Journal of Psychological Science
Print ISSN: 1735-7462 Online ISSN: 2676-6639
Homepage: http://www.psychologicalscience.ir

Artificial intelligence (machine learning) in the psychology of learning: Unveiling new insights and directions

Nora Darjazini[1], Mohammad Hossein Zarghami[2], Reza Ghorban Jahromi[3], Leila Shobeiry[4]
1. Ph.D Candidate in Educational Psychology, Science and Research Unit, Islamic Azad University, Tehran, Iran. E-mail: n.darjazini16@gmail.com
2. Assistant Professor, Behavioral Sciences Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran. E-mail: Zar100@gmail.com
3. Assistant Professor, Department of Educational and Personality Psychology, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: rrghorban@gmail.com
4. Assistant Professor, Department of French Language, Faculty of Literature, Humanities and Social Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran. E-mail: shobeiri@srbiau.ac.ir

ARTICLE INFO

Article type:
Research Article

Article history:
Received 18 February 2024
Received in revised form 17 March 2024
Accepted 19 April 2024
Published Online 21 November 2024

Keywords:
machine learning,
reading comprehension,
natural language processing,
syntactic and lexical features,
learning english

ABSTRACT

Background: The intersection of artificial intelligence (AI), psychology and applied linguistics particularly in the realm of language learning, has opened up a fascinating avenue for exploring the intricate processes and mechanisms underlying human cognition. Machine learning algorithms have the potential to shed light on the fundamental principles of learning language specially on reading comprehension as the core language learning variable.

Aims: In this research, we used supervised machine learning techniques in order to discover the most important syntactic and lexical features affecting the reading comprehension of English language learners.

Methods: The design of present study is causal comparative type (ex post facto). the population includes all second secondary level students who learn English in language training institutions. To select the participants, language training institutes in Tehran were referred. The participants (n=360) answered BALA exam (Young, 2022) questions in written and spoken form.

Results: 260 features were extracted from the computer texts prepared from the speech and writing learners responses by natural language processing (NLP) algorithms. We used learning models of decision tree, nearest neighbor, support vector method, neural network and regularized linear method to predict reading comprehension using extracted linguistic features.

Conclusion: The results showed that the variance of language learners' reading comprehension can be well modeled using the extracted grammatical and lexical features, and in addition, twenty features that play the most important role in explaining the variance were identified. This study shows that ML methods can determine the detailed investigation of language processes related to reading comprehension.

Citation: Darjazini, N., Zarghami, M.H., Ghorban Jahromi, R., & Shobeiry, L. (2024). Artificial intelligence (machine learning) in the psychology of learning: Unveiling new insights and directions. *Journal of Psychological Science*, 23(141), 179-197. 10.52547/JPS.23.141.179

Journal of Psychological Science, Vol. 23, No. 141, 2024
 DOI: 10.52547/JPS.23.141.179

✉ **Corresponding Author:** Mohammad Hossein Zarghami, Assistant Professor, Behavioral Sciences Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran.
E-mail: Zar100@gmail.com, Tel: (+98) 9122263167

Extended Abstract

Introduction

In the contemporary landscape of cognitive research, the convergence of artificial intelligence (AI), psychology, and applied linguistics has given rise to a compelling exploration, particularly within the domain of language learning. This study embarks on a nuanced investigation into the intricate processes and mechanisms that underlie human cognition, with a specific focus on the intersection of artificial intelligence, psychology, and applied linguistics in the context of language acquisition. Notably, the utilization of machine learning algorithms introduces a paradigm shift, providing a novel lens through which we can discern the fundamental principles governing language learning, with a particular emphasis on the pivotal variable of reading comprehension (Poulsen & Gravgaard, 2016).

The backdrop of this research is framed within the dynamic interplay of AI and psychology, where the field of machine learning contributes significantly to our understanding of the cognitive intricacies involved in language acquisition. Applied linguistics, as a complementary discipline, amplifies the relevance of this research by contextualizing it within the realm of language learning, accentuating its practical implications (Truckenmiller et al., 2021).

The central aim of this study is to employ supervised machine learning techniques to unravel the essential syntactic and lexical features that exert a discernible impact on the reading comprehension abilities of English language learners. Through this meticulous exploration, we endeavor to unveil new insights and directions in the psychology of learning, thus contributing to the burgeoning body of knowledge at the confluence of AI, psychology and applied linguistics (Graesser & Sabatini, 2022).

