

# Algorithm for Detecting Anomalous Student Activities in the Online Learning Process Based on Box Plots

Liliya A. Demidova

*Institute for Information Technologies  
Federal State Budget Educational Institution of Higher  
Education «MIREA – Russian Technological University»  
Moscow, Russia  
liliya.demidova@rambler.ru*

Anastas A. Misailidi

*Institute for Information Technologies  
Federal State Budget Educational Institution of Higher  
Education «MIREA – Russian Technological University»  
Moscow, Russia  
misailidi.a@gmail.com*

**Abstract**—The proposed research presents an algorithm for detecting anomalous student activities during online Python programming education, based on box plot analysis. Throughout their learning journey, students submit solutions to unique exercises across a range of tasks for assessment within the Digital Teacher Assistant (DTA) system. Anomalous activities are defined as those involving an excessively high number of solution submissions for a particular task or those characterized by extremely prolonged durations of task-solving. In both cases, students require timely assistance from instructors to clarify challenging course material. Upon each new submission of an exercise solution into the DTA system, the characteristics of the box plots for that specific task are recalculated. Decisions are made regarding whether the student’s submission constitutes an anomaly (outlier), either in terms of the number of submissions or the duration of task-solving. The experimental results, using a dataset collected during the spring semester of 2023 at RTU MIREA, indicate the practical viability of implementing the proposed algorithm for the real-time detection of various anomalous activities (outliers) in students’ interactions with the DTA system.

**Keywords**—*Anomalous activity, outlier, box plot, algorithm, Digital Teacher Assistant+*

## I. INTRODUCTION

Today education is undergoing revolutionary changes thanks to modern technologies and innovations. One of the key components of this transformation is educational data analytics, which opens up unique opportunities for educational institutions and educators to improve the quality of learning and adapt to the needs of each student. Educational data analytics is a powerful tool in education, enabling the extraction of valuable insights from vast amounts of information collected during the learning process and using them to optimize teaching strategies, create personalized courses, and enhance the overall effectiveness of education.

Educational Data Mining (EDM) is a field that focuses on applying data analysis methods to extract valuable knowledge and patterns in the educational environment. EDM utilizes various methods and techniques to analyze data collected from various educational sources such as e-textbooks, online courses, and learning platforms.

As of the current moment, tracking underperforming students and the field of educational data mining remain pivotal in the realm of education. Modern technologies and analytical methods continue to play a crucial role in the analysis of vast datasets collected during online learning and courses [1]. Presently, the primary focus is on identifying complex patterns and trends related to student performance, as well as providing individual recommendations to optimize the learning process. State-of-the-art algorithms are designed to proactively identify student challenges, enhance their motivation and engagement in learning, and reduce student attrition from educational programs [2]. This area of scientific research continues to evolve actively, offering endless prospects for improving the quality of education and achieving greater efficiency in the field of education.

In EDM, the following data analysis tools are considered.

- **Cluster Analysis.** This tool allows for the grouping

of students into clusters based on their activity and learning performance. It can help identify different learning styles and student needs [3].

- **Classification and Prediction.** This tool enables the creation of models using machine learning algorithms to predict student success, identify outlier risks, or recommend individual learning paths.

- **Association Rules.** This tool helps discover relationships and dependencies between student actions in the educational environment.

- **Factor Analysis.** This tool helps uncover hidden factors influencing student success, such as motivation, preferences, and the educational environment.

- **Text Data Mining.** This tool is used for analyzing textual data, such as essays, reviews, and student comments, to identify trends and sentiments.

- **Data Visualization.** This tool allows educators and administrators to present information about students and educational processes in a more visual manner, aiding in making well-informed decisions [4].

- **Data Analysis in Recommender Systems.** This tool uses data analysis to provide students with course recommendations, materials, and learning strategies based on their previous experiences and interests [5].

- **Graph Analytics.** This tool enables the analysis of students’ social networks and interactions within the educational environment.

These tools are used in EDM to optimize the educational process, improve learning outcomes, and create more personalized learning experiences for students [6].

Massive online education, especially in the field of programming languages, is rapidly evolving. With the advent of online platforms and courses such as Coursera, edX, Codecademy, and many others, students and professionals worldwide have gained access to extensive libraries of educational resources, allowing them to learn programming languages at their own pace and at a convenient time. This has led to the democratization of education and increased accessibility of programming knowledge for everyone, regardless of their location and financial resources. Many of these courses offer interactive tasks, projects, and practical exercises, which contribute to a deeper understanding of programming languages and the enhancement of IT skills. Massive online education in the field of programming languages has truly transformed people’s ability to learn and apply technological skills, making programming more accessible and inclusive for all [7].

Timely identification of learners experiencing difficulties in mastering specific topics within a course is of significant interest, with the aim of enabling timely intervention by educators to address emerging issues, such as providing additional materials or explanations.

## II. PROBLEM STATEMENT

The discipline “Python Programming” at RTU MIREA provides students with the opportunity to solve problems online on the corresponding website, demonstrating the modern direction of online education at its best. This discipline serves

as a shining example of the convergence of modern technology and education, offering students a more interactive and practically oriented educational experience [8].

Studying this discipline involves using DTA, which offers students unique exercises for various types of tasks. This enables students not only to study theory but also to immediately apply the acquired knowledge in practice. Interactive assignments and practical exercises, created in a digital environment, foster a deep understanding of the programming language and its application in real-life situations. Moreover, the online format allows students to learn at their convenience and pace, which is particularly important in today's fast-paced life.

This discipline also exemplifies the principle of educational accessibility since it is accessible online and can be studied by students from different geographical regions, contributing to the global democratization of programming knowledge. Online resources of this kind provide unique opportunities for students, helping them develop sought-after skills in the field of information technology in an interactive and engaging manner.

What makes DTA particularly valuable is that each problem variant is autogenerated. This means that students are provided with different versions of the same problem, ensur-

ing the fairness and reliability of assessment. It also encourages students to approach problem-solving based on their understanding of the material rather than mechanical repetition of answers [9].

When a student completes an exercise for a problem and submits their solution to the DTA system, it is automatically checked by the system for correctness. This ensures quick feedback and instant results. Furthermore, students' solutions are automatically clustered based on their approaches, further motivating students to explore all possible ways to solve a single problem. If a student successfully completes exercises for all 11 mandatory tasks, they have the opportunity to undergo intermediate assessment in the form of a credit, confirming their knowledge and skills in programming [10].

Data collection and dataset creation based on learning outcomes are important steps that can provide valuable information for analysis and decision-making in the educational environment. This article analyzes a dataset containing information about the interactions of second-year students at RTU MIREA with the DTA system while studying the "Python Programming" course in the spring of 2023. The dataset examined in the proposed research includes the following columns [11].