# 15th IEEE International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC-2022)

## Impact of Programming Language Skills in Programming Learning

Md Faizul Ibne Amin, Md Mostafizer Rahman, Yutaka Watanobe, and Muepu Mukendi Daniel
The University of Aizu


December 21, 2022

# Outline

- Introduction
- Background Study
- Motivation
- Objective
- Related Works
- Methodology
- Experimental Results and Analysis
- Conclusion and Future Work
- Reference

# Introduction

- Big Data: An extensive amount of digital data are being generated by the internet of things and information technology consistently.

- Research Scope: With this big amount of ICT data, research and development are being done and consistently extracting valuable features, knowledge, and information.

- Practical Skill: To improve logical and innovative thinking, problem-solving skills, implementation skills, and practical-based learning play a vital role.

- Programming Learning: The importance of programming learning does not only draw attention to the growing demand for ICT skills and programming education but also plays an important role in almost all other practical-based disciplines.

- Importance of Programming Language Skills: In programming learning, the impact of programming language is also important to enrich the programming and technical skill.

# Background

- The traditional programming learning environment in educational institutions is not enough to prepare skilled programmers due to the limited number of exercise classes, limited scope for practices, and lack of individual tutoring.

- The OJ systems have been introduced to cope up with this issue to enhance the scope and opportunities for practice and individual learning in addition to traditional classroom-based learning.

- Many educational institutes are being used OJ systems for conducting courses related to programming, software engineering, and ICT-related Disciplines.

- Consequently, a mentionable amount of programming-related data (e.g., submission logs, scores, source codes) are being generated consistently which can be worthy raw materials for further programming education research and development.

- Therefore, the main research objective of this paper is to use programming-related data for the analysis.

# Motivation

- Students can use to make their learning more meaningful and study effectively to reach their desired goal, there is no parallel of more programming practice to gain problem-solving skills and logical thinking.

- Our main motivation is to explore the impact of programming language skills in programming learning by comprehensive analysis through real-world e-learning platform data.

- We consider a practice-based exercise class of a basic computer programming course that consists of programming exercises and coding tests.

- Moreover, the analytical result of this paper can help students and programmers as well as the improvement programming learning.

# Objective

Research Objective: The key points of this research are as follows:

- The statistical analysis of the used programming languages for solving problems are presented.
- Single and multiple languages used for problem-solving and acceptance rate for both cases are considered.
- The average acceptance rate of users' based on AOJ systems, single and multiple languages used for solutions acceptance rate are presented.

- We expect that the analysis is helpful to students and novice programmers to improve their programming learning.

# Related Works

- Some recent research work related to Educational big data analysis and the development of programming learning has been taken into account.

| Numbers | Authors | Year of publication | Paper title | Publication |
|---|---|---|---|---|
| 1 | T. Saito and Y. Watanobe | 2020 | Learning Path Recommendation System for Programming Education Based on Neural Networks | International Journal of Distance Education Technologies |
| 2 | M. M. Rahman, Y. Watanobe, R. U. Kiran, T. C. Thang and I. Paik | 2021 | Impact of Practical Skills on Academic Performance: A Data-Driven Analysis | IEEE Access |
| 3 | M. M. Rahman, Y. Watanobe, T. Matsumoto, R. U. Kiran and K. Nakamura | 2022 | Educational Data Mining to Support Programming Learning Using Problem-Solving Data | IEEE Access |
| 4 | L. Belcastro, R. Cantini, F. Marozzo, et al | 2022 | Programming big data analysis: principles and solutions | J Big Data |
| 5 | W. X. Zhao, W. Zhang, Y. He, X. Xie, and J. Wen | 2018 | Automatically Learning Topics and Difficulty Levels of Problems in Online Judge Systems | ACM Transaction |

# Methodology

Data Used for the experiment: This research has conducted by using the data from the Aizu Online Judge (AOJ). AOJ and its general features are introduced As follows:

- The AOJ system is a very popular OJ system in Japan as well as worldwide. It contains around 3,000 problems, about 100,000 registered users, 6 million source codes, and submission logs and started its functionality in 2004.