This research, characterized by its methodological rigor and interdisciplinary approach, stands poised to illuminate hitherto unexplored facets of language acquisition. By delving into the intricate interplay between artificial intelligence, psychology, and applied linguistics, it seeks to not only advance our theoretical understanding but also to offer practical implications for educators, psychologists, and researchers alike. As we embark on this intellectual journey, the pursuit of uncovering the intricacies of reading comprehension in language learning shall unfold, guided by the principles of scientific inquiry and the commitment to enriching our comprehension of the human cognitive landscape.

The foundational underpinnings of language across diverse domains encompass two primary structures: vocabulary, encapsulating the depth and breadth of the lexical repertoire an individual can comprehend and produce, and syntax, governing the intricacies and precision of grammatical structures within language comprehension and production. Numerous studies have been undertaken to gauge the variance in reading comprehension, manifested in both oral and written discourse, with a notable example being the work by Brimo et al. in 2017. However, many studies in this domain have concentrated on metrics that do not provide profound insights, often focusing on assessing knowledge at rudimentary levels rather than delving into deeper cognitive processes (Ellis & Roever, 2021).

Language proficiency and literacy demand cognitive processes that extend beyond the scope of conventional multiple-choice response assessments, necessitating exploration of both constructive and deconstructed language responses (Pearson & Hamm, 2005). This prompts an essential inquiry into how advanced technologies, such as Machine Learning (ML) and Natural Language Processing (NLP), can facilitate the evaluation of both constructive and deconstructed language, thus contributing to a comprehensive understanding of language skills. Integrating ML and NLP in addressing this question can significantly influence the overall quality of data analysis in educational research and play a pivotal role in validating empirical models.

In machine learning, a set of data, known as the training set, enables algorithms to identify variables correlated with reading comprehension as an output variable. The nature of this correlation is characterized by direction and strength. Furthermore, a separate set of data, referred to as the test set, is employed to assess the performance and evaluate the applicability of the employed algorithm. This study aims to comprehend the development of language

skills in relation to reading proficiency and investigate how Natural Language Processing and Machine Learning contribute to a deeper understanding beyond traditional analytical indicators and techniques. Researchers also delve into determining what proportion of variance in learners' reading comprehension can be accounted for by lexical and syntactic features extracted through natural language processing. This study aims to ascertain the explanatory power of lexical and syntactic features derived from NLP in understanding the variance in learners' reading comprehension.

Method

The research methodology employed in this study entails quantitative modeling within the framework of a quasi-experimental design. The focus of this investigation lies within the methodological domain of language learning, with data collected to model the English language reading comprehension of high school students. The comparative foundation involves the utilization of various machine learning techniques, all grounded in relational paradigms, rendering correlation the central perspective of inquiry. To comprehend the diverse methodological aspects of this research, it is imperative to outline the research design. The primary variable (dependent variable) in this study is the reading comprehension of language learners. Research indicates that the proficiency level of language learners is correlated with their adoption of comprehension strategies. Consequently, participants in this study are categorized into two groups: those with "high proficiency" and those with "low proficiency" in language skills, based on scores assigned by language learning institutions. Participants' responses to reading skill questions are obtained through both oral and written formats, providing a comprehensive evaluation of their language comprehension abilities. For this research, a total of 180 proficient language learners with high levels of reading comprehension skills and 180 students with low levels of reading comprehension skills volunteered from language learning centers in Tehran, with their educational level being secondary school. In similar studies, a maximum sample size of 50 individuals per group has been considered (Yap et al., 2021; Gomez-Marino et al., 2021; Trocano Miller et al., 2021). However, due to potential test fatigue, increased reliability of estimated statistical parameters, and enhanced generalizability of results, this research opted for a sample size of 180 participants in each group (high and low proficiency). Students positioned in advanced language learning levels constitute the high proficiency group, while those in introductory language learning levels constitute the low proficiency group. The researcher's hypothesis posits that high-proficiency language learners necessitate greater language skills compared to their low-proficiency counterparts.

Results

The first step in data analysis is preprocessing, where the oral and written language of language learners is converted into digital text using software. The primary goal of preprocessing is to extract lexical and syntactic features. To achieve this, algorithms from the COVFEFE software package (Comly et al., 2019) under the Python environment are utilized. Features serve as predictor variables, representing factors aimed at reducing the volume of data without losing essential information. The preprocessing output, conducted through the lexicosyntactic pipeline of the COVFEFE package, yields a file with comma-separated values. These values represent the 260 extracted lexical and syntactic variables for each input file (generated based on written and spoken responses of language learners).

