Comparing Perception and Reality: Exploring Test Complexity and Student Performance in Higher Education

Ivica Pesovski[1], Ricardo Santos[2], Vladimir Trajkovik[3], Roberto Henriques[2]

[1] Brainster Next College, Skopje, Macedonia

[2] Nova Information Management School, Universidade Nova de Lisboa, Campus de Campolide, Lisboa, Portugal

[3] Faculty of Computer Science & Engineering, Ss. Cyril and Methodius University, Macedonia

This is the author accepted manuscript version of the article published by IATED:

Pesovski, I., Santos, R., Trajkovik, V., & Henriques, R. (2023). Comparing Perception and Reality: Exploring Test Complexity and Student Performance in Higher Education. In L. G. Chova, C. G. Martínez, & J. Lees (Eds.), 16th International Conference of Education, Research and Innovation (pp. 6654-6660). (ICERI Proceedings). IATED Academy. https://doi.org/10.21125/iceri.2023.1660

COMPARING PERCEPTION AND REALITY: EXPLORING TEST COMPLEXITY AND STUDENT PERFORMANCE IN HIGHER EDUCATION

I. Pesovski[1], R. Santos[2], V. Trajkovik[3], R. Henriques[2]

[1]Brainster Next College, Skopje (MACEDONIA)
[2]NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, 1070-312, Lisbon (PORTUGAL)
[3]Faculty of Computer Science & Engineering, Ss. Cyril and Methodius University (MACEDONIA)

Abstract

Building on previous work where we investigated the correlation between perceived test complexity and student performance using a subset of data from a single subject, this study expands the scope to provide a more comprehensive analysis. This time, we engage with a more robust dataset that spans an entire academic year, including 1801 records from 47 students across 9 subjects, covering 89 distinct activities.

Our aim remains constant – to explore whether there is an alignment between the professor's and students' perceptions of assignment complexity and how these perceptions influence the students' final performance. Professors were asked to rate the difficulty level of each assignment they administered, and students were requested to provide their perceived complexity upon submitting their solutions. Notably, these perceptions were collected independently, with neither party privy to the other's assessment. Student time on assignment was also measured by looking at the assignment start date, submission date and assignment due date. Two variables were extracted from this to help better understand student perception of assignment difficulty by looking at the time each student spent on a given assignment.

We employed the *k*-means clustering algorithm to discern patterns in the relationship between the differences in complexity perceptions and student performance. The findings revealed a clear trend: students who perceive an assignment to be more difficult than their professor's rating tend to achieve lower scores. Although exploratory in nature, these findings have significant implications. They hint at the potential of leveraging perceptions of complexity as a tool for early identification of students likely to excel or struggle in a subject. This research reaffirms the importance of considering perception data as a valuable adjunct to performance metrics in higher education.

These conclusions emphasize the importance of monitoring not only students' performance but also their perception of test complexity. This can provide a more holistic view of their learning process and help educators create a more supportive and effective learning environment.

Keywords: Student performance, student performance prediction, exam complexity, learning management systems, k-means clustering, student procrastination, cluster analysis, assignment difficulty, student assignment perception.

1 INTRODUCTION

Understanding how well students are doing in the world of higher education is frequently more complicated than simply looking at exam results. Nonetheless, a student's performance on exams is almost always the most important factor considered when determining whether or not the student will be recognized, receive a scholarship, or be admitted to a variety of schools. Although each student has a different approach to learning [1], grades are most often formed using uniform scales, not considering variance in student personalities. Some researchers have already proposed measures to evaluate academic achievements in terms of an individual student’s goals and aspirations [2, 3]. Most higher education institutions also tend to evaluate student’s knowledge on semestral exams, a practice that offers a great room for improvement by utilizing a more frequent evaluation window [4, 5]. Studies analyzing predictors of student success at university have shown that these predictors vary from faculty to faculty [6]. Moreover, the grade performance has proven to be influenced by various parameters like student’s motivation, beliefs or attitude [7]. In today’s world most educational institutions utilize some kind of learning management systems (LMS) to enhance their student's experience, which can also influence the student's academic achievement [8, 9]. There is also a huge change in who visits universities as today's students belong to generation Z, a generation that completely differs from its predecessors - the Millenials [10]. The different mindset and approach that these students have toward learning and toward life in general, can be used to further adapt the way students are examined and have their knowledge graded.