- The AOJ system is officially used for programming and algorithm-related courses at the University of Aizu, Japan.

- The source code of the AOJ system has been used in IBM's research project “Project CodeNet”

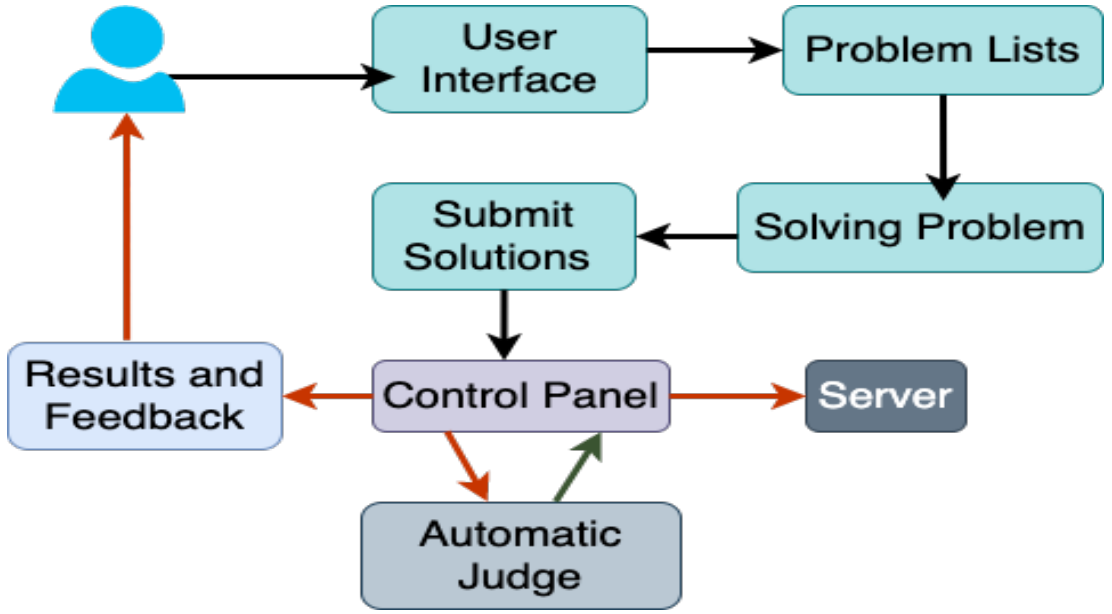


Figure: Overview of the OJ System

# Methodology (Cont.)

## Proposed Approach

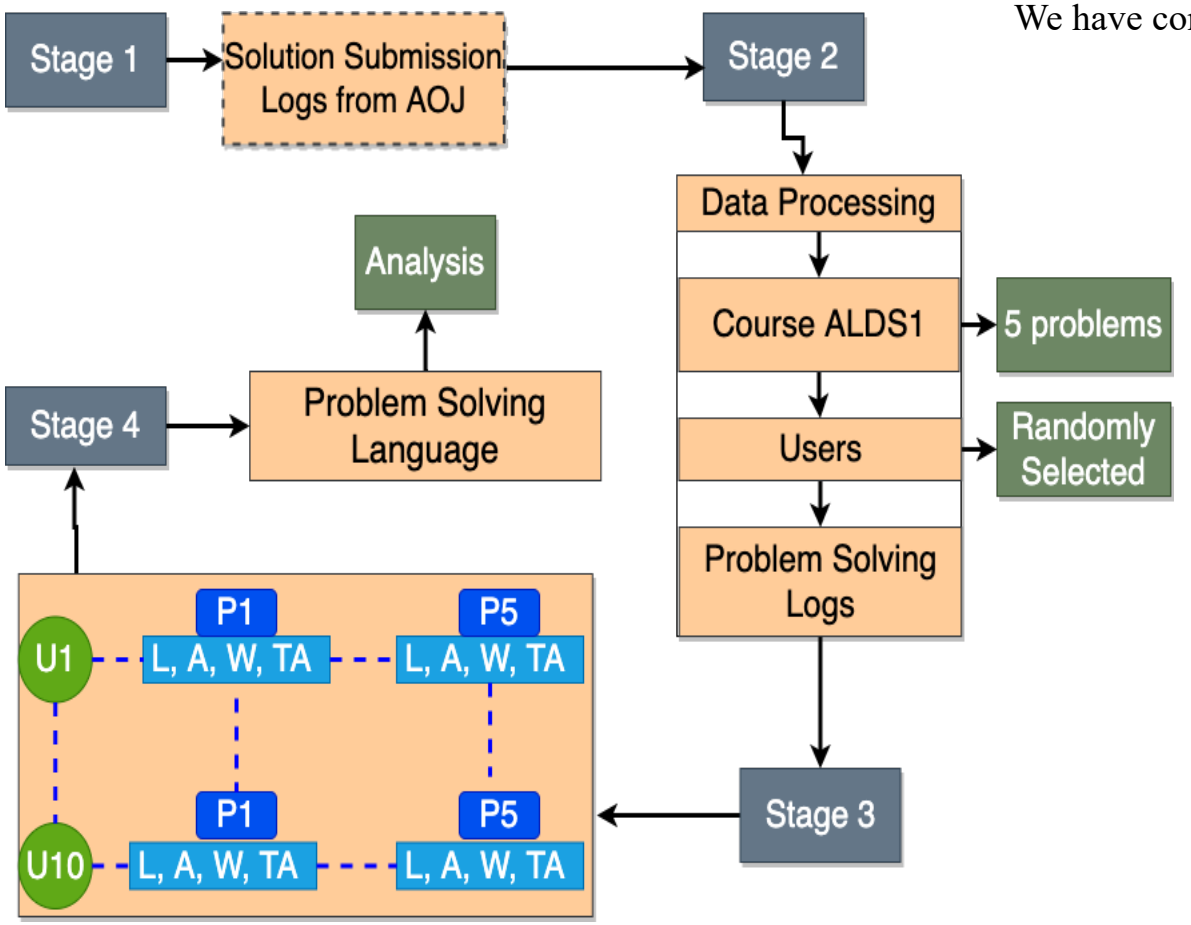


Figure: Overview of the proposed approach

We have conducted our analysis by following four stages as shown in the Figure.

**In stage 1**: The data are collected from AOJ.

- The targeted or main dataset as a total value of **3,932,501** is used, among them, a value of **825,401** data are used related to Algorithm and Data Structure I (ALDS1) course.
- Five problems of ALDS1 are selected and the total attempting value for problem-solving data is **343,205**.
- Then selected the unique user as a value of **26,338** who solved all the five problems.
- Finally, 10 users have picked up for our analysis initially.

# Methodology (Cont.)

**In Stage 2:** The data have been processed**.**

- The data from course ALDS1 is selected for the experiment.

- As a part of processing and to get efficient data, only User id, Problem id, used Languages for problem-solving, and Verdict is taken into account among many attributes.

- As five problems of the ALDS1 course has selected which are shown in Table

| Course | Problem Id | Problem Name |
|---|---|---|
| ALDS1 | ALDS1_1_A | Insertion Sort |
| | ALDS1_1_B | Greatest Common Divisor |
| | ALDS1_1_C | Prime Numbers |
| | ALDS1_1_D | Maximum Profit |
| | ALDS1_2_A | Bubble Sort |