The preprocessing involves spoken responses related to Image Description 1 (168=n), Image Description 2 (168=n), Story Retelling 1 (168=n), and Story Retelling 2 (140=n), where participants orally express their interpretations of two images and two stories. Additionally, the dataset pertaining to the written language domain includes features related to learners' explanatory responses to questions (154=n) and their written features (155=n) (where n denotes the number of valid responses after data screening in each section). For all open-ended questions, the NLP output includes results related to predicting participants' reading comprehension, their expressed interest in the topic, questions they formulated, and their responses to high-level comprehension questions. In total, 1567 units of analysis related to

reading comprehension (with a common set of 260 lexicosyntactic variables in all of them) were obtained for 154 participating students. Given that machine learning does not perform well in handling missing data and considering the need to manage the sparsity and diversity of learners' responses, the dimensions of this data are reduced.

To achieve this, the average of each feature across each individual's open-ended responses with common features and both reading texts is computed. Subsequently, each feature from this set of averaged features is extracted from the output of the writing task (text-based) by taking the average. This approach prevents the introduction of missing values into the analysis and avoids negatively impacting the analysis domain of a missing element for a specific task. After processing the two sets of features, each consisting of 260 variables, 165 variables are extracted from spoken responses and 166 variables are extracted from written and textual responses (with some overlap between the extracted variables). Following this process, variables with no variance are removed using the 'nearZeroVar' command in R. The comprehension output (BALA) is then combined with the oral and written datasets for each participant. For analysis purposes, the obtained datasets are divided into two categories: those obtained through the regular version and those obtained through the modified version.

Table 1. Linguistic features extracted through natural language processing

Constituent grammatical components at the word level	Oral	Oral/corrected	Written	Written/corrected
Number of words	0/09	0/17	0/34	-0/06
Adjectives	0/01	0/35	0/01	0/09
Adverbs	0/32	-0/02	-0/06	-0/13
Coordinates	0/06	-0/17	0/04	-0/18
Demonstratives	0/03	0/03	0/28	-0/1
Determiners	-0/09	-0/17	0/13	0/15
Inflected verbs	0/2	0/04	<.01	0/18
Light verbs (be, have, come, go, give, take, make, do, get, move, put)	-0/01	<.01	0/04	0/02
Nouns	-0/14	-0/04	-0/03	0/05
Function words	0/1	-0/06	0/06	-0/07
Prepositions	0/07	0/12	0/09	0/03
Personal pronouns	0/18	0/06	-0/14	-0/13
Subordinating conjunctions	0/23	0/1	0/18	0/13
Verbs	0/25	0/08	-0/16	0/06
Ratio of nouns to nouns and verbs	-0/24	-0/07	0/12	<.01
Ratio of nouns to verbs	-0/15	0/14	0/05	-0/06
Ratio of Personal pronouns to Personal pronouns and nouns	0/16	0/07	-0/09	-0/11
Adverbial phrase consisting of an adverb	0/22	0/09	0/01	-0/02

Conclusion

The current research aimed to identify the most influential syntactic and lexical features affecting the English language comprehension of learners using supervised machine learning techniques. This study may be one of the first to utilize entirely granular features from Natural Language Processing (NLP) in response to uniformly obtained language learner speech and writing, enabling the comparison of modeling approaches across different response methods.

Supervised machine learning methods were employed to investigate whether 260 grammatical and lexical features extracted from NLP could predict reading comprehension. To assess the performance of various models, results were compared with their baseline averages, and models with the least error in each of the four datasets were selected. For both text-based datasets (normal and modified), the Random Forest algorithm proved to be the best. Through cross-validation, there was a 12.75% relative reduction in error for written/normal data, while for written/modified data, there was only a 2.4% reduction in error. In the oral/normal method, the Support Vector Machine model exhibited the best performance (6.07% relative error reduction), whereas the Gradient Boosting algorithm demonstrated the best performance for oral/modified data, reducing the error by approximately 17.21%. Although there was no distinct pattern indicating

which reading comprehension version or language domain was superior in terms of relative error reduction among the four models, it can be generally stated that the oral/modified, written/normal, oral/normal, and written/modified methods experienced the highest to lowest reduction in errors, respectively.

