DOI: 10.20535/2410-8286.248091

AUTOMATIC CLASSIFICATION OF EFL LEARNERS' SELF-REPORTED TEXT DOCUMENTS ALONG AN AFFECTIVE CONTINUUM

Derya Uysal[1], **Alper Kürşat Uysal**[2] (Corresponding author)

[1]*School of Foreign Languages, Alanya Alaaddin Keykubat University, Alanya, Turkey*
[2]*Department of Computer Engineering, Alanya Alaaddin Keykubat University, Alanya, Turkey*

alper.uysal@alanya.edu.tr

This study aims to place EFL learners along an affective continuum via machine learning methods and present a new dataset about affective characteristics of EFL learners. In line with the purposes, written self-reports of 475 students from 5 different faculties in 3 universities in Turkey were collected and manually assigned by the researchers to one of the labels (positive, negative, or neutral). As a result, two combinations of the same dataset (AC-2 and AC-3) including different numbers of classes were used for the assessment of automatic classification approaches. Results revealed that automatic classification confirmed the manual classification to a great extent and machine learning methods could be used to classify EFL students along an affective continuum according to their affective characteristics. Maximum accuracy rate of automatic classification is 90.06% on AC-2 dataset including two classes. Similarly, on AC-3 dataset including three classes, maximum accuracy rate of classification is 71.79%. Last, the top-10 features/words obtained by feature selection methods are highly discriminative in terms of assessing student feelings for EFL learning. It could be stated that there is not an existing study in which feature selection methods and classifiers are used in the literature to automatically classify EFL learners' feelings.

Keywords: affective factors; EFL learning; text classification; feature selection; EFL students; higher education; affective barriers.

Introduction

There is a notion called the affective domain and positive feelings are driving forces to attain language learners' goal of language acquisition. The idea behind this study is to propose an alternative machine learning-based approach in order to place learners along a continuum based on their affective characteristics as EFL (English as a foreign language) learners instead of qualitative, or quantitative approaches.

The notion of the affective domain was initially developed by Krathwohl et al. (1964) and it has received considerable contributions from researchers and experts in the field of education so far (Sousa, 2016). Affective domain addresses to feelings and emotions, as well as their outward expression. To provide a more operational definition, the affective domain has three subcomponents: behaviour, feeling, and cognition. Feeling is the personal sensation experienced. Cognition is the personal judgements that accompany feelings, and behaviour is the observable reaction which involves both cognition and feelings (Brett, Smith, & Huitt, 2003). Emotions, on the other hand, involve expressive behaviour, bodily reactions and subjective feelings as a result of one's interaction with the context (Cahour, 2013).

Affective variables are significant for learning in general (Schutz & Lanehart, 2002) and for language learning in particular (Méndez López & Peña Aguilar, 2013). Unlike any other subject matters, a level of personal engagement is necessary to learn a foreign language. As recognized by language teachers, learners have to cope with the ambiguities and stress of interaction within the parameters of an unfamiliar culture while conveying conversationally appropriate and personally meaningful messages through unfamiliar phonological, semantic, and syntactic systems. That is why most learners find this process inherently stressful (Horwitz, 1995; Kębłowska, 2012). Language learning is potentially the most threatening experience of all other school subjects since learners are expected to express themselves using a language they are not proficient in (Kębłowska, 2012). It must be acknowledged that while all the cognitive factors may be optimally operating in the learning process, learners can fail because of an affective block (Brown, 1973; Griffith & Nguyen; 2006; Garrett & Young, 2009).

The idea that affect plays a significant role in language learning was first put forward by Gardner and Lambert (1972). After that, Gardner's (1979, 1985) socio-educational model focused on the effect of motivation and attitude on second language learning, and then further studies expanded or revised the original model (Gardner & Clément, 1990; Gardner, 2001). Motivation could be defined as willingness pushing an individual to achieve a predetermined goal (Anderson & Bourke, 2000). Recently, motivation in second language learning has been studied by second-generation researchers (Dörnyei, 1990, 2005, 2015; Dörnyei & Al-Hoorie, 2017; Noels, 2005).

In the socio-educational model, attitude is associated with approaches of language learners to learning situation such as school, language teacher, classmates and so on (Gardner & Lambert, 1972). According to Anderson & Bourke (2000), learning a foreign language entails the attitude of learners towards self, learning context and target culture. Concepts associated with learner's self are self-respect, self-confidence, self-efficacy and self-perception. Recently, attitude in language learning has still been investigated from various aspects such as attitude of language learners toward pronunciation (Huensch & Thompson, 2017) and learner attitudes toward learning English (Zulfikar, Dahliana & Sari, 2019).

Another affective variable widely explored is anxiety. Studies as to language anxiety have been initiated by Horwitz et al. (1986) and the concept was defined as a situation specific worry or nervousness about using a second language. Initial studies have been expanded with further studies demonstrating the correlation between language anxiety and various components of the learning process (Horwitz, 1995, 2000, 2001). Recently, another concept investigated in relation to anxiety is enjoyment (Boudreau et al., 2018; Dewaele, Magdalena & Saito, 2019; De Smet et al. 2018).

Previous research in language learning has either underestimated the feelings or their relevance to learning or studied them as stable and isolated individual variable (Garrett & Young, 2009; Sampson, 2020). Although studies researching affective variables in a holistic manner are few in number, the most influential theory as to second language learning in this regard was proposed by Krashen as "Affective Filter Hypothesis" (1982). According to the hypothesis, positive feelings remain driving forces to attain language learners' goal of language acquisition whereas negative feelings disallow them to perform successfully in learning because negative affective variables disallow information about language to reach language areas of the mind and build barriers into the acquisition of the language. The content, functions and nature of the "affective filter" suggested for second language learning context were tested by Laine (1988) in a foreign language learning context. Results revealed that significant affective filter lowers or raisers were situation related attitudes, personality traits, foreign language self-concept, and target language related attitudes. Studies that aim to explore the affective filter hypothesis exist in the recent body of literature, as well (Nath, Mohamad & Yamat, 2017; Wang, 2020).

