Chapter 6

Prediction of Early Quitting Students in a Speed-Reading Course[1]

İnanç Kabasakal [2]

Abstract

Speed-reading courses are designed to improve their students' reading speed and comprehension. The use of e-learning environments enables data collection that helps in the assessment of students and predictive analyses. Student dropout prediction is among the popular problems in this context. This study presents a neural network-based prediction model that identifies impending dropouts from a speed-reading course. Assessment scores obtained in the last ten sections were analyzed to predict dropouts who will not proceed to the next level. Despite the challenge of predicting short-term dropouts, the tests resulted in an accuracy of 78.24% on average. Moreover, 56.58% of predicted students dropped out before the next level, while 52.67% of students were successfully identified just before the dropout.

1. Introduction

Speed-reading courses are designed to improve their students' reading speed and comprehension. These courses help to coordinate brain and eye movements so that the speed of visualization is increased as well as eyesight is enlarged (Öztahtalı, 2011), both helping to achieve progress in reading. There is a variety of online speed-reading courses that aim to improve learners' reading habits.

Web-based courses offered in *Learning Management Systems* provide the ability to track and assess learners' activity in a cost-effective manner

---

1 This chapter is a substantially revised and extended version of the paper entitled "Predicting Quitting in Students in Speed-Reading Course: A Case Study" presented in the 10th InTraders International Conference on Social Sciences & Education E-Conference held between the dates 21-22 June, 2023.

2 Assoc. Prof. Dr., Ege University, Department of Business Administration / Quantitative Methods, İzmir, Türkiye, inanc.kabasakal@ege.edu.tr, https://orcid.org/0000-0003-0098-0144

https://doi.org/10.58830/ozgur.pub574.c2357

(Psaromiligkos et al., 2011). Various data mining studies have focused on identifying similarities of learning progress and unusual patterns in learners' behavioral actions (Rodrigues et al., 2018). E-learning environments for speed reading help to collect data that primarily helps in tracking and assessing students. The use of collected data enables further analyses on learners' progress and take action.

The data analyses on records that represent users' behavior in web-based learning have been helpful in predicting the students who are likely to leave the course. In particular, this problem has been a popular focal point in *Educational Data Mining* studies. Educational dropout prediction has been described as *"a major, important, and challenging task for every education institution's administrator, policy maker, and educators"* by Kumar et al. (2017). Predicting such students is vital to prevent such incidents, as well as discovering the factors that might lead to that decision. Considering that educational institutions are organizations that serve students, predicting dropouts might be regarded as a part of managerial task that is analogous to customer retention.

This chapter focuses on the student prediction problem in an online speed-reading course. The course offers 28 stages of reading activities, where student activity in reading sessions is tracked on the web-based system. The web-based system conducts an assessment at the end of each reading activity, and the progress of students in both accuracy and speed are measured. Available data is used to develop a predictive model that makes use of artificial neural networks. A distinguishing feature of the predictive model is its short-term orientation. In a 28-stage course, the model is trained to predict only the students that will quit the course until the end of the current stage. The results demonstrate that the model predicts that 53% of the students that will quit the course in a short while. More generally, 57% of "dropout alerts" raised by the prediction model were found to be accurate.

The chapter is organized as follows: initially, the use of data mining in education is briefly discussed, and the opportunities in learning systems are explored. The student dropout prediction problem is then introduced with reference to popular approaches in prior studies. Next, the benefits of speed-reading courses are illustrated, and the trade-off between accuracy and speed is introduced as a challenge in this context. Subsequently, a case study is introduced along with details of the dataset being analyzed, and the methodology being utilized. The predictive model based on Artificial Neural Networks is presented along with the training stage and the test results. The predictive power of the model is presented in detail using accuracy, precision, and recall indicators. Finally, the conclusion includes practical implications of the study and further improvements for speed-reading courses equipped with dropout prediction capabilities.

2. Literature Review

The use of data mining techniques has been a popular field of study, often mentioned as Educational Data Mining (EDM). Data mining studies highlight best practices and benefits of data mining, especially in e-learning platforms where various types of activities are collected. This section briefly introduces the use of data mining techniques in education. Moreover, the section also covers 'student dropout prediction' as one of the prominent tasks in EDM studies.

Since the chapter presents a student dropout prediction model in a speed-reading course, this section also covers the 'speed-reading' concept along with the objectives, challenges, and benefits reported in prior studies.

2.1. Educational Data Mining

Educational Data Mining (EDM) can be briefly described as the use of data mining techniques in problems related to educational tasks. Within the concept of 'Knowledge Discovery from Databases', Data Mining is a step where useful and interesting patterns within data are discovered using algorithms from statistics, machine learning, or pattern recognition (Fayyad et al., 1996).