Taking into consideration these challenges for giving a fair and correct grade to each student on each subject, we tried to look for various self-reporting measures and characteristics that can be used to better predict and evaluate student performance. One of these included analyzing student's personality traits to forecast their success in programming courses [11]. The current study was inspired by other researchers who found a clear correlation between student perceived test difficulty and exam results [12, 13, 14]. Students and professors tend to evaluate the exam complexity differently [15]. Students usually consider the majority of exams to be more difficult than they actually are [16]. Still, it looks like students are generally more accurate judges of task difficulty than teachers. For controlling the quality of the assignments, their reflection of the subject, and the degree of skill students should have to pass it, consistency between students and professors is crucial.

In an effort to build upon our previous work on how student perception of exam difficulty correlates to exam grades [17], we decided to replicate the study by using a larger dataset and evaluating the results on different study subjects. In our previous research, we focused on analyzing the correlation between exam complexity perception and exam grade on students attending the structured programming subject at a software engineering college program. In that work, we identified 2 groups of students who are statistically different among each other regarding their success on the programming course. One group rated most of the exams as being harder than what the professor rated them to be, and the other group rated most of the exams as being easier than what the professor rated them to be. The first group tended to score worse, while the latter group that perceived exams as easier than the professor did, tended to get greater results on the same exams. These two clusters of students were clearly divided and there were no larger ambiguities between each cluster's members. These results are useful for better adjusting the learning material, early spotting huge discrepancies in test complexity and providing better overall experience for students. In this paper we are looking at the complexity levels and exam grades students got on 9 subjects, including programming, mathematical and social science subjects. Because the 9 subjects include different activity types throughout their duration like homeworks, laboratory experiments, exams, tests, quizzes and similar, in the wording of this paper, the words exam and assignment are used interchangeably throughout this paper.

We included the time variable in our research after being further motivated by the work of Yang [18], who investigated how the habit of student procrastination affects the overall performance of the students. In his research, Yang took into consideration the start date of the assignment, the date when students first viewed the task, the submission date, and the due date in order to evaluate the amount of time that each student spent working actively the amount of spare time each had on a certain assignment and how this affected their final marks. In the current study that we are conducting, we took into account both the amount of time that each student spent working on each individual assignment as well as the amount of time that remained until the deadline. This was done in order to gain a better understanding of the clusters as well as the ways in which differences in perceived difficulty, in conjunction with the amount of time that an individual devotes to a specific task, reflect on the overall performance of the student on that particular task or assignment.

Cluster analysis has been chosen as the main methodology for performing the analysis, since it has been established as a suitable technique for analyzing educational scenarios [19, 20]. Clustering offers an appropriate way to analyze the huge amounts of server log data to understand student learning and thus enable educational data mining (EDM). In this paper we are using k-means, a non-hierarchical clustering algorithm, to analyze characteristics of student behavior [21]. This type of analysis is suitable for testing our hypotheses as it groups the students into clusters or groups so that students within a given group are similar to each other and different from the students in other groups. We look into difficulty perception so we expect to have a group of students who consider assignments as easier than the professor and a group of students who consider it as more difficult. Clusters like these can be further divided in more granular ones, by introducing other variables. In our study we are looking into the time it took each student to submit a solution from the time the assignment became available and expect to have the initial cluster further divided regarding the proportion of time each student spent on a given assignment.

2 METHODOLOGY

The experiment that forms the basis of this study was carried out by 47 first-year students who were enrolled in a software engineering program over the entirety of the academic year. The students were enrolled in a total of 9 classes that required them to turn in graded assignments, and they were responsible for a combined 89 of these assignments throughout the course of the school year. The dataset contains 1801 records, with each record containing information on a student's grade and the level of difficulty they rated each assignment as having. Important dates for each assignment were also logged. These dates include the date each assignment became available (or the assignment start date), the date it was due and the date on which each student submitted the solution.