Considering the utmost importance of affect in language learning, foreign language teachers need methods for affective assessment and measurement. While assessment means gathering information about affective variables of students, measurement means giving numerals to students according to the degree of their affective attributes (McCoach, Gable & Madura, 2013). Affective characteristics form an important part of the lens through which students perceive and react to the school context. Thus, early identification of affective variables may enable teachers to provide the type of assistance students need (Anderson & Bourke, 2000). Assessment informs teachers about weaknesses, strengths, and abilities of students, which is necessary to improve students' learning outcomes (Setiawan & Mardapi, 2019).

Most researchers or practitioners find affective assessment and measurement problematic for some reasons such as vagueness of affective outcomes, privacy and intangibility of affective characteristics (Anderson & Bourke, 2000). Affective outcomes are not easy to measure or teach because they vary from internally consistent qualities of character to simple attention or selected phenomena (Pierre & Oughton, 2007). Teachers do not have sufficient competence in designing assessment instruments for the affective aspect because it has a complicated construct. To put it in another way, affective domain is hard to measure and define since it is abstract and vague (Setiawan & Mardapi, 2019).

McCoach, Gable & Madura (2013) and Anderson & Bourke (2000) suggest an affective continuum (Figure-1) for affective assessment. According to them, learners could be placed along an affective continuum based on the affective characteristics they develop toward a target. Strength or degree is related to the intensity of the feelings and they might be strong, weak or moderate. Also, the direction of these feelings might be positive, neutral or negative.

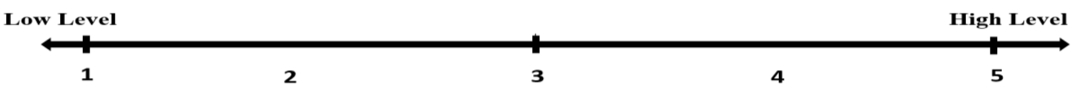

Figure 1. *An Affective Continuum (McCoach, Gable & Madura, 2013, p.36)*

In order to achieve this aim, Anderson & Bourke, 2000; McCoach, Gable & Madura, 2013 propose to gather data about affective characteristics through direct observations or self-reports (preferably self-reports) and follow a number of steps in order to reveal latent constructs.

The initial aim of this study is to propose an alternative method, machine learning, in order to place learners along a continuum based on their affective characteristics as EFL (English as a foreign language) learners. Research aiming to reveal affective characteristics of students is carried out with qualitative (interviews, observations), or quantitative (surveys, questionnaires or scales) approaches (Buissink-Smith, Mann & Shephard, 2011). Although these existing methods are useful for both revealing affective characteristics of students and finding out their relations with various factors, they may require immense time and effort of researchers and practitioners. Unlike these methods, this study aims to present a practical and time-efficient method to define the place of students along an affective continuum and categorize them into two (positive or negative) or three (positive, neutral or negative) groups along this continuum via machine learning. Furthermore, this study displays how the objectivity of affective assessment is increased by means of computerized methods.

The affective measurement suggested here could be used for various purposes. For example, a teacher may have a desire to find out the affective stance of his/her learners on EFL learning for designing effective learning courses or for predicting possible affective domain-related difficulties or challenges for the entire year. Defining affective characteristics at the outset of an EFL course or program is highly significant in terms of detecting students' negative feelings and taking precautions to replace these negative feelings with positive ones.

Moreover, the study aims to present a new dataset about affective characteristics of EFL learners. Thus, researchers may conduct further studies by using computerized methods. The study could be evaluated as a first step of analyzing affective variables through computerized methods. In the next subsections, information about the dataset and automatic classification scheme based on machine learning is given in addition to the results of the study including accuracy and feature set analysis.

Methods

Creating the dataset

The first step of the study is creating the dataset about affective characteristics of EFL students. In this step, self-reporting technique was used to gather detailed information as to the target topic. Participants are required to assess and reflect on their own internal traits or states through self-report instruments which are effective in inferring a person's level on the affective characteristic of interest (McCoach, Gable & Madura, 2013; Anderson & Bourke, 2000; Buissink-Smith, Mann & Shephard, 2011). Initially, a question was prepared to gather information about feelings of students for EFL learning. The question was as follows: "please write a paragraph that describes your feelings for English learning considering your past and present English courses". As the question and answers were in Turkish language, the dataset created is in Turkish.

The question was delivered to EFL students from 5 different faculties (engineering, tourism, business and management, and education) in 3 state universities in Turkey via online and hardcopy forms during two consecutive academic years (2019-2020 and 2020-2021). English is the main medium of instruction in engineering departments and ELT department in education faculty while Turkish is the medium of instruction in the faculties of tourism, business and management, and other departments in faculty of education. At the end of this period, total 475 paragraphs about the affective characteristics of Turkish EFL students were obtained. While 200 students answered the question in hardcopy forms, 275 students answered it in online forms.

After gathering the paragraphs, they were transferred to text files. Then, all of the 475 paragraphs were read through by the researchers and manually assigned the following labels: 1:positive, 2:negative, and 3:neutral. The paragraphs assigned to label-1 entails totally or mostly positive feelings toward learning EFL; label-2 entails totally or mostly negative feelings and label-3 entails both of them or the ones expressing no feeling. The dataset could be reached from https://drive.google.com/file/d/1dgUBivSYQRTQx43NS2ldf3uPYXDh0d0o/view?usp=sharing. As a result of the dataset creation step, two combinations of the same dataset including different numbers of classes were used for the assessment of automatic classification approaches. While the first combination consists of positive and negative classes, the second combination of the dataset consists of positive, negative, and neutral classes. The detailed information concerning the dataset is presented in Table 1 below. These two combinations of the dataset will be referred as AC-2 and AC-3 in the next parts of the study.