In summary, machine learning algorithms exhibited varying degrees of success across datasets, and while there was no interpretable method for their success, the research expanded the foundational relationships between spoken language and reading comprehension. Notably, considerable variance in reading comprehension can be modeled through syntactic and lexical features extracted from language learner speech and writing, highlighting their potential significance in understanding language acquisition.

Ethical Considerations

Compliance with ethical guidelines: This article is taken from the doctoral dissertation of the first author in the field of educational Psychology in the Faculty of Psychology, Islamic Azad University Science and Research Branch. In order to maintain the observance of ethical principles in this study, an attempt was made to collect information after obtaining the consent of the participants. Participants were also reassured about the confidentiality of the protection of personal information and the presentation of results without mentioning the names and details of the identity of individuals

Funding: This study was conducted as a PhD thesis with no financial support.

Authors' contribution: The first author was the senior author, the second were the supervisors and the third was the advisors.

Conflict of interest: the authors declare no conflict of interest for this study.

Acknowledgments: I would like to appreciate the supervisor, the advisors, in the study.

مجله علوم روانشناختی

شاپا چاپی: ۷۴۶۲-۱۷۳۵ شاپا الکترونیکی: ۶۶۳۹-۲۶۷۶

Homepage: http://www.psychologicalscience.ir

کاربرد هوش مصنوعی (یادگیری ماشینی) در روانشناسی یادگیری: رونمایی از بینش‌ها و جهت‌گیری‌های جدید

نورا درجزینی[۱]، محمدحسین ضرغامی[۲]✉، رضا قربان جهرمی[۳]، لیلا شوبیری[۴]

۱. دانشجوی دکتری روانشناسی تربیتی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

۲. استادیار، مرکز تحقیقات علوم رفتاری، دانشگاه علوم‌پزشکی بقیه‌الله، تهران، ایران.

۳. استادیار، گروه روانشناسی تربیتی و شخصیت، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

۴. استادیار، گروه تخصصی فرانسه، دانشکده ادبیات، علوم‌انسانی و اجتماعی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

مشخصات مقاله

نوع مقاله:
پژوهشی

تاریخچه مقاله:
دریافت: ۱۴۰۲/۱۱/۲۹
بازنگری: ۱۴۰۲/۱۲/۲۷
پذیرش: ۱۴۰۳/۰۱/۳۱
انتشار برخط: ۱۴۰۳/۰۹/۰۱

کلیدواژه‌ها:
یادگیری ماشین،
درک مطلب،
پردازش زبان طبیعی،
ویژگی‌های نحوی و واژگانی،
یادگیری زبان انگلیسی

چکیده

زمینه: تلاقی هوش مصنوعی (AI)، روانشناسی و زبان‌شناسی کاربردی به ویژه در حوزه یادگیری زبان، مسیری جذاب را برای کاوش در فرآیندها و مکانیسم‌های پیچیده زیربنایی شناخت انسان باز کرده است. الگوریتم‌های یادگیری ماشینی این پتانسیل را دارند که اصول اساسی یادگیری زبان به‌ویژه درک مطلب را به عنوان متغیر اصلی یادگیری زبان روشن کنند ولی بررسی ادبیات این حوزه نشان داد تاکنون مطالعه‌ای به بررسی مهم‌ترین ویژگی‌های نحوی و واژگانی مؤثر بر درک مطلب زبان آموزان انگلیسی با تکنیک‌های یادگیری ماشین انجام نشده است.

هدف: هدف پژوهش حاضر شناسایی مهم‌ترین ویژگی‌های نحوی و واژگانی مؤثر بر درک مطلب زبان‌آموزان انگلیسی با تکنیک‌های یادگیری ماشین نظارت شده بود.

روش: طرح پژوهش حاضر از نوع علّی مقایسه‌ای بود. جامعه آماری در این مطالعه شامل همه دانش‌آموزان متوسطه دوم شهر تهران در سال ۱۴۰۲ بود که زبان انگلیسی را در مؤسسات آموزش زبان یاد می‌گیرند. از این میان با استفاده از روش نمونه‌گیری در دسترس ۳۶۰ نفر از آن‌ها انتخاب شدند. ابزار گردآوری داده‌ها شامل آزمون BALA (یانگ، ۲۰۲۲) به صورت کتبی و گفتاری بود. برای تحلیل داده‌ها از الگوریتم‌های موجود در بسته نرم‌افزاری COVFEFE (کمیلی و سایرین، ۲۰۱۹) تحت نرم‌افزار Python استفاده شد.