For delivering the activities to the students, the professors used a custom-built learning management system (LMS). When scheduling an assignment through this platform, the professors had to fill certain fields, like which students should see the assignment, when it should become available, and then a required field to choose the assignment difficulty on a 1-10 scale. When the students received the assignment, they were not aware of the difficulty level their professors chose. When uploading the assignment solution, the students had to fill a required field indicating the assignment difficulty. The professor would then proceed to check and grade each assignment. While reviewing the assignment, the professor could see the difficulty level each student chose. Each professor had to insert the grade for each assignment in the same learning management system. For each assignment the LMS tracks the assignment start date, due date and student submission date. We use these dates to analyze the time each student spent on a given assignment. The workflow described here left us with data in the data format given in Table 1.

Table 1. The data format of the dataset used in the research

Activity	Student difficulty level	Professor difficulty level	Points available	Score	Start date	Due date	Submission date

The first step that we took was to normalize the final score to a maximum of 10 points. This was necessary both because various assignments were assessed differently and so that we could maintain consistency throughout the analysis. Assignments that were evaluated on a scale ranging from 0 to 100 as well as a scale ranging from 0 to 5 were normalized in this manner together with the results that students obtained on them. As a consequence, every assignment score is displayed on a scale ranging from 0 to 10.

The length of time that was allocated to do each assignment and hand in the solutions varied from one assignment to the next. Some assignments, such as laboratory exercises, quizzes, or semester exams, lasted for two to three hours, while others, such as homeworks, lasted for several days or even longer. We extracted two additional variables in order to examine the amount of time each student spent working on a certain assignment in a consistent manner, beginning with the date the assignment was assigned, continuing with the date it was submitted, and ending with the date it was due. One of these new variables is the percentage of time students spent on assignment, and the other was the percentage of time that remained before the assignment was due. This way, regardless of whether the task was due in a few hours or many days, we may determine, for example, that a student used forty percent of the time allotted to complete the work, while the remaining sixty percent of the time was free time.

After preprocessing the data to have it in the desired format and anonymizing it to preserve student's data privacy, we utilized the *k*-means clustering algorithm to partition the students into a certain number of clusters, each containing most similar students. Clustering algorithms have been applied in the context of education to pinpoint the various learning strategies students utilize on LMS and how these patterns connect to their general performance [22, 23]. Additionally, clustering has been used to categorize students according to their preferences and use of various kinds of educational materials [24]. Clustering approaches have been used to predict the difficulty of particular programming issues, with a high accuracy score [25], in a somewhat different context. Our implementation of *k*-means focused on adopting 2 different perspectives: perception of exam difficulty, featuring the variables associated with how both professors and students perceive the difficulty of the grading event, and time spent on the assignments containing variables measuring how much time it took for them to submit their assignments relative to the time they had available. Both perspectives were combined to obtain our final clustering solution.

3 RESULTS

In this section, we reveal and discuss the results of our analysis. The analysis had a qualitative component, where we looked to discover in which variables the resulting clusters would be different from one another. This analysis was complemented with a quantitative analysis featuring how the different clusters are distributed across 3 different variables. The first variable was *level_difference* which measures the difference between the professors' perceived level of difficulty of the assignment and the student's perceived level of difficulty. The second variable, *time_spent_exercise_%* measures the proportion of the total time available that had passed until submission relative to the total time available (meaning that values greater than 1 represent submissions made after the due date). The third variable, *points_acquired*, refers to the total amount of points received by the student for that assignment (normalized between 0 and 10).

3.1 Qualitative analysis

Figure 1 presents the qualitative profiles of the distinct clusters obtained from our approach. We identified four unique groups. Clusters 1 and 2 consist of students who perceived the assignment as easier than their professor did and, on average, scored higher grades than their peers. The key difference between these clusters is the time spent until submission, with students in cluster 1 dedicating significantly more time than those in cluster 2.

Clusters 0 and 3 exhibit similar patterns in terms of difficulty and outcome but differ in the time elapsed from start to submission relative to the total time available for assignment completion. These findings align with our previous research [17], suggesting that discrepancies between student and professor assessments of assignment difficulty could be valuable indicators for predicting student performance.

Interestingly, contrary to our expectations, late submissions, who are usually associated with higher degrees of procrastination behavior, did not necessarily translate into poorer assignment performance.

Figure 1. General qualitative profiling of the clusters resulting from k-means. Left: vertical axis plot showing the qualitative results obtained from the final clustering solution (for the sake of comparison, all variables were normalized to the range [0,1], Right: Number of assessment submissions attributed to each cluster.

3.2 Quantitative analysis

To better understand the real-world implications of our qualitative differences, we analyzed three variables: *level_difference*, *time_spent_exercise_%*, and *points_acquired*. Figure 2 shows that students in clusters 1 and 2 perceived the assessments as less difficult than their professors did, by nearly 2 points out of 10. These students also scored higher on the assessments compared to those in other clusters. In contrast, students in clusters 0 and 3 perceived the assessments as more difficult than their professors did and scored lower grades.

The introduction of the percentage of available time used until submission adds another layer of analysis. While this variable doesn't seem to impact the final assessment result, it helps characterize different student behaviors. Cluster 0 includes submissions that used nearly 100% of the available time but performed relatively worse. Submissions in Cluster 1 took just as long but tended to receive near-perfect grades. Submissions in Cluster 2 were completed rapidly and also achieved better grades, while Cluster 3 represents fast submissions that resulted in poor performance.

We hypothesize that Cluster 0 submissions were made by students prone to procrastination, while Cluster 3 submissions were from students who may have answered impulsively without reflecting on their approach. Cluster 2 may include submissions from well-prepared students who are up-to-date with their materials, and Cluster 1 may represent students who are perfectionists and dedicate all available time to ensure they provide a correct answer. Further research is needed to confirm these hypotheses.

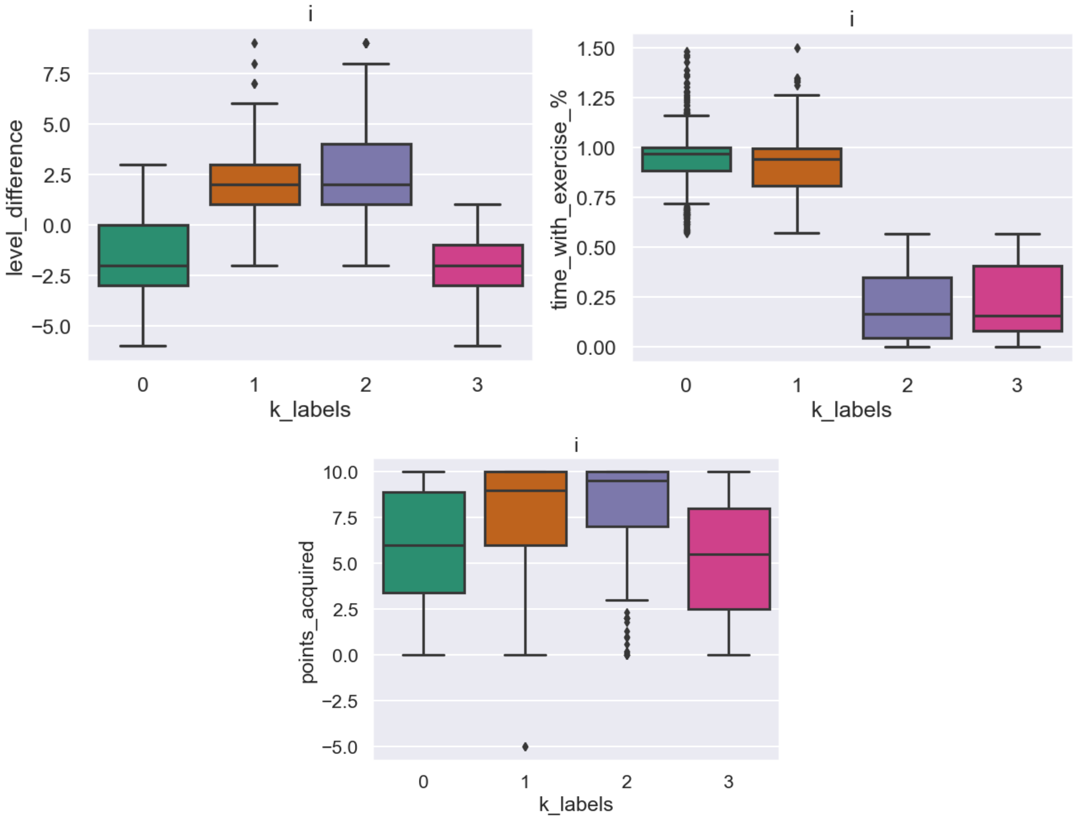

Figure 2. Boxplots showcasing the distribution of variables level_difference, time_with_exercise_%, and points_acquired across the different groups.