The emergence of EDM has been attributed to the increase in availability of educational data (Kumar et al., 2017), which provides the fundamental resource for knowledge discovery in the education domain. In the *Encyclopedia of Emerging Industries* (Gale Publishing, 2017), the use of data mining in education was related to identifying patterns to determine students who might have problems and drop out. The data used in this process was described as a "wide range of data in education" including attendance, demographics, discipline and medical records, grades, and immigration status. According to Guleria and Sood (2014), EDM is useful in discovering trends in educational data to cluster students based on their needs in education, improve students' learning and performance, and improvement in academic institutions by organizing in-house trainings.

EDM studies are related to various disciplines, including cognitive sciences, computer science, human-computer interaction, artificial intelligence, cognitive psychology (Abdulahi Hasan and Fang, 2021), web mining, e-learning, and text mining (Guleria and Sood, 2014).

According to Romero and Ventura (2007), knowledge discovery with data mining is described as a 'formative evaluation' technique that helps educators to design and improve learning environments and techniques in e-learning. According to the authors, data mining in web-based education systems deals with both academic data (courses, curriculum, etc.) and usage data originating from the interaction between the system and students.

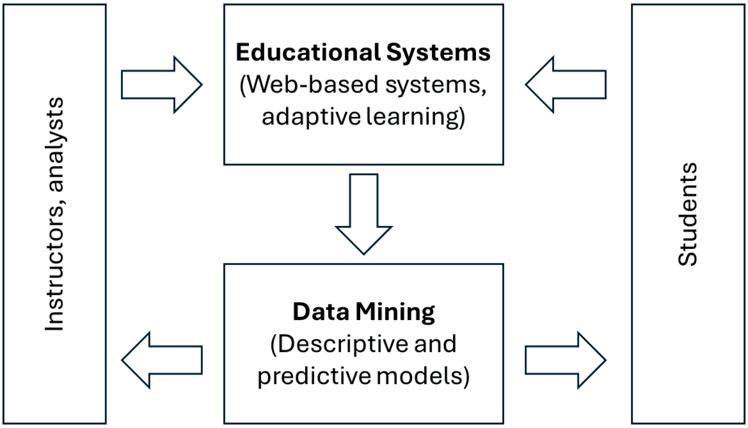

Figure 1 Interaction of students and instructors with EDM (adopted from Romero and Ventura, 2007)

As in Figure 1, EDM helps decision makers design and improve the education system and helps to personalize content through student-based recommendations to improve learning (Romero and Ventura, 2007). A significant difference in education from other domains is the organization of data in multi-level hierarchies since the educational activities take place in classes, departments, faculties, institutions that form a hierarchy (O'Connell and McCoach, 2008), which might be taken into account in EDM (Baker and Yacef, 2009). According to Dutt et al. (2017), this nature of educational data should be considered in data preparation and selection of appropriate algorithms in clustering tasks.

Student success prediction is one of the popular tasks in EDM. Various attributes that represent student's academic grades and attributes, characteristics of academic programs and courses, and external factors are used for this purpose. A data mining study by Lopes et al. (2017) reported that it is possible to predict the performance of master's students by utilizing classification techniques such as random forest, classification tree, support vector machines, classification and regression tree. Sequential records indicating students' exam performance were analyzed by Kuzilek et al. (2021) to predict student success in the academic year.

Predicting student dropouts is another popular EDM task that received extensive interest in prior research. The next subsection introduces several models for this task.

2.2. Student Dropout Prediction

Students quit their courses or education due to individual reasons or various contextual factors. According to Kinnunen and Malmi (2006), student dropout happens because of difficulties encountered while studying. In their study, the authors suggested that the primary reasons for dropout in a computer science course were the time constraints and lack of motivation in students.

In a study that investigated student dropout in an online course K12 students, De la Varre et al. (2014) identified reasons that relate to factors including time constraints, parental influence, under-preparation for academic rigor, and problems originating from online platforms. Another study (Goldschmidt and Wang, 1999) highlights that student dropout in schools might be related to being held back from school, misbehaving, family socioeconomic status, employment status, and intensity, as well as school-related factors.

A comparison of dropout rates recorded in online education and traditional education was reported by Lykourentzou et al. (2009). The findings of the study revealed that dropout rates were higher in web-based courses than those held in traditional education. According to Goldschmidt and Wang (1999), dropout rates differ across public and private schools, which also holds across secular and non-secular schools.