یافته ها: ۲۶۰ ویژگی از متون کامپیوتری تهیه‌شده از پاسخ‌های گفتاری و نوشتاری زبان‌آموزان با استفاده از الگوریتم‌های پردازش زبان طبیعی (NLP) استخراج شد. سپس از مدل‌های یادگیری درخت تصمیم، نزدیک‌ترین همسایه، روش بردار پشتیبان، شبکه عصبی و روش خطی منظم برای پیش‌بینی درک مطلب با استفاده از ویژگی‌های زبانی استخراج‌شده استفاده شد.

نتیجه‌گیری: نتایج نشان داد که تغییرات درک مطلب خواندن زبان‌آموزان را می‌توان با استفاده از ویژگی‌های دستوری و واژگانی استخراج شده به خوبی مدل‌سازی کرد. علاوه بر این، بیست ویژگی که بیشترین نقش را در تبیین واریانس دارند، شناسایی شد. این مطالعه نشان می‌دهد که روش‌های ML می‌توانند بررسی دقیق فرآیندهای زبانی مربوط به درک مطلب را تعیین کنند.

استناد: درجزینی، نورا؛ ضرغامی، محمدحسین؛ قربان جهرمی، رضا؛ و شوبیری، لیلا (۱۴۰۳). کاربرد هوش مصنوعی (یادگیری ماشینی) در روانشناسی یادگیری: رونمایی از بینش‌ها و جهت‌گیری‌های جدید. مجله علوم روانشناختی، دوره ۲۳، شماره ۱۴۱، ۱۷۹-۱۹۷.

مجله علوم روانشناختی، دوره ۲۳، شماره ۱۴۱، ۱۴۰۳. DOI: 10.52547/JPS.23.141.179

✉ **نویسنده مسئول:** محمدحسین ضرغامی، استادیار، مرکز تحقیقات علوم رفتاری، دانشگاه علوم‌پزشکی بقیه‌الله، تهران، ایران. رایانامه: zar100@gmail.com

تلفن: ۰۹۱۲۲۲۶۳۱۶۷

مقدمه

زمانی که موضوع هوش مصنوعی[1] مطرح شد بسیاری فکر نمی‌کردند این تکنولوژی تا چه حد می‌تواند در مسیر تاریخ بر زندگی انسان‌ها تأثیر بگذارد. با وجود این وقتی کاربردهایی از هوش مصنوعی ارائه شد بر آن شدند تا درک درستی از این حوزهٔ نوظهور پیدا کنند (بالوم و همکاران، ۲۰۲۰). این افراد تنها پژوهشگران و فعالان در حوزه‌های علمی و دانشگاهی نبودند بلکه بسیاری از سیاست‌مداران، شرکت‌ها و شاغلین در حرفه‌های مختلف را در بر می‌گرفت به قول برایان بلاها استاد دانشگاه ایلینویز «سونامی هوش مصنوعی» اتفاق افتاده است و باید برای آن آماده شد. تعبیر سونامی برای هوش مصنوعی نشان می‌دهد که جوامع باید برای برخورد آن آمادگی لازم را کسب کنند. در هر صورت در امواج بلند این سونامی می‌توان از موج سواری لذت برد و یا شاهد شکستن شدن و خرد شدن بود (یانگ، ۲۰۲۲). حوزه‌ی تحقیق، پژوهش و آموزش نیز مانند سایر حوزه‌ها تحت تأثیر و نفوذ گسترش روزافزون هوش مصنوعی قرار گرفته است. هوش مصنوعی (AI) به سرعت حوزه روانشناسی تربیتی را متحول کرده و می‌کند و راه‌های جدید و نوآورانه‌ای را برای افزایش بازده آموزش، یادگیری و سنجش و ارزیابی عادلانه ارائه می‌دهد (کامیلی و همکاران، ۲۰۱۹). ابزارها و فناوری‌های مبتنی بر هوش مصنوعی بینش ارزشمندی را در مورد رفتار دانش‌آموز، الگوهای یادگیری و نقاط قوت و ضعف شناختی در اختیار معلمان و مربیان و دست‌اندارکاران آموزش و پژوهش قرار می‌دهند. از این اطلاعات می توان برای شخصی سازی تجربیات یادگیری، ارائه پشتیبانی هدفمند و بهبود نتایج کلی آموزشی استفاده کرد (تراکمیلر و همکاران، ۲۰۲۱). به عنوان مثال می‌توان به مواردی مانند یادگیری شخصی شده[2] که الگوریتم‌های هوش مصنوعی می‌توانند حجم وسیعی از داده‌ها را در مورد عملکرد دانش‌آموز، سبک‌های یادگیری و ترجیحات فرد یادگیرنده برای ایجاد برنامه‌های یادگیری شخصی‌سازی شده تجزیه و تحلیل کرده و بر اساس برنامه‌ی آموزشی پیشنهادی وی را با نیازهای فردی اش انطباق دهند و تکالیف تمرینی سازگار با قابلیت‌های و توانایی‌های یادگیرنده ارائه دهند و به منظور تقویت و توسعه‌ی این توانمندی‌ها منابع و تکالیف مربوطه را توصیه کنند. سیستم‌های آموزشی هوشمند[3]: سیستم‌های آموزشی مبتنی بر هوش مصنوعی می‌توانند آموزش‌ها و بازخوردهای شخصی‌شده را در زمان واقعی به دانش‌آموزان ارائه دهند (گومزمیرنو و همکاران، ۲۰۲۱). این سیستم‌ها می‌توانند به سرعت با سطح توانایی یاد گیرنده منطبق شوند، در صورت نیاز نکات و توضیحاتی را ارائه دهند و مسیرهای یادگیری جایگزین را بر اساس پیشرفت آن‌ها مشخص نمایند. ارزیابی خودکار[4]: هوش مصنوعی می‌تواند امتیازدهی، سنجش و آزمون گیری و همچنین تکالیف هر یادگیرنده را خودکار نماید و در مورد عملکرد دانش‌آموزان به معلمان بازخورد فوری ارائه دهد. این موضوع می تواند به شناسایی حوزه‌هایی که یادگیرنده به حمایت بیشتر نیاز دارد کمک کند و به تصمیم‌گیرندگان در حوزه‌ی آموزشی گزارش دهد. ارائه‌ی تحلیل‌های پیش‌بینی‌کننده[5]: هوش مصنوعی می‌تواند داده‌های مربوط به متغیرهای مختلف با ابعاد بالا یادگیرندگان را برای پیش‌بینی عملکرد آینده و شناسایی چالش‌های بالقوه آموزشی و تربیتی استفاده کند. از این اطلاعات می‌توان برای ارائه مداخلات اولیه، جلوگیری از عقب افتادن یاد گیرندگان و هدایت برنامه‌های یادگیری شخصی استفاده کرد و چت ربات‌های آموزشی[6]: چت ربات‌های مجهز به هوش مصنوعی می‌توانند به یادگیرنده‌گان دسترسی ۲۴ ساعته را برای پشتیبانی، پاسخ به سؤالات و ارائه راهنمایی فراهم کنند. چت ربات‌ها همچنین می‌توانند به می‌تواند به منظور تولید بازی‌ها با هدف یادگیری، ارائه بازخورد شخصی و تشویق مشارکت یادگیرندگان استفاده شوند (بریمو و همکاران، ۲۰۱۷). با وجود این استفاده از هوش مصنوعی در روانشناسی تربیتی هنوز در مراحل اولیه است ولی این پتانسیل را دارد که روش آموزش و یادگیری را متحول کند. همانطور که فناوری هوش مصنوعی به توسعه خود ادامه می‌دهد، می‌توانیم انتظار داشته باشیم که حتی برنامه‌های نوآورانه‌تری را ببینیم که تجربه یادگیری را برای همه یادگیرندگان عمیق‌تر می‌کند. به طور کلی در حوزه‌ی تربیتی هوش مصنوعی روی دو موضوع عمده تمرکز داشته است: آموزش و یادگیری[7]، و سنجش و بازخورد تحصیلی[8].

[1]. Artificial intelligence
[2]. Personalized Learning
[3]. Intelligent Tutoring Systems
[4]. Automated Assessment
[5]. Predictive Analytics
[6]. Predictive Analytics
[7]. Teaching and Learning
[8]. Teaching and Learning

زمانی که موضوع مطالعه یک فرآیند آموزشی چند بعