scores, and execution time.
4. **Key Result:** XGBoost achieved the best performance with 61.70% accuracy and F1 macro of 0.69.

Outcome: predict how much the probability of each career is preferred for a CSE undergraduate student.

SIGNIFICANCE AND IMPLICATIONS OF THE STUDY:

Why the Problem is Worth Researching: With many CSE graduates in Bangladesh facing difficulties in choosing suitable careers, this study provides data-driven career guidance to help align student skills with job opportunities.

Why the Research is Unique: This research integrates technical and soft skills, using advanced machine learning models like XGBoost to predict career paths, offering a more comprehensive solution than previous studies.

How This Research Advances Knowledge: It enhances career counseling by incorporating multiple factors(skills, preferred-job, and work-life balance) and applying multi-label classification recommendations.

Limitations:

- Focused only on CSE students in Bangladesh.
- Relies on self-reported data, which may introduce biases.
- Less amount of data.

Timeline/ Schedule:

| Activity Information | Time (Month) |
|---------------------------------|--------------|
| Activity-1: Proposal submission | Month-1 |
| Activity-2: Data Collection | Month-2 |
| Activity-3: Analyze data | Month-3 |
| Activity-4: Report writing | Month-5 |
| Activity-5: Report writing | Month-5 |

.....
Supervisor Signature and Seal

Signature of the Board Members: (Use for presentation's Board)

1).....2)-----3)-----4) -----

Comment: i) Accept ii) Contitional Accept iii) Not Accept

Recommad for Project Exhibition: Yes / No

³ [Efficient Knowledge Transformation System Using Pair of Classifiers for Prediction of Students Career Choice](#)

⁴ [Predicting Students' Employability using Support Vector Machine: A SMOTE-Optimized Machine Learning System](#)