4 CONCLUSIONS

This research presented in this paper is a continuation of our earlier research in which we looked into how discrepancies in perception between students and instructors about the difficulty of assignments can reflect on the score that students earn on each assignment. In the previous study, only students who had taken the structured programming class in their first year of college were analyzed. In the current study, we work with the same students, but we take into account the difficulty perception differences and the score they got on assignments from all 9 subjects they had taken in their first year of college. Programming, mathematics, and various aspects of social science are included in these subject areas.

We also considered including the assignment start time and due date in addition to the time the student submitted the solution in order to extract two additional variables that we include in our analysis: time_with_exercise_% and time_left_%. This idea was inspired by other researchers working in the same field who also used the time variable in their analysis (which is described in the introduction section). The first thing to take into consideration is the amount of time that students actually spent working on their assignments out of the total amount of available time for solutions. It is expressed as a percentage due to the fact that different kinds of assignments from different classes typically have different allotted amounts of time to complete them; some of them allow hours, while others allow days for working on a solution. The second variable is the reciprocal of the first one and indicates the proportion of the available time for the assignment that students did not use. Alternatively, it indicates the amount of time that remains between the time the assignment was submitted and the date it is due. It is represented in percentages for the same reasons.

The results discussed in the previous section confirm our earlier findings, this time across all ranges of different subjects. Students who perceive the assignment as more difficult than the professor, once more prove to be worse on assignment results. On the contrary, students who perceive them as easier, tend to score better. The time variable helped us explain the 4 clusters better, that there are students who perceive the task as easier and either spent little or spend much time solving it and then students who perceive the task as more difficult and either spent only a little or they spend much time solving it. However, the time to solve the assignment doesn't seem to influence the assignment scores, which is counterintuitive since we would expect procrastinators and students who submit work near the deadline to have worse results than students who submitted in a timely fashion.

While doing this study, we identified certain areas that can be further analyzed in more detail in order to have more conclusive results. One of them is analyzing the clusters in the context of groups of subjects. In the current study we took the holistic approach to look at the complete dataset containing data from various subjects. There is an opportunity to divide the subjects by the area they cover, for example subjects covering programming topics, subjects covering mathematical skills, subject converting social sciences etc. and to perform the analysis in that manner. By doing so, certain patterns may prove to be valid only for a certain study field and more tailored conclusions can be drawn regarding the nature of the subject at hand. Another area where the current experiment can be extended is analyzing the assignments by their type. In this study, again, using the holistic approach, we treated different types of assignments in the same way which was the reason the time variables had to be represented in percentages. In our dataset the assignments are already separated by types like homeworks, laboratory exercises, tests, exams etc., each taking different time to complete. Different assignment types require different skills and approach to solve, which is why analyzing different assignment types separately can lead toward different conclusions regarding assignment type.

Finally, the importance of this study is that it confirmed the effect that perception toward a given assignment difficulty has on the success of solving the same assignment. In educational context, this can be utilized to better adjust the course curricula, offer more personalized learning approach to each student regarding his/her perception, early identify struggling students, provide a more targeted support to them, identify high performers, offer additional materials or learning opportunities to excellent students and by doing so improve the complete learning experience in general.

REFERENCES

[1] J. T. Richardson, "Students' approaches to learning and teachers' approaches to teaching in higher education," Educational psychology, vol. 25, no. 6, pp. 673–680, 2005.

[2] O. Guterman, "Academic success from an individual perspective: A proposal for redefinition," International Review of Education, vol. 67, no. 3, pp. 403–413, 2021

[3] U. Ononye, F. Ndudi, D. Bereprebofa, and I. Maduemezia, "Academic resilience, emotional intelligence, and academic performance among undergraduate students," Knowledge and Performance Management, vol. 6, no. 1, p. 1, 2022.

[4] I. Pesovski, A. Klashninovski, and A. M. Makeska, "The effect of frequent testing on student performance," 2022.

[5] I. Pesovski and V. Trajkovik, "The influence of frequent quiz-based exams on learning performance among computer science students," in ICERI2022 Proceedings. IATED, 2022, pp. 5693–5699.

[6] K. Danilowicz-Gosele, K. Lerche, J. Meya, and R. Schwager, "Determinants of students' success at university," Education economics, vol. 25, no. 5, pp. 513–532, 2017.

[7] J. M. Flom, K. Y. Green, and S. Wallace, "Helping higher education students succeed: An examination of student attributes and academic grade performance," Active Learning in Higher Education, vol. 24, no. 2, pp. 221–235, 2023.

[8] I. Pesovski, T. Dimitrova, and V. Trajkovik, "Educational achievement and satisfaction during traditional, online and combined classes among students following intensive workplace related courses," in 2021 19th International Conference on Information Technology Based Higher Education and Training (ITHET). IEEE, 2021, pp. 1–9.

[9] I. Pesovski, A. Klashninovski, A. Makeska, and V. Trajkovik, “The impact of learning management systems on student achievement and engagement in online classrooms,” in EDULEARN22 Proceedings. IATED, 2022, pp. 6451–6456.

[10] I. Pesovski, A. Kulakov, and V. Trajkovikj, “Differences in cognitive ability assessment results between millennial and generation z cohorts,” 2022.

[11] I. Pesovski, “Using personality traits to forecast grades of computer programming students,” in INTED2023 Proceedings. IATED, 2023, pp. 4286–4286.

[12] D. Scasserra, “The influence of perceived task difficulty on task performance,” 2008.

[13] F. Hammonds and G. Mariano, “Student test grades in college: A study of possible predictors.” International Journal of Teaching and Learning in Higher Education, vol. 27, no. 1, pp. 114–118, 2015.

[14] T. Stenlund, P.-E. Lyŕen, and H. EkÏof, “The successful test taker: exploring test-taking behavior profiles through cluster analysis,” European Journal of Psychology of Education, vol. 33, pp. 403–417, 2018.

[15] P. Tavakoli, “Investigating task difficulty: Learners’ and teachers’ perceptions,” International Journal of Applied Linguistics, vol. 19, no. 1, pp. 1–25, 2009.

[16] G. van de Watering and J. van der Rijt, “Teachers’ and students’ perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items,” Educational Research Review, vol. 1, no. 2, pp. 133–147, 2006.

[17] I. Pesovski, R. Santos, and S. Kagan, “Perspective matters: A preliminary look into the differences between professor and student perception of test difficulty and their relationship with test performance,” in EDULEARN23 Proceedings. IATED, 2023, pp. 5163–5169.

[18] Y. Yang, D. Hooshyar, M. Pedaste, M. Wang, Y.-M. Huang, and H. Lim, “Predicting course achievement of university students based on their procrastination behaviour on moodle,” Soft Computing, vol. 24, pp. 18 777–18 793, 2020.

[19] S. Ullrich-French and A. Cox, “Using cluster analysis to examine the combinations of motivation regulations of physical education students,”Journal of Sport and Exercise Psychology, vol. 31, no. 3, pp. 358–379, 2009.

[20] J. Kabok, S. Radisic, and B. Kuzmanovic, “Cluster analysis of higher-education competitiveness in selected european countries,” Economic research-Ekonomska istraˇzivanja, vol. 30, no. 1, pp. 845–857, 2017.

[21] P. D. Antonenko, S. Toy, and D. S. Niederhauser, “Using cluster analysis for data mining in educational technology research,” Educational Technology Research and Development, vol. 60, pp. 383–398, 2012.

[22] R. Cerezo, M. Sanchez-Santillan, M. P. Paule-Ruiz, and J. C. Nunez, “Students’ lms interaction patterns and their relationship with achievement: A case study in higher education,” Computers & Education, vol. 96, pp. 42–54, 2016.

[23] M. Riestra-Gonzalez, M. del Puerto Paule-Ruız, and F. Ortin, “Massive lms log data analysis for the early prediction of course-agnostic student performance,” Computers & Education, vol. 163, p. 104108, 2021.

[24] N. A. Stites, E. Berger, J. Deboer, and J. F. Rhoads, “A cluster-based approach to understanding students’ resource-usage patterns in an active, blended, and collaborative learning environment,” International Journal of Engineering Education, vol. 35, no. 6, pp. 1738–1757, 2019.

[25] C. M. Intisar and Y. Watanobe, “Cluster analysis to estimate the difficulty of programming problems,” in Proceedings of the 3rd International Conference on Applications in Information Technology, 2018, pp. 23–28.