As reported by Xenos et al. (2002), it is a significant objective to prevent or reduce the rate of dropouts in distance education. Predictive data mining and machine learning models have often been utilized in student dropout prediction in distance education. A study by Kotsiantis et al. (2003) underlines that such models can achieve considerably high accuracy scores in limited datasets consisting of only demographic variables (63%), and in datasets involving student progress (83%) during the academic period. According to Kurtyılmaz and Ergün-Başak (2023), since student dropouts often reflect similar characteristics, dropout prediction is possible through analyzing school experiences, absenteeism, discipline problems, and a decrease in academic performance. Detecting students who have a great risk of dropping out is a crucial task that would provide significant benefits even after achieving tiny improvements in accuracy of student dropouts (Kurulgan, 2023).

2.3. Speed Reading Techniques

Speed reading techniques have been a necessity for individuals who need to catch up with the rapid transformation in the society, thus having too much to read and learn within limited time periods (Soysal, 2021). Speed reading techniques and training are designed to help individuals to overcome this problem and improve their reading skills. In this regard, speed reading can be described as a practical skill that facilitates learning more quickly. On the other hand, while speed-reading is often praised as an essential skill, Macalister (2010) quoted that many teachers have skepticism towards speed-reading and its usefulness. In a study that explored the attitudes of teachers towards speed reading, Kemiksiz (2019) noticed that optional courses that aim to present speed-reading techniques is necessary in formal education.

Despite its potential benefits, developing speed reading skills has difficulties. Readers have habits that limit their reading speed, comprehension, or both. One such habit is subvocalization, which corresponds to vocalizing words by readers in their minds while reading (Nushi, 2017). As reported in a study (Öztahtalı, 2011) in Türkiye, reading words one by one is another major habit that forces readers to spend their effort and time on each word.

Speed reading requires a combination of cognitive ability with motor competence, including eye movements (Hilmi and Alwi, 2023). Speed reading training is designed to improve the reading of the students in terms of speed and comprehension.

The rate of reading, or the reading speed, is an indicator of fluency in readers (Macalister, 2010). Adults with a college education often read at a rate of 200-400 words per minute (Rayner et al., 2016). Similar scores for average reading speed have been reported in various studies, including 250 words/minute (Guo and Wang, 2018; Hilmi and Alwi, 2023), 250-300 words/minute (Grabe, 2009; Oyamada, 2014), 150-250 words/minute (Tsvetkova, 2017), 244 words/minute (Dyson and Haselgrove, 2000).

Increasing the reading speed involves a trade-off between accuracy and speed (Rayner et al., 2016) or accuracy and earliness (Pan et al., 2024), where users occasionally prioritize one over the other. Another study (Dyson and Haselgrove, 2000) reported a trade-off between speed and comprehension level, particularly when reading text from a screen. Considering this trade-off, it might be argued that speed-reading should aim for an acceptable level of speed without sacrificing the level of comprehension. Prior research also suggests an "optimal" reading speed of approximately 300 words per minute (Carver, 1982), while readers can read quickly while achieving a satisfactory level of comprehension measured in terms of efficiency and accuracy.

Positive effects of speed-reading techniques on the reading speed and degree of comprehension of students have been reported in prior studies (Dedebali & Saracaloğlu, 2010). Such an effect was also confirmed in gifted students (Soysal, 2022). Similarly, Durukan's study (2020) reported that speed-reading training significantly improves the reading speeds and comprehension levels of secondary school students.

Education programs to improve reading ability are often aimed at increasing readers' visual perception, as well as eliminating negative reading habits, improving vocabulary, and organizing stimulants (Akçamete and Güneş, 1992). Towards assessment and training of reading speed, various technologies have been developed since the beginning of 20th century. The study by Abdul-Rab et al. (2023) presents a list of machines and tools including pacing machines, eye-tracking machines, reading rateometers, video-based tracking devices. Besides, there is a variety of web-based or stand-alone tools including *Rocket Reader* (rocketreader.com), *SuperReading* (superreading.com), *Spreeder* (spreeder.com), *Iris Reading* (irisreading.com), *Rev It Up Reading* (revitupreading.com), *Legentas* (legentas.com), BeeLine Reader (beelinereader.com), and *Kwik Reading* (https://www.kwikbrain.com/products/reading) that aim to improve the reading effectiveness of users. According to Nushi (2017), Spreeder helps to avoid some reading habits that have negative effect on reading speed.

Tsvetkova (2017) criticized speed-reading by highlighting the possibility of information over-saturation or information overload, which leads to obtaining too much information quickly without having the chance to digest new inputs with critical thinking. In this regard, "slowing down and understanding" is one of the recommendations by Bawden and Robinson (2020) to overcome with information overload, since understanding includes processing necessary information is a requirement that takes time after information consumption.