Table 1. *Affective continuum (AC) dataset*

Class Label	Number of Samples
positive	164
negative	153
neutral	158

Automatic classification scheme used to classify students along the continuum

In this study, machine learning approaches including feature selection and classification stages were used. To obtain a successful automatic classification scheme, it is crucial to apply efficient dimension reduction approaches such as feature selection methods (FSMs). Feature selection is a significant step for tasks including classification of text documents because it allows the selection of relevant features/terms and removal of the irrelevant ones. After selecting the discriminative features, it is necessary to apply appropriate classification algorithms for tasks including the classification of text documents. FSMs and classifiers employed in this study are presented in the next subsections.

Feature selection methods (FSM)

Filter-based FSMs are widely used for the purpose of classifying texts as they are advantageous in terms of computation time (Parlak & Uysal, 2020). Five well-known filter-based FSMs, distinguishing feature selector (DFS), Gini index (GI), Information Gain (IG), discriminative features selection (DFSS), and chi-square (CHI2), used in the experiments are explained below.

Gini index (GI), a global FSM, produces a single score for each feature in the training set (Shang et al., 2007). The GI formula is given below.

$$GI(t) = \sum_{i=1}^{M} P(t|C_i)^2 P(C_i|t)^2 \quad (1)$$

The GI formula entails two conditional probabilities. In this formula, $P(t|C_i)$ means the probability of feature t in class C_i and $P(C_i|t)$ means the probability of class C_i when feature t presents. Also, M is the number of classes in the dataset.

One of the well-known filter-based FSMs for classifying texts is Distinguishing feature selector (DFS) (Uysal & Gunal, 2012). It relies on some pre-defined criteria about discriminative characteristics of features. The DFS formula is as following.

$$DFS(t) = \sum_{i=1}^{M} \frac{P(C_i|t)}{P(\bar{t}|C_i) + P(t|\bar{C}_i) + 1} \quad (2)$$

DFS formula entails some conditional probabilities. The probability $P(C_i|t)$ means the probability of class C_i when feature t presents and M is the number of classes in the dataset. $P(\bar{t}|C_i)$ is the probability of lack of feature t given class C_i. $P(t|\bar{C}_i)$ is the probability of feature t when all other classes except C_i present.

Information gain (IG) is utilized to measure the impact of the presence or absence of a feature on the correct classification decision (Deng et al., 2019). Obtaining a high score for a feature with IG method means that the feature is highly discriminative for classification. IG formula is given below.

$$\begin{aligned} IG(t) = -\sum_{i=1}^{M} P(C_i)\log P(C_i) + P(t)\sum_{i=1}^{M} P(C_i|t)\log P(C_i|t) \\ + P(\bar{t})\sum_{i=1}^{M} P(C_i|\bar{t})\log P(C_i|\bar{t}), \end{aligned} \quad (3)$$

In the IG formula, M means the total number of classes and $\boldsymbol{P(C_i)}$ is the probability of class C_i in the training set. While $\boldsymbol{P(t)}$ refers to the probability of existence of feature t, $\boldsymbol{P(\bar{t})}$ is the probability of the absence of the feature t in the training set. Also, the meaning of the statements $\boldsymbol{P(C_i|t)}$ and $\boldsymbol{P(C_i|\bar{t})}$ is the probability of the existence of class C_i for two separate cases where feature t is present or absent, respectively.

Discriminative features selection (DFSS) is another filter-based FSM utilized for classifying text documents. DFSS aims to select features with a higher document and average term frequency in documents belonging to a certain class (Zong et al., 2015). The DFSS formula is presented as following.

$$DFSS(t,C) = \frac{tf(t,C)/df(t,C)}{tf(t,\bar{C})/df(t,\bar{C})} \times \frac{a}{(a+b)} \times \frac{a_i}{(a+c)} \times \left|\frac{a}{(a+b)} - \frac{c}{(c+d)}\right| \quad (4)$$

In the formula, $tf(t, C)$and $tf(t, \bar{C})$ refer to the frequencies of feature t in class C and in the other classes, respectively. While $df(t, C)$ refers to the number of text documents feature t occurs in class C, $df(t, \bar{C})$ refers to the number of text documents feature t occurs in other classes. While a refers to the number of text documents in class C including feature t, b refers to the number of text documents in class C not including feature t. While c refers to the number of text documents in all classes except class C that includes feature t, d is the number of text documents in all classes except class C not including feature t.

Chi-square (CHI2) FSM is based on a well-known statistical test to investigate how far two events are independent of each other (Deng et al., 2019). The CHI2 formula is presented below.

$$CHI2(t, C) = \sum_{t \in \{0,1\}} \sum_{C \in \{0,1\}} \frac{(N_{t,C} - E_{t,C})^2}{E_{t,C}}, \quad (5)$$

In the formula of CHI2, E means the expected frequency and N means the observed frequency for each state of class C and feature t.

Classifiers

Filter-based FSMs are utilized in the experiments. These kinds of dimension reduction methods do not depend on the learning model. For this reason, two different well-known classifiers were used in order to reveal how much selected features contribute to the accuracy of classification. These two classifiers are Support Vector Machines (Joachims, 1998) and Naïve Bayes (Chen et al., 2009). It should be noted that these kinds of classifiers were used in some previous studies related to education and computation (Liu et al., 2021; Pei & Xing, 2021). These methods are briefly explained in the next subsections.

Support Vector Machines (SVM) classifier is a successful and well-known classifier for classifying text documents. Margin maximization concept (Joachims, 1998) is the main theory that lies behind SVM classifier. The margin of the SVM classifier is the distance between the closest data point in the training set and the decision surface. In this study, linear version of SVM classifier, one of the most popular classifiers for tasks including classification of text documents, was used in the experiments.

Naïve Bayes (NB) classifier is based on Bayes theorem that assumes the features are independent from each other. Therefore, a score of probability is found by multiplying the conditional probabilities in a classification algorithm based on the independence assumption (Jiang et al., 2013). NB classifiers can be implemented using various event models such as multi-variate Bernoulli and Multinomial. Multi-variate Bernoulli event model considers document frequencies; however, multinomial event model takes term frequencies into account, instead of during the calculations (Uysal, 2016). In the present study, the former model, multi-variate Bernoulli, is used while implementing NB classifier.

Results

This section presents the results of the in-depth investigation done with the purpose of evaluating the performance of FSMs and classifiers on the two combinations of the dataset including a different number of classes. Experiments were carried out for two cases including no stemming or stemming. Zemberek (Akın & Akın, 2007), one of the common frameworks for Turkish stemming, was used in the experiments. It should also be asserted that lowercase conversion was used as the pre-processing step besides weighting terms with term frequency-inverse document frequency. The results were obtained using 10-fold cross validation to measure the datasets objectively. Besides, an accuracy measure is adopted so as to evaluate the performance of classification as the class distribution of the datasets is nearly balanced.

Accuracy analysis

Different numbers of the features selected by five diverse selectors were given as input to two classifiers that are SVM and NB. Dimension reduction was performed via constructing feature sets with 10, 50, 100, 300, 500, 800, and 1000 features. Accuracies obtained in the experiments with SVM classifier are listed in Table 2 for the AC-2 dataset. In the tables, the highest scores are indicated in bold.

Table 2. *Accuracies for the AC-2 dataset obtained with SVM classifier*

	Accuracy (Zemberek stemming)							Accuracy (No stemming)						
Feature Size	10	50	100	300	500	800	1000	10	50	100	300	500	800	1000
DFS	69.33	85.64	88.34	85.03	87.12	85.74	86.68	73.33	80.64	84.39	82.68	85.02	84.00	85.71
GI	57.66	79.69	85.73	87.01	**90.06**	88.08	88.06	53.76	79.99	78.71	83.63	82.63	84.96	85.63
IG	70.35	83.28	84.37	86.06	87.07	85.76	85.69	75.34	82.33	82.05	84.68	85.00	85.36	**87.03**
CHI2	68.36	85.30	83.71	88.37	87.73	88.72	85.69	75.67	80.71	82.42	84.99	85.65	84.70	85.72
DFSS	73.31	79.64	85.30	86.68	85.75	88.03	89.07	66.39	79.69	79.65	80.03	78.32	80.35	76.67

For SVM classifier, the highest accuracy was obtained with 500 features using GI FSM and Zemberek stemming on AC-2 dataset. While the highest accuracy was 90.06% with Zemberek stemming, the highest accuracy with no stemming was 87.03% using SVM classifier on AC-2 dataset. It should be noted that the highest accuracy using no stemming was obtained with IG FSM and SVM classifier on AC-2 dataset. Accuracies obtained in the experiments with NB classifier are listed in Table 3 for the AC-2 dataset where the highest scores are indicated in bold.

Table 3. *Accuracies for the AC-2 dataset obtained with NB classifier*

	Accuracy (Zemberek stemming)							Accuracy (No stemming)						
Feature Size	10	50	100	300	500	800	1000	10	50	100	300	500	800	1000
DFS	70.37	82.35	83.35	**84.32**	81.00	80.97	80.99	74.34	**86.67**	84.71	82.99	83.32	79.68	76.66
GI	59.98	75.05	80.36	82.04	81.35	79.64	80.64	55.05	80.35	79.70	83.69	84.37	82.70	80.67
IG	72.69	80.33	81.70	82.68	81.32	79.32	80.99	76.02	83.68	82.69	82.65	81.38	81.37	77.68
CHI2	70.33	81.32	82.04	82.33	81.64	80.31	80.99	77.02	82.38	84.38	83.98	81.34	81.69	77.68
DFSS	69.75	79.02	80.66	79.36	79.03	79.99	80.96	68.70	78.04	78.25	79.00	79.00	77.32	77.36

For NB classifier, the highest accuracy was obtained with 50 features using DFS FSM and no stemming on AC-2 dataset. While the highest accuracy was 86.67% with no stemming, the highest accuracy with Zemberek stemming was 84.32% using NB classifier on AC-2 dataset. It should be noted that the highest accuracies were obtained with DFS FSM using both no stemming and Zemberek stemming with NB classifier on AC-2 dataset. Accuracies obtained in the experiments with SVM classifier are listed in Table 4 for the AC-3 dataset where the highest scores are indicated in bold.