Table: Problem list

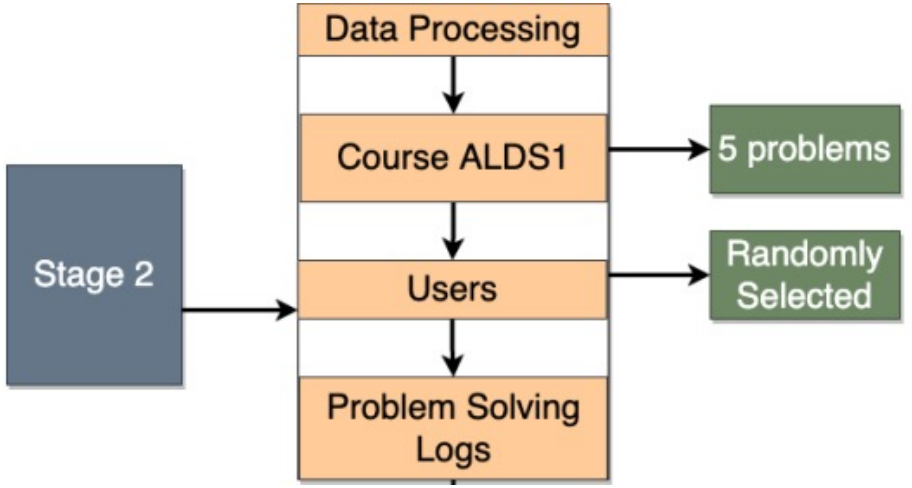


Figure: Stage 2

- After that, 10 users are selected randomly who solved all the five problems and collected their problem-solving logs for the experiment.

# Methodology (Cont.)

- **In stage 3**: All the 10 user's total attempts, total unique languages used for the problem-solving, verdict (Accepted, Wrong Answer), and languages used for the acceptance of all 5 problems are considered.

- **In stage 4:** we have focused to do an analysis by considering two points such as acceptance rate based on AOJ and acceptance rate based on languages used. For the first case, the equation for the calculation of the general acceptance rate according to AOJ is as follows: $\boldsymbol{U_n = \sum_{i=1}^{n} \frac{AC_i}{TA_i}}$

Where, $AC$= Acceptance rate, $TA$= Total Attempts

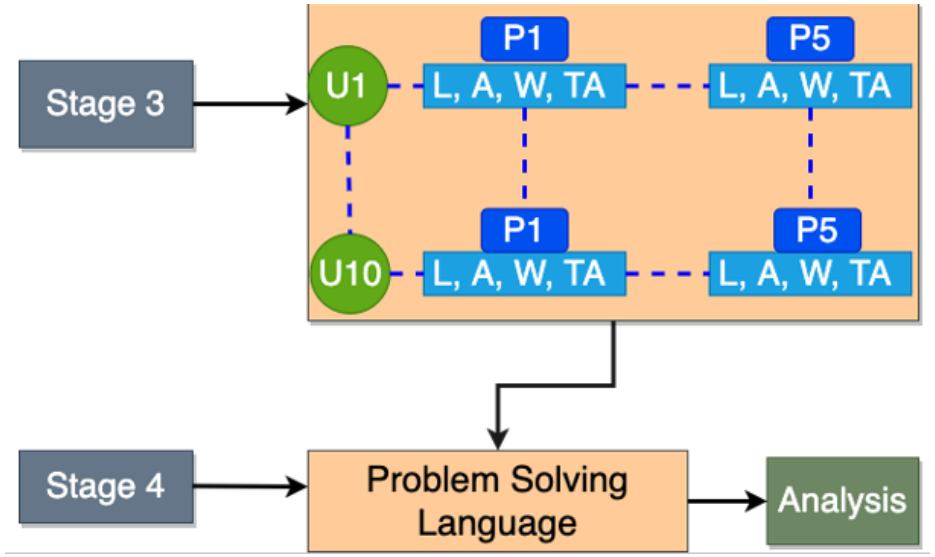


Figure: Stage 3 and 4

After that, the acceptance rate based on the languages used for the problem-solving is calculated as follows: $\boldsymbol{U_n = \sum_{i=1}^{n} \frac{\frac{AC_i}{TA_i} + \frac{LU_i}{TUL_i}}{2}}$

Where, $LU$ = Language Used, $TUL$ = Total Unique Language Used.

# Experimental Results and Analysis

The experimental results are presented here, and based on the result, the features and impact of the programming language skills have been analyzed.

The overall average acceptance rate of all the 10 users is **39.95%.** Next, we have considered the single and multiple languages used for problem-solving.

**For the case of single language used for problem-solving**

| User | Acceptance rate based on AOJ | Acceptance Rate based on Language |
|---|---|---|
| $U_2$ | 33.3% | 29.16% |
| $U_6$ | 21.73% | 23.36% |
| $U_8$ | 54.16% | 39.58% |
| $U_9$ | 40.54% | 32.77% |
| **Average** | 37.43% | 31.21% |

Table: Single language used for solving

- The average AOJ acceptance rate is **37.43%** and the average acceptance rate based on language is **31.21%.**

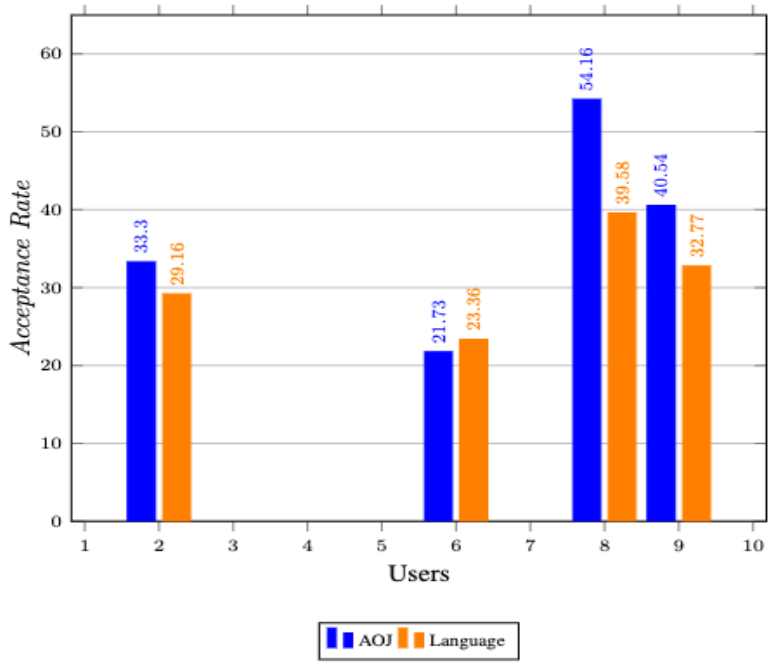


Figure: Comparisons of acceptance rate between AOJ and Language for single language used

# Experimental Results and Analysis (Cont.)

**For the case of multiple language used for problem-solving**

| User | Acceptance rate based on AOJ | Acceptance Rate based on Language |
|---|---|---|
| $U_1$ | 35.29% | 42.64% |
| $U_3$ | 25.80% | 50.40% |
| $U_4$ | 68.75% | 71.87% |
| $U_5$ | 63.63% | 56.81% |
| $U_7$ | 20.83% | 36.41% |
| $U_{10}$ | 35.48% | 42.74% |
| **Average** | 41.63% | 50.14% |

Table: Multiple languages used for solving

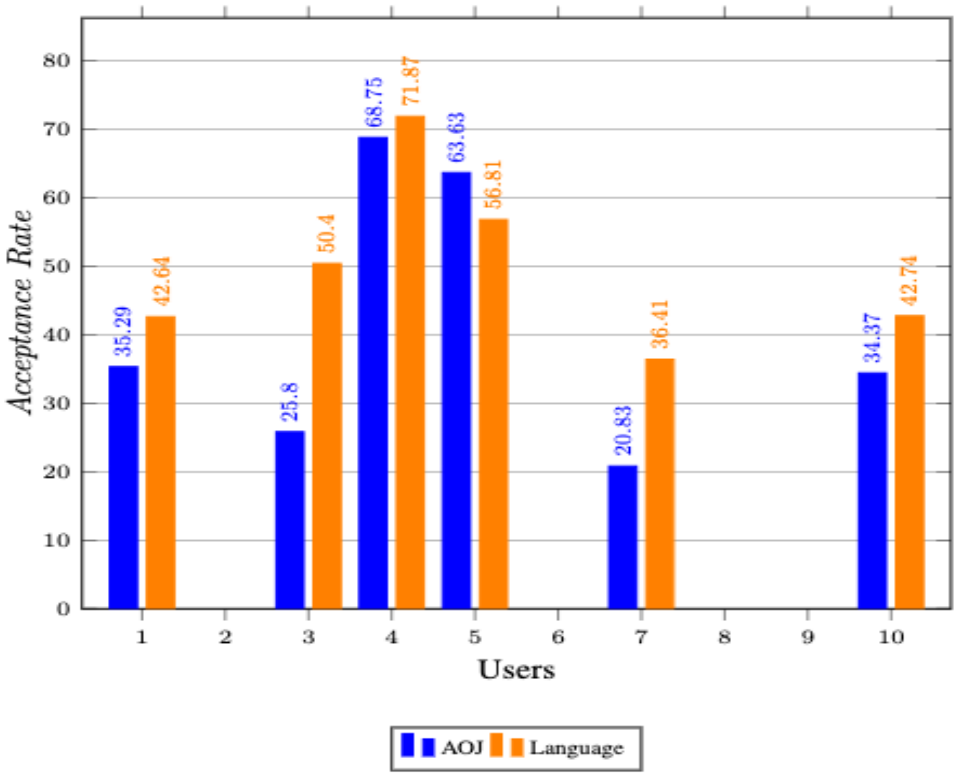


Figure: Comparisons of acceptance rate between AOJ and Language for multiple languages used

- The average AOJ acceptance rate is **41.63%.** On the other hand, in the case of used language for problem-solving, the average acceptance rate is **50.14%.**

# Experimental Results and Analysis (Cont.)

- Based on the experimental results according to single and multiple languages used for problem-solving, the acceptance rate is higher for the users who have used multiple languages than the users who have used single language in both AOJ-based and Language-based cases.

| Used Language | Acceptance rate based on AOJ | Acceptance Rate based on Language |
|---|---|---|
| **Single** | 37.43% | 31.21% |
| **Multiple** | 41.63% | 50.14% |

Table: Comparison of single and multiple used languages based on acceptance rate

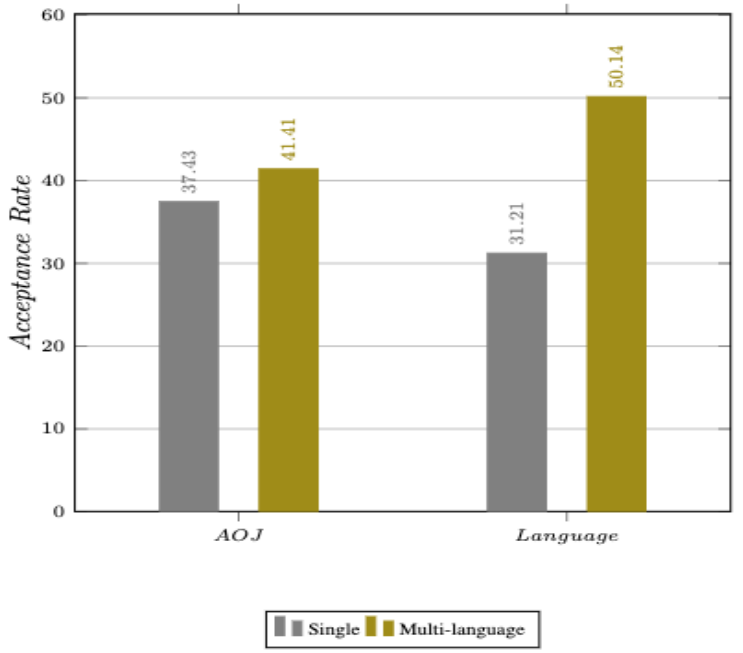


Figure: Comparisons of acceptance rate between AOJ and Language

# Experimental Results and Analysis (Cont.)

- In addition, when only one language was used to solve the problem, the average acceptance rate by the AOJ was **34.81%** and the acceptance rate by Language was **29.90%.** When more than one language was used to solve the problem, the acceptance rate by AOJ was **47.39%** and the acceptance rate by Language was **54.18%.**

| Single language used for solving | | | Multiple language used for solving | | |
|---|---|---|---|---|---|
| User | Acceptance rate based on AOJ | Acceptance rate based on Language | User | Acceptance rate based on AOJ | Acceptance rate based on Language |
| $U_{11}$ | 33.3% | 29.16% | $U_{12}$ | 46.42% | 48.21% |
| $U_{13}$ | 21.21% | 23.10% | $U_{15}$ | 26.08% | 38.04% |
| $U_{14}$ | 21.73% | 23.36% | $U_{16}$ | 52.58% | 66.27% |
| $U_{19}$ | 50% | 37.5% | $U_{17}$ | 83.3% | 66.6% |
| $U_{20}$ | 47.82% | 36.41% | $U_{18}$ | 28.57% | 51.78% |
| **Average** | **34.81%** | **29.90%** | - | **47.39%** | **54.18%** |

Table: Information of additional 10 users

# Experimental Results and Analysis (Cont.)

- Analysis shows that most of the users (among 10) have used one language (e.g., C) to solve the problem. With proceeding to the subsequent problems, different programming languages have been used (e.g., C++, Python, and Java) for problem-solving.

- The users who have used a single language for problem-solving, the attempt rate is reduced in solving next problems.

- It can be seen that language skills are improving in subsequent problem-solving.

- In the case of multiple languages used, users are not maintained a particular set of programming languages.

- The learning orders of programming languages can be impactful. We have found some language orders such as C to Python and C to C++. The preliminary results and analysis of the current study can be a good basis for the future research to determine the language learning order.

# Conclusion and Future Work

- A statistical analysis has been presented for exploring the effects of programming language skills in programming learning. In a computer programming course data from AOJ have been selected for analysis.

- The analysis has been conducted based on single and multiple languages used for problem-solving including the acceptance rate for both cases.

- Based on the results, we have realized that multiple languages used for the solution have a higher impact on the acceptance rate than the single language used.

- These analyses can effectively contribute to the improvement of students, novice programmers as well as overall programming learning.

- The preliminary results and analysis of this study can be a good basis for the future research scope to determine the language learning order and language recommendations for the specific problems.

- In future work, we will consider more data (e.g., users and problems) for exploring additional features and co-relation using the machine learning model.

# References

[1] T. Saito and Y. Watanobe. (2020), “Learning Path Recommendation System for Programming Education Based on Neural Networks, " International Journal of Distance Education Technologies. 18. 36-64. 10.4018/IJDET.2020010103.

[2] M. M. Rahman, Y. Watanobe, R. U. Kiran, T. C. Thang and I. Paik, ”Impact of Practical Skills on Academic Performance: A Data-Driven Analysis," in IEEE Access, vol. 9, pp. 139975-139993, 2021, doi: 10.1109/ACCESS.2021.3119145.

[3] M. M. Rahman, Y. Watanobe, T. Matsumoto, R. U. Kiran and K. Nakamura, “Educational Data Mining to Support Programming Learning Using Problem-Solving Data," in IEEE Access, vol. 10, pp. 26186-26202, 2022, doi: 10.1109/ACCESS.2022.3157288.

[4] L. Belcastro, R. Cantini, F. Marozzo, et al. “Programming big data analysis: principles and solutions," J Big Data 9, 4 (2022). https://doi.org/10.1186/s40537-021-00555-2

[5] W. X. Zhao, W. Zhang, Y. He, X. Xie, and J. Wen. 2018. “Automatically Learning Topics and Difficulty Levels of Problems in Online Judge Systems," ACM Trans. Inf. Syst. 36, 3, Article 27 (July 2018), 33 pages. https://doi.org/10.1145/3158670

Thank you very much for your kind attention!