A study by Klimovich et al. (2023) reports that the positive effect of speed-reading training is limited to approximately 35 words/minute, and this improvement is mostly caused by the increase of readers' awareness of the reading process and avoiding the re-reading behavior. Another study (Sirait et al., 2020) which compared reading comprehension in students before and after speed reading exercises has reported statistically significant improvements with speed reading techniques.

2.4. Dropout Prediction in Speed-Reading

During the literature review, no publication was matched that specifically attempts to predict student dropouts in online speed-reading courses. However, a variety of student dropout prediction models exist, some of which have been mentioned beforehand.

In a relevant study, disengagement from reading instructional texts has been examined by Mills et al. (2014). In this paper, the authors examined features obtained from reading times, such as total reading time and maximum page reading time. Moreover, in addition to such measurements of time, the paper also includes progress data such as the location of quit, the proportion of text seen, and whether text is completed, or not.

3. Methodology

The methodology in this study includes analyzing student progress data in speed-reading courses to train a classification model that predicts whether a student will drop the course. As in the study by Mills et al. (2014), measuring reading times is taken as an indicator of student dropout. In particular, measurements regarding the reading behavior and the result (completed or quit) are available in an online speed-reading context, as in this study.

In addition to the reading speed, we also notice the need for assessing comprehension. As in several prior studies mentioned above (Rayner et al., 2016; Dyson and Haselgrove, 2000; Pan et al., 2024), the comprehension rate is considered an essential indicator that complements the speed when assessing the progress in speed-reading. By analyzing both progress indicators, we aim to train a neural network-based model that predicts student dropout.

The dropout prediction is considered as a prediction task that requires training and assessing classification models. Classification in data mining involves a supervised learning approach that involves creating a classification model from training data and assessing the model by predicting class labels for given data (Han et al., 2012). Training data is selected randomly from a dataset, where a portion of the existing data is excluded from the training step and reserved for testing. The predictive power of classification models is measured by generating predictions for the test data and checking predicted values with actual class labels. The assessment of predictions is often presented in confusion matrices (Xu et al., 2020), where predicted and actual class labels are compared as below:

Table 1 A confusion matrix

		Actual	
		Positive	**Negative**
Predicted	**Positive**	*True Positives (TP)*	*False Positives (FP)*
	Negative	*False Negatives (FN)*	*True Negatives (TN)*

Based on the numbers in the matrix; performance indicators, namely the precision, recall and accuracy, can be calculated as below:

$$\boldsymbol{Precision} = \frac{\sum \boldsymbol{TP}}{\sum \boldsymbol{TP} + \boldsymbol{FP}}$$

The precision criterion indicates the success rate in making correct predictions when predicting positive ones.

$$\boldsymbol{Recall} = \frac{\sum \boldsymbol{TP}}{\sum \boldsymbol{TP} + \boldsymbol{FN}}$$

The recall criterion is relevant to the success in identifying the positives as much as possible in predictions. Since recall gives the ratio of positives correctly identified, it is described as a measure of "completeness" (Han et al., 2012:368). Both terms 'recall' and 'sensitivity' are often used in this context. We have used the term 'sensitivity' in our results.

$$\boldsymbol{Accuracy} = \frac{\sum \boldsymbol{TP} + \boldsymbol{TN}}{\sum \boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{FN}}$$

3.1. Case Study

This chapter examines the student dropouts in an online speed-reading course, namely, *DNA Speed Reading Courses*, presented by Dijital Nesil Akademisi. The objective is to predict students who are likely to quit the course as early as possible. In this regard, data collected during the students' progress has been analyzed to develop a classification model.

Speed reading courses usually present a set of texts, preferably in similar lengths, to the learners. Next, multiple-choice questions are used for the assessment of reading comprehension (Macalister, 2010). The speed-reading training offered by DNA follow this approach. The speed-reading training consists of 28 chapters (stages), each of which consists of 2 or 3 subchapters. The duration of all stages of the training takes approximately three months.

During this, students might quit the training due to various reasons, which are out of the context for this study.

The objective of the study is to achieve early prediction of dropouts, which corresponds to predicting the students who will quit at the end of their current stage in their progress. To achieve this, predictor variables were derived from the reading activities completed, and assessments made. In this regard, indicators for comprehension and reading obtained through the last 10 stage were used to train predictive models to identify students who will quit training in the subsequent stage. Therefore, a limitation of the study is to predict dropouts in the students who has completed the first 10 stages (among 28) in the training.

3.2. Dataset Analyzed

The dataset analyzed involves 244 students who have enrolled in DNA speed-reading training. As mentioned earlier, this number contains only those who completed the first ten stages of the training. A comparison of regular students who continue training with those who quit has been presented in Table 2.

Table 2 Comparison of students that dropped out with entire registrants

	Count	Ratio (%)
Dropouts	61	25%
Regulars	183	75%
All	244	100%

The distribution of dropout students and the last stage they completed were listed in Table 3. Regular students who continue their training are also listed according to their current progress.

Table 3 Count of registered students and dropouts over stages

Stage	Dropouts	% of Dropouts	Regulars
27	9	14.75%	27
26	3	4.92%	9
25	2	3.28%	6
24	5	8.20%	15
23	5	8.20%	15
22	5	8.20%	15
21	4	6.56%	12
20	3	4.92%	9
19	2	3.28%	6
18	5	8.20%	15
17	2	3.28%	6
16	4	6.56%	12
15	3	4.92%	9
14	1	1.64%	3
13	1	1.64%	3
12	0	0.00%	0
11	3	4.92%	9
10	4	6.56%	12

As shown in Figure 2, the dropout counts fluctuate, and a surprisingly high number of students drop out after stage 27, which is only one stage before the finish.

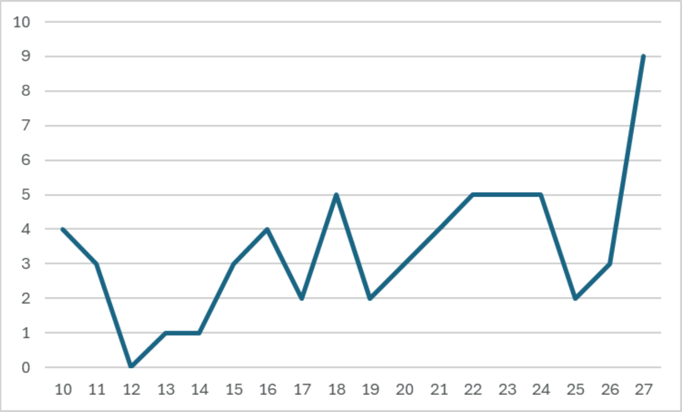

Figure 2 Number of students that drop out over stages of course

The dataset involves 244 students, each of which was represented with sequential numbers starting from 1. In addition to this identifier number, all the columns in the dataset represent the progress of students with scores achieved in various stages of the course. We note that the dataset involves no columns that might signify the identity of students, such as their name, surname, location, course registration number, national identity number, address, e-mail, or any other contact information in any form.

The dataset constructed for analysis consists of the following variables:

Table 4 Variables in Data

Variable	**Type**	**Description**
Max_Stage	Numeric (integer)	The last stage in student progress
Total_Activities	Numeric (integer)	Total number of activities completed
SpeedT01, SpeedT02, ..., SpeedT09, SpeedT10	Numeric (with decimals)	Reading scores measured after speed-reading sessions, numbered from 1 to 10.
CompT01, CompT02, ..., CompT09, CompT10	Numeric (with decimals)	Comprehension scores measured after speed-reading sessions, numbered from 1 to 10.
Status	Binary	Target variable to be predicted. - 0: regular student that continues to the speed-reading sessions - 1: dropout behavior

The speed variables represent the average number of words per minute measured for the last ten activities. The indices that follow T signify the recency of the score. The formulation of variables (numbered as SpeedT01, CompT01, …) are detailed as below:

Letting *S* as the last stage completed by a student, and letting that *X* is a number where $1 \leq X \leq 10$:

SpeedT*X* is defined as the average reading speed recorded during the stage numbered as *(S-X)* of speed-reading training

CompT*X* is defined as the average comprehension percentage achieved during the stage numbered as *(S-X)* of speed-reading training

In this regard, SpeedT02 for a student who is at stage 24 corresponds to the average speed recorded during the activities within training stage 22.

The comprehension rates above are measured by assessment through multiple-choice questions after reading. Besides, speed is calculated simply by dividing the total number of words by reading duration. As noted by Dyson and Haselgrove (2000), reading speed and degree of comprehension are used together to calculate reading efficiency.

If W is the number of words, D is the reading duration, and c is the comprehension rate, reading speed will be equal to W/D. According to this notation, reading efficiency is calculated as follows (Guo and Wang, 2018):

$$E = \frac{W}{D}c$$

Below is a subset of training data consisting of 10 records. Each row demonstrates the most recent ten reading and comprehension scores of students, either active or dropped out.