Table 4. *Accuracies for the AC-3 dataset obtained with SVM classifier*

	Accuracy (Zemberek stemming)							Accuracy (No stemming)						
Feature Size	10	50	100	300	500	800	1000	10	50	100	300	500	800	1000
DFS	54.48	62.92	66.26	64.91	69.34	69.58	68.92	55.36	61.78	63.84	61.08	62.43	64.66	64.23
GI	40.68	58.23	67.40	67.79	68.46	70.26	70.69	42.26	60.92	62.73	64.25	63.10	**67.38**	65.34
IG	52.24	63.56	65.37	65.98	67.13	70.27	70.21	55.11	61.97	63.81	64.25	63.13	64.92	65.38
CHI2	56.04	63.15	64.69	63.56	70.26	69.78	71.12	55.35	62.88	64.24	62.41	59.56	66.04	64.48
DFSS	47.14	59.56	66.50	67.78	70.03	71.14	**71.79**	46.68	60.68	61.17	60.47	62.04	62.05	63.36

For SVM classifier, the highest accuracy was obtained with 1000 features using DFSS FSM and Zemberek stemming on AC-3 dataset. While the highest accuracy was 71.79% with Zemberek stemming, the highest accuracy with no stemming was 67.38% using SVM classifier on AC-3 dataset. It should be noted that the highest accuracy using no stemming was obtained with GI FSM and SVM classifier on AC-3 dataset. Accuracies obtained in the experiments with NB classifier are listed in Table 5 for the AC-3 dataset where the highest scores are indicated in bold.

Table 5. *Accuracies for the AC-3 dataset obtained with NB classifier*

	Accuracy (Zemberek stemming)							Accuracy (No stemming)						
Feature Size	10	50	100	300	500	800	1000	10	50	100	300	500	800	1000
DFS	53.80	59.12	**63.83**	59.11	57.77	54.21	55.14	54.47	60.93	62.06	**64.05**	64.04	59.39	54.96
GI	40.53	53.13	58.25	59.57	58.49	56.67	54.25	38.76	57.15	60.31	63.64	63.18	61.62	58.28
IG	52.25	60.91	60.04	59.56	57.80	55.80	53.80	53.77	60.72	62.27	63.60	59.37	59.84	57.85
CHI2	54.26	59.36	59.39	59.33	58.46	54.67	56.25	55.14	61.58	63.60	62.27	59.81	59.15	56.93
DFSS	46.24	55.79	58.25	58.25	58.03	57.59	56.26	42.74	58.71	59.86	63.17	61.86	61.86	60.08

For NB classifier, the highest accuracy was obtained with 300 features using DFS selection method and no stemming on AC-3 dataset. While the highest accuracy was 64.05% with no stemming, the highest accuracy with Zemberek stemming was 63.83% using NB classifier on AC-3 dataset. It should be noted that the highest accuracies using both no stemming and Zemberek stemming were obtained with DFS selection method with NB classifier on AC-3 dataset.

The highest accuracies obtained on AC-2 and AC-3 datasets were 90.06% and 71.79%, respectively. So, the performance of classification degrades after including the class ‘neutral’ into the dataset. While the highest classification performance was obtained with the combination of GI selection method, Zemberek stemming, and SVM classifier on AC-2 dataset, the highest classification performance was obtained with the combination of DFSS feature selection, Zemberek stemming, and SVM classifier on AC-3 dataset. Although the highest classification performances were obtained with GI and DFSS FSM on two datasets, it should be noted that DFS was the best performer for 4 out of 8 cases in the experiments. While the highest accuracies obtained using Zemberek stemming were always better than the results obtained with no stemming on two datasets for SVM classifier, the highest accuracies obtained using no stemming were always better than the results obtained with Zemberek stemming on two datasets for NB classifier.

Feature set analysis

Top-10 features for AC-2 and AC-3 datasets obtained are presented in Tables 6–9. The translation of the features to English was also shown in the brackets. The tables given also illustrate feature sets regarding five diverse FSMs. The features specific to an individual FSM are written in bold, in the tables. It should be noted that there are not so many features selected by only one feature selection according to these tables. Also, the top-10 features selected by corresponding FSMs using the same stemming setting differs for AC-2 and AC-3 datasets.

Table 6. *Top-10 features for the AC-2 dataset using Zemberek stemming*

No	1	2	3	4	5	6	7	8	9	10
DFS	hiç (never)	eğitim (education)	sev (like)	kötü (imperfect)	öğrenci (student)	yanlış (inefficient)	güzel (good)	biz (we)	okul (school)	gör (see)
GI	**ingilizce (english)**	ol (be)	**bir (one)**	**ve (and)**	ders (course)	**öğren (learn)**	bu (this)	dil (language)	**için (for)**	**çok (many)**
IG	hiç (never)	eğitim (education)	sev (like)	kötü (imperfect)	yanlış (inefficient)	öğrenci (student)	güzel (good)	biz (we)	okul (school)	gör (see)
CHI2	hiç (never)	eğitim (education)	sev (like)	kötü (imperfect)	öğrenci (student)	gör (see)	yanlış (inefficient)	biz (we)	okul (school)	güzel (good)
DFSS	eğitim (education)	ders (course)	sev (like)	dil (language)	hiç (none)	bu (this)	gör (see)	**konuş (speak)**	ol (be)	okul (school)

Table 7. *Top-10 features for the AC-2 dataset using no stemming*

No	1	2	3	4	5	6	7	8	9	10
DFS	seviyorum (I like)	hiç (never)	eğitim (education)	konuşma (speaking)	**yetersiz (insufficient)**	**güzel (good)**	kötü (imperfect)	bize (to us)	sınıftan (since the grade)	yanlış (inefficient)
GI	**ingilizce (english)**	**bir (one)**	ve (and)	**için (for)**	**çok (many)**	bu (this)	dil (language)	seviyorum (I like)	de (as well)	hiç (never)
IG	hiç (never)	eğitim (education)	konuşma (speaking)	kötü (imperfect)	bize (to us)	sınıftan (since grade)	yanlış (inefficient)	rağmen (although)	seviyorum (I like)	**eğlenceli (joy)**
CHI2	hiç (never)	eğitim (education)	konuşma (speaking)	seviyorum (I like)	kötü (imperfect)	bize (to us)	sınıftan (since the grade)	yanlış (inefficient)	rağmen (although)	**zor (difficult)**
DFSS	hiç (never)	ve (and)	eğitim (education)	de (as well)	dil (language)	bu (this)	**öğrenmek (learn)**	**ama (yet)**	konuşma (speaking)	**dersi (course)**

Table 8. *Top-10 features for the AC-3 dataset using Zemberek stemming*

No	1	2	3	4	5	6	7	8	9	10
DFS	hiç (never)	ama (yet)	kötü (imperfect)	eğitim (education)	**zevk (enjoyment)**	sev (like)	fakat (but)	yap (do)	ver (give)	**yanlış (inefficient)**
GI	**ingilizce (english)**	ol (be)	**bir (one)**	ders (course)	**ve (and)**	öğren (learn)	**bu (this)**	çok (many)	**için (for)**	ama (yet)
IG	hiç (never)	ama (yet)	eğitim (education)	sev (like)	kötü (imperfect)	**güzel (nice)**	fakat (but)	öğrenci (student)	yap (do)	ver (give)
CHI2	ama (yet)	hiç (never)	eğitim (education)	kötü (imperfect)	sev (like)	fakat (but)	yap (do)	ver (give)	ders (course)	öğrenci (student)
DFSS	ama (yet)	ol (be)	ders (course)	**dil (language)**	eğitim (education)	öğren (learn)	yap (do)	sev (like)	çok (many)	hiç (never)

Table 9. *Top-10 features for the AC-3 dataset using no stemming*

No	1	2	3	4	5	6	7	8	9	10
DFS	seviyorum (I like)	hiç (never)	yetersiz (insufficient)	ama (yet)	kötü (imperfect)	fakat (but)	yanlış (ineffective)	sınıftan (since grade)	eğitim (education)	**zevkli (joyful)**
GI	**ingilizce (english)**	bir (one)	ve (and)	çok (many)	için (for)	ama (yet)	**bu (this)**	dil (language)	**de (as well)**	daha (more)
IG	hiç (never)	ama (yet)	eğitim (education)	fakat (but)	konuşma (speaking)	kötü (imperfect)	yanlış (inefficient)	sınıftan (since grade)	**bize (to us)**	seviyorum (I like)
CHI2	ama (yet)	hiç (never)	seviyorum (I like)	fakat (but)	kötü (imperfect)	eğitim (education)	konuşma (speaking)	sınıftan (since grade)	yanlış (inefficient)	yetersiz (insufficient)
DFSS	ama (yet)	hiç (never)	seviyorum (I like)	çok (many)	bir (one)	dil (language)	ve (and)	için (for)	daha (more)	**iyi (well)**

The top 10 features that were obtained as a result of feature set analysis are highly discriminative in terms of assessing student feelings toward EFL learning. For instance, the top features associated with positive feelings are güzel=nice, iyi=good/well/nice well, zevk=joy, eğlenceli=joyful, seviyorum=I like. These features show that students throughout the right end (positive end) of the affective continuum frequently describe EFL learning process with these words. Regarding the features associated with negative feelings, the most frequent words are kötü=imperfect, yanlış=inefficient, zor=difficult, yetersiz=insufficient and the students placed throughout the left end (negative end) of the affective continuum use these words to describe their feelings for EFL learning. Although, some features among the top 10 seem neutral, when the dataset is examined, these features may be associated with the negative end of the continuum. For example, the features “hiç=never, sınıftan=since grade, eğitim=education, konuşma=speaking, okul=school, and ders=course” are mostly used by the students to express that “they have *never* learned English language because of imperfect/insufficient/inefficient *education* system/teachers/courses although they have been learning it *since* the second *grade* of primary *school”* and “they have never improved their *speaking* skills as teachers don’t prefer *speaking* activities, so they want more *speaking*-centred English *courses*”. The linkers expressing contrast like ama=yet or fakat=but and the linkers expressing a parallel idea like de=as well, or ve=and show neither negative nor positive feelings of students. Likewise, some verb stems like sev=like, ol=be/become, yap=do, öğren=learn were assessed neutral as they do not express any feelings either negative or positive.

Discussion

This study has two main purposes. The initial purpose of the present study is to create a new Turkish dataset about the affective characteristics of EFL learners. The second purpose is to place learners along a continuum based on their affective characteristics as EFL learners via machine learning. In line with these purposes, written self-reports of 475 students from 5 different faculties (engineering, tourism, business and management and education) in 3 state universities in Turkey were collected and then manually assigned by the researchers one of three labels: 1:positive, 2:negative, and 3:neutral.

Considering the manual classification of learners' self-reports, these 3 groups of students could be placed along an affective continuum as illustrated in figure 2.

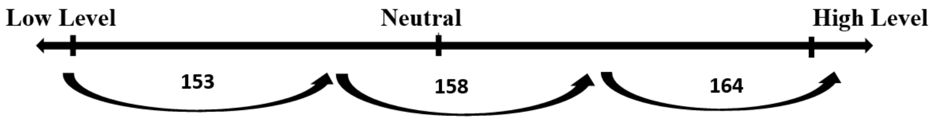

Figure 2. *Classification of learners' self-reports along an affective continuum*

Regarding the automatic classification of EFL learners' self-reported text documents, the accuracy rate of classification via five different FSMs and two different classifiers changes from 90.06% to 84.32% on AC-2 dataset including two classes that are positive and negative. Similarly, on AC-3 dataset including three classes that are positive, negative, and neutral, the accuracy rate of classification via five different FSMs and two different classifiers changes from 71.79% to 63.83%. The results of the FSMs show that automatic classification confirms the manual classification to a great extent and machine learning methods could be used to classify students along an affective continuum according to their affective characteristics as EFL learners. While the accuracy rate is quite high for the two-class dataset, it is slightly lower for the three-class dataset. Moreover, feature set analyses for both datasets have proved that the most frequent top-10 words are highly discriminative in terms of assessing student feelings toward EFL learning. Thus, machine learning methods are appropriate for discriminating expressions students use to express their feelings for EFL learning.

Considering the significance of affective characteristics in EFL learning, methods helping practitioners or researchers detect affective features of students will contribute both to the related body of literature and in-class practices. This study could be assessed as the first step of using computerized methods in defining affective factors and needs to be elaborated with further studies. Also, it presents a time-and-effort-efficient method for practitioners to detect the affective characteristics of their students. Another point is that as far as we know, there isn't an existing study in which FSMs and classifiers are used in order to automatically classify text documents including EFL learners' feelings. Hence, we could state that the present study is unique and is expected to fill a gap regarding data science applications in EFL learning.

Regarding the future work, similar investigations could be carried out for different languages and language learning settings, e.g. ESL (English as a Second Language) or ESP (English for Specific Purposes). In particular, feelings of students learning Turkish as a second language in TOMERs (Center for Teaching Turkish Language) in state universities in Turkey seem worthy of concern. Moreover, similar studies could be carried out for individual affective characteristics such as attitude, motivation, or anxiety. For instance, EFL students could be placed along a continuum according to their level of positive or negative attitudes toward EFL learning process. Last, as a follow-up study, in addition to placing students along the continuum in groups, studies investigating placement of individual students along a continuum could be conducted.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest statement

The authors declare no conflicts of interest.

References:

Akın, A. A., & Akın, M. D. (2007). Zemberek, an open source nlp framework for Turkic languages. *Structure*, *10*, 1-5. https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.556.69

Anderson, L. W., & Bourke, S. F. (2000). *Assessing affective characteristics in the schools*. Routledge.

Boudreau, C., MacIntyre, P., & Dewaele, J. M. (2018). Enjoyment and anxiety in second language communication: An idiodynamic approach. *Studies in Second Language Learning and Teaching*, *8*(1), 149-170. https://doi.org/10.14746/ssllt.2018.8.1.7

Brett, A., Smith, M., & Huitt, W. (2003*)*. Overview of the affective domain. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. http://www.edpsycinteractive.org/brilstar/chapters/affectdev.pdf

Brown, H. D. (1973). Affective variables in second language acquisition. *Language learning*, *23*(2), 231-244. https://doi.org/10.1111/j.1467-1770.1973.tb00658.x

Buissink-Smith, N., Mann, S., & Shephard, K. (2011). How do we measure affective learning in higher education?. *Journal of Education for Sustainable Development*, *5*(1), 101-114. https://doi.org/10.1177/097340821000500113

Cahour, B. (2013). Characteristics, emergence and circulation in interactional learning. In M. Baker, J. Andriessen and S. Jarvela (eds), *Affective learning together* (pp.52-70). London: Routledge. https://doi.org/10.4324/9780203069684

Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications*, *36*(3), 5432-5435. https://doi.org/10.1016/j.eswa.2008.06.054

De Smet, A., Mettewie, L., Galand, B., Hiligsmann, P., & Van Mensel, L. (2018). Classroom anxiety and enjoyment in CLIL and non-CLIL: Does the target language matter?. *Studies in Second Language Learning and Teaching*, *8*(1), 47-71. https://doi.org/10.14746/ssllt.2018.8.1.3

Deng, X., Li, Y., Weng, J., & Zhang, J. (2019). Feature selection for text classification: A review. *Multimedia Tools and Applications*, *78*(3), 3797-3816. https://doi.org/10.1007/s11042-018-6083-5

Dewaele, J. M., Magdalena, A. F., & Saito, K. (2019). The effect of perception of teacher characteristics on Spanish EFL learners' anxiety and enjoyment. *The Modern Language Journal*, *103*(2), 412-427. https://doi.org/10.1111/modl.12555

Dörnyei, Z. (1990). Conceptualizing motivation in foreign language learning. *Language Learning*, *40,* 46-78. https://doi.org/10.1111/j.1467-1770.1990.tb00954.x

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* London: Erlbaum.

Dörnyei, Z., Ryan, S. (2015). *The psychology of the language learner revisited*. New York, NY: Routledge.

Dörnyei, Z., & Al-Hoorie, A. H. (2017). The motivational foundation of learning languages other than global English: Theoretical issues and research directions. *Modern Language Journal, 101*(3), 455-468. https://doi.org/10.1111/modl.12408

Gardner, R.C. (1979). Social-psychological aspects of second language acquisition. In: Giles, H. & St. Clair, R. (eds.), *Language and Social Psychology* (pp.193–220). Oxford: Blackwell.

Gardner, R. C. (1985). *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. London: Edward Arnold.

Gardner, R. C. (2001). Integrative motivation and second language acquisition. *Motivation and second language acquisition, 23*(1), 1-19.

Gardner, R. C., & Clément, R. (1990). *Social psychological perspectives on second language acquisition*. John Wiley & Sons.

Gardner, R.C. & Lambert, W.E. (1972). Motivational variables in second language acquisition. In R.C. Gardner & W. Lambert (eds.) *Attitudes and motivation in second language learning*. (pp. 119-216). Rowley, MA: Newbury House.

Garrett, P., & Young, R. F. (2009). Theorizing affect in foreign language learning: An analysis of one learner's responses to a communicative Portuguese course. *The Modern Language Journal*, *93*(2), 209-226. https://doi.org/10.1111/j.1540-4781.2009.00857.x

Griffith, K. G., & Nguyen, A. D. (2006). Are educators prepared to affect the affective domain. *National forum of teacher education journal. 16* (3), 1-4.

Horwitz, E. K. (1995). Student affective reactions and the teaching and learning of foreign languages. *International Journal of Educational Research*, *23*(7), 573-579. https://doi.org/10.1016/0883-0355(96)80437-x

Horwitz, E. K. (2000). It Ain't over'til It's Over: On Foreign Language Anxiety, First Language Deficits, and the Confounding of Variables. *Modern Language Journal*, *84*(2), 256-259. https://doi.org/10.1111/0026-7902.00067

Horwitz, E. K. (2001). Language anxiety and achievement. *Annual review of applied linguistics*, *21*, 112-126. https://doi.org/10.1017/s0267190501000071

Horwitz, E. K., Horwitz, M. B., ve Cope, J. (1986). Foreign language classroom anxiety. *The modern language journal*, *70*(2), 125-132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Huensch, A., & Thompson, A. S. (2017). Contextualizing attitudes toward pronunciation: Foreign language learners in the United States. *Foreign language annals*, *50*(2), 410-432. https://doi.org/10.1111/flan.12259

Jiang, L., Cai, Z., Zhang, H., & Wang, D. (2013). Naive Bayes text classifiers: a locally weighted learning approach. *Journal of Experimental & Theoretical Artificial Intelligence*, *25*(2), 273-286. https://doi.org/10.1080/0952813x.2012.721010

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137-142). Springer, Berlin, Heidelberg.

Kębłowska, M. (2012). The place of affect in second language acquisition. In *New perspectives on individual differences in language learning and teaching* (pp. 157-167). Springer, Berlin, Heidelberg.

Krashen, S. (1982). *Principles and practice in second language acquisition.* ISO 690

Krathwohl, D. R., Bloom, B. S., and Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook II: Affective domain*, Allyn and Bacon, Boston, Mass.

Laine, E. J. (1988). Report on The Affective Filter in Foreign Language Learning and Teaching. Report 2: A Validation Study of Filtering Factors with a Focus on the Learner's FL Self-Concept. Jyvaskyla Cross-Language Studies, No. 15.

Liu, B., Xing, W., Zeng, Y., & Wu, Y. (2021). Quantifying the Influence of Achievement Emotions for Student Learning in MOOCs. *Journal of Educational Computing Research*, 59(3), 429-452. https://doi.org/10.1177/0735633120967318

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain*. New York, NY: Springer.

Méndez López, M. G., & Peña Aguilar, A. (2013). Emotions as learning enhancers of foreign language learning motivation. *Profile Issues in Teachers Professional Development*, *15*(1), 109-124. https://files.eric.ed.gov/fulltext/EJ1053728.pdf

Nath, P. R., Mohamad, M., & Yamat, H. (2017). The effects of movies on the affective filter and English acquisition of low-achieving English learners. *Creative Education*, *8*(08), 1357. https://doi.org/10.4236/ce.2017.88096

Noels, K. A. (2005). Orientations to learning German: Heritage language background and motivational processes. *Canadian Modern Language Review, 62*, 285–312. https://doi.org/10.3138/cmlr.62.2.285

Parlak, B., & Uysal, A. K. (2020). On classification of abstracts obtained from medical journals. *Journal of Information Science*, *46*(5), 648-663. https://doi.org/10.1177/0165551519860982

Pei, B., & Xing, W. (2021). *An Interpretable Pipeline for Identifying At-Risk Students*. Journal of Educational Computing Research, https://doi.org/10.1177/07356331211038168

Pierre, E., & Oughton, J. (2007). The Affective Domain: Undiscovered Country. *College Quarterly*, *10*(4), 1-7. https://eric.ed.gov/?id=EJ813766

Sampson, R. J. (2020). Interacting Levels and Timescales in the Emergence of Feelings in the L2 Classroom.In Richard J. Sampson and Richard S. Pinner (Eds.), *Complexity Perspectives on Researching Language Learner and Teacher Psychology* (pp.35-51). Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781788923569-004

Schutz, P. A., & Lanehart, S. L. (2002). Emotions in education. *Educational Psychologist*, *37*(2), 67-68. https://doi.org/10.1207/s15326985ep3702_1

Setiawan, A., & Mardapi, D. (2019). The Development of Instrument for Assessing Students' Affective Domain Using Self-and Peer-Assessment Models. *International Journal of Instruction*, *12*(3), 425-438. https://doi.org/10.29333/iji.2019.12326a

Shang, W., Huang, H., Zhu, H., Lin, Y., Qu, Y., & Wang, Z. (2007). A novel feature selection algorithm for text categorization. *Expert Systems with Applications*, *33*(1), 1-5. https://doi.org/10.1016/j.eswa.2006.04.001

Sousa, D. A (2016). *How the Brain Works.* Crowin Press.

Uysal, A. K. (2016). An improved global feature selection scheme for text classification. *Expert systems with Applications*, *43*, 82-92. https://doi.org/10.1016/j.eswa.2015.08.050

Uysal, A. K., & Gunal, S. (2012). A novel probabilistic feature selection method for text classification. *Knowledge-Based Systems*, *36*, 226-235. https://doi.org/10.1016/j.knosys.2012.06.005

Wang, L. (2020). Application of Affective Filter Hypothesis in Junior English Vocabulary Teaching. *Journal of Language Teaching and Research*, *11*(6), 983-987. https://doi.org/10.17507/jltr.1106.16

Zong, W., Wu, F., Chu, L. K., & Sculli, D. (2015). A discriminative and semantic feature selection method for text categorization. *International Journal of Production Economics*, *165*, 215-222. https://doi.org/10.1016/j.ijpe.2014.12.035

Zulfikar, T., Dahliana, S., & Sari, R. A. (2019). An Exploration of English Students' Attitude toward Learning English. *English Language Teaching Educational Journal*, *2*(1), 1-12. https://doi.org/10.12928/eltej.v2i1.947

Received: 20 December, 2021
Accepted: 01 February, 2022