Table 5 10 records in the training dataset constructed with student performance indicators with class labels

Predictor Variables											Target Variable
	Progress		*Reading Speed*				*Reading Comprehension (%)*				
Student ID	**Max Stage**	**Total Activities**	**SpeedT10**	**SpeedT09**	**...**	**SpeedT01**	**CompT10**	**CompT09**	**...**	**CompT01**	**Status**
71	10	26	145.25	148		152.5	70	80		80	1
170	10	21	70	72.5		96	80	80		65	1
176	10	34	115.29	132.5		164.5	60	62.5		95	1
223	10	23	77.33	78.67		89	70	73.33		60	1
0	27	71	678.5	691		2418.25	85	80		75	0
1	27	61	77.5	102		127.5	80	85		80	0
2	27	61	118	128		183.5	76.67	85		90	0
3	27	80	162.5	138		122.5	85	73.33		90	0
4	27	70	265.67	310.5		175.5	66.67	90		90	0
7	27	80	313	347.5		573.5	80	80		80	0

The rightmost column is the class label to be predicted, and others except for the Student ID are predictor variables to be used in classification model.

3.3. Artificial Neural Networks

We utilize Artificial Neural Networks (ANN's) to train a classification model out of training data. ANN's have been designed to work like brain cells, training ANN's help to establish links between inputs and outputs by optimizing neuron-like processors organized in one or multiple layers. Neural network is a set of connected units where inputs are converted to outputs with using weights (Han et al., 2012:398), which includes activation functions that enable non-linear outputs that extend the capability of classifiers. The number of network layers and hidden neurons can be adjusted to develop classifiers with custom neural structures.

3.4. Test Configuration

The initial dataset of 244 rows was split into a training and a test dataset. As usual in many machine learning models, 75% of data was randomly selected for training classification model, while the remaining 25% was held for testing.

Multiple tests were conducted to predict dropouts in students. Before each test, the ratio of dropout students among the training and test dataset was checked to ensure a balance in both datasets.

10 Neural Networks were trained using the following topology:

- Network Layers = 1
- Hidden Neurons = 5

The predictive performance of each test is presented in confusion matrices, which simply compare the predictions with actual values. Precision, recall and accuracy criteria, which were presented in the previous section, were used as the performance indicators for predictive models. The results of the predictions are presented in the next subsection.

4. Findings

In this section, we present the performance of predictive models being trained as a result of ten tests. The predictions by each ANN trained are compared with the actual values in confusion matrices, separately as below:

Table 6 Confusion matrices for 10 tests, combined

Test	Confusion Matrices				Performance Metrics		
			Prediction		Accuracy	Precision	Sensitivity
1			*Regular*	*Quit*			
	Actual	*Regular*	42	4	83.61%	69.23%	60.00%
		Quit	6	9			
2			*Regular*	*Quit*			
	Actual	*Regular*	41	5	78.69%	58.33%	46.67%
		Quit	8	7			
3			*Regular*	*Quit*			
	Actual	*Regular*	38	8	73.77%	46.67%	46.67%
		Quit	8	7			
4			*Regular*	*Quit*			
	Actual	*Regular*	39	8	72.58%	42.86%	40.00%
		Quit	9	6			
5			*Regular*	*Quit*			
	Actual	*Regular*	39	7	77.05%	53.33%	53.33%
		Quit	7	8			
6			*Regular*	*Quit*			
	Actual	*Regular*	38	8	73.77%	46.67%	46.67%
		Quit	8	7			
7			*Regular*	*Quit*			
	Actual	*Regular*	41	5	80.33%	61.54%	53.33%
		Quit	7	8			
8			*Regular*	*Quit*			
	Actual	*Regular*	43	3	81.97%	70.00%	46.67%
		Quit	8	7			
9			*Regular*	*Quit*			
	Actual	*Regular*	40	6	78.69%	57.14%	53.33%
		Quit	7	8			
10			*Regular*	*Quit*			
	Actual	*Regular*	38	8	81.97%	60.00%	80.00%
		Quit	3	12			
				Average:	**78.24%**	**56.58%**	**52.67%**

Based on the test results presented in the findings, the models trained have achieved an average accuracy of 78.24%. Accordingly, 78.24% of predictions on whether a student will quit the course in the following stage were accurate.

On average, 56.68% of students who were predicted to quit in the next stage had been found to quit at the predicted occasion. More importantly, 52.67% of students who would quit the course in the following stage were correctly identified by the classification models.

Although both percentages might be considered low, we notice that the predictions do not only signify a dropout during the remaining of the training program. Instead, the training data and class labels are adjusted so that the model is trained to predict only the students who are likely to quit just after their current stage in their progress within an entire 28-stage training program. As noted by Romero et al. (2008), a good classifier model in EDM needs to be accurate, as well as comprehensible for decision makers and instructors. In this regard, being able to predict more than half of students who would quit in considerably a short while can be considered a good insight for instructors. More importantly, the results demonstrate that this is achieved with acceptable sensitivity, thus without making too many predictions.

5. Conclusion

This chapter presents the application of artificial neural networks for the prediction of students who will quit an online speed-reading course. Data was collected through the assessment of participants attending a speed-reading course, as well as logging progress data on the e-learning system. The dataset constructed for analysis involves reading speeds and comprehension scores achieved by each student in each stage of the course.

The student dropout problem is a widely known problem in *Educational Data Mining* studies. The use of classification models for predicting students who are likely to quit has been regarded as a necessity in various studies.

The results presented in this chapter compare the predictive performance of the models being tested. In a case where the dropout rate is ~25%, we achieved predicting ~53% of dropouts just before they quit. This was possible by involving current and a finite set of most recent observations when preparing the training data. This corresponds to the data preprocessing and transformation activities held before data mining (Fayyad et al., 1996).

In this regard, we note that the ability to identify those that quit in a short while is a valuable capability that helps taking necessary actions, identifying and resolving problems, and thereby preventing students from leaving the course. The use of predictive models and algorithms is a promising practice, especially in retaining students and leading them to continue training programs.

Acknowledgement

The author would kindly present thanks to Mr. Ömer MEŞELİ, founder of *Dijital Nesil Akademisi (https://dijitalnesilakademisi.com)*, for his extensive support for this study.

References

Abdul-Rab, S. D., Abdul-Hamid, S., Romly, R., Toti, U. S., & Mohamed, A. W. A. (2023). Transformational Development of Speed-Reading Technology: Tools, Machines and Software Applications. Theory and Practice in Language Studies, 13(6), 1452-1463.

Abdulahi Hasan, A., & Fang, H. (2021, May). Data Mining in Education: Discussing Knowledge Discovery in Database (KDD) with Cluster Associative Study. In 2021 2nd International Conference on Artificial Intelligence and Information Systems (pp. 1-6).

Akçamete, G., & Güneş, F. (1992). Üniversite Öğrencilerinde Etkili ve Hızlı Okumanın Geliştirilmesi. Ankara University Journal of Faculty of Educational Sciences (JFES), 25(2), 463-471.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of educational data mining, 1(1), 3-17.

Bawden, D., & Robinson, L. (2020, June 30). Information Overload: An Introduction. Oxford Research Encyclopedia of Politics. Retrieved from https://oxfordre.com/politics/ew/10.1093//.001.0001/-e-1360, Access Date: December 5, 2024.

Carver, R. P. (1982). Optimal Rate of Reading Prose. Reading Research Quarterly, 18(1), 56-88.

Dedebali, N. C. (2010). Hızlı okuma tekniğinin sekizinci sınıf öğrencilerinin okuma hızlarına ve okuduğunu anlama düzeylerine etkisi. Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 27(27), 171-183.

De la Varre, C., Irvin, M. J., Jordan, A. W., Hannum, W. H., & Farmer, T. W. (2014). Reasons for student dropout in an online course in a rural K–12 setting. Distance Education, 35(3), 324-344.

Durukan, E. (2020). Impact of speed reading training on reading speeds and comprehension skills of secondary school students. Cypriot Journal of Educational Science. 15(2), 184–193.

Dyson, M., & Haselgrove, M. (2000). The effects of reading speed and reading patterns on the understanding of text read from screen. Journal of research in reading, 23(2), 210-223.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34.

Gale Research Inc. (2017). Data Mining. In the Encyclopedia of Emerging Industries, 7th Edition, ISBN: 978-1410363237.

Goldschmidt, P., & Wang, J. (1999). When can schools affect dropout behavior? A longitudinal multilevel analysis. American Educational Research Journal, 36(4), 715-738.

Grabe, W. (2009). Reading in a second language: Moving from theory to practice. New York, USA: Cambridge University Press.

Guleria, P., & Sood, M. (2014). Data mining in education: A review on the knowledge discovery perspective. International Journal of Data Mining & Knowledge Management Process, 4(5), 47.

Guo, W., & Wang, J. (2018, October). Towards attentive speed reading on small screen wearable devices. In Proceedings of the 20th ACM International Conference on Multimodal Interaction (pp. 278-287).

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and. Techniques, Waltham: Morgan Kaufmann Publishers.

Hilmi, I. F., & Alwi, N. A. (2023). Reading Speed Effectiveness in Improving Literary Culture in the Industrial Revolution 4.0. Modality Journal: International Journal of Linguistics and Literature, 3(1), 38-47.

Kemiksiz, Ö. (2019). Türkçe öğretmeni adaylarının "hızlı okuma" becerisine yönelik metafor algıları. Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi, 7(1), 71-84.

Kinnunen, P., & Malmi, L. (2006, September). Why students drop out CS1 course?. In Proceedings of the second international workshop on Computing education research (pp. 97-108).

Klimovich, M., Tiffin□Richards, S. P., & Richter, T. (2023). Does speed□reading training work, and if so, why? Effects of speed□reading training and metacognitive training on reading speed, comprehension and eye movements. Journal of Research in Reading, 46(2), 123-142.

Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003). Preventing student dropout in distance learning using machine learning techniques. In Knowledge-Based Intelligent Information and Engineering Systems: 7th International Conference, KES 2003, Oxford, UK, September 2003. Proceedings, Part II 7 (pp. 267-274). Springer Berlin Heidelberg.

Kumar, M., Singh, A. J., & Handa, D. (2017). Literature survey on educational dropout prediction. International Journal of Education and Management Engineering, 7(2), 8.

Kurtyılmaz, Y., & Başak, B. E. (2023). Mediating Role of Academic Self-Efficacy between Insufficient Self-Control and School Dropout. International Journal of Contemporary Educational Research, 10(1), 157-170.

Kurulgan, M. (2024). A bibliometric analysis of research on dropout in open and distance learning. Turkish Online Journal of Distance Education, 25(4), 200-229.

Kuzilek, J., Zdrahal, Z., & Fuglik, V. (2021). Student success prediction using student exam behaviour. Future Generation Computer Systems, 125, 661-671.

Lopes, R. A., Rodrigues, L. A., & Brancher, J. D. (2017, June). Predicting master's applicants performance using KDD techniques. In 2017 12th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-6). IEEE.

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. Computers & Education, 53(3), 950-965.

Macalister, J. (2010). Speed reading courses and their effect on reading authentic texts: A preliminary investigation. Reading in a Foreign Language, 22(1), 104-116.

Mills, C., Bosch, N., Graesser, A., & D'Mello, S. (2014). To quit or not to quit: predicting future behavioral disengagement from reading patterns. In Proceedings of Intelligent Tutoring Systems: 12th International Conference (ITS 2014), Honolulu, USA, June 5-9, 2014. Springer International Publishing, 19-28.

Nushi, M. (2017). Spreeder: A web app to develop and enhance reading speed. Reading in a Foreign Language, 29(1), 178-184.

Oyamada, A. (2014). Using computer technology to develop reading speed. TESOL Working Paper Series, 12, 58-71.

Öztahtalı, İ. İ. (2011). Speed-reading Techniques and the Implementability in Turkish. Uludağ Üniversitesi Fen Edebiyat Fakültesi Sosyal Bilimler Dergisi, 12(20), 81-89.

Pan, F., Zhang, H., Li, X., Zhang, M., & Ji, Y. (2024). Achieving optimal tradeoff for student dropout prediction with multi-objective reinforcement learning. PeerJ Computer Science, 10, e2034.

Psaromiligkos, Y., Orfanidou, M., Kytagias, C., & Zafiri, E. (2011). Mining log data for the analysis of learners' behaviour in web-based learning management systems. Operational Research, 11, 187-200.

Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & Treiman, R. (2016). So much to read, so little time: How do we read, and can speed reading help?. Psychological Science in the Public Interest, 17(1), 4-34.

Rodrigues, M. W., Isotani, S., & Zarate, L. E. (2018). Educational Data Mining: A review of evaluation process in the e-learning. Telematics and Informatics, 35(6), 1701-1717.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert systems with applications, 33(1), 135-146.

Romero, C., Ventura, S., Espejo, P. G., & Hervás, C. (2008). Data mining algorithms to classify students. In Proceedings of the 1st International Conference on Educational Data Mining, Montréal, Québec, Canada, June 20-21, 2008.

Sirait, M. F., & Hutauruk, B. S. (2020). The Effect of Using Speed Reading Technique to the Students' Ability in Comprehending a Text. Cetta: Jurnal Ilmu Pendidikan, 3(3), 485-498.

Soysal, T. (2021). Okuma ve anlama bağlamında hızlı okuma teknikleri. Uluslararası Karamanoğlu Mehmetbey Eğitim Araştırmaları Dergisi, 3(2), 129-137.

Soysal, T. (2022). A Mixed Method Study on Improving Reading Speed and Reading Comprehension Levels of Gifted Students. International Journal of Education and Literacy Studies, 10(1), 147-155.

Tsvetkova, M. (2017). The speed reading is in disrepute: Advantages of slow reading for the information equilibrium. European Journal of Contemporary Education, 6(3), 593-603.

Xenos, M., Pierrakeas, C., & Pintelas, P. (2002). A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. Computers & Education, 39(4), 361-377.

Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. Information sciences, 507, 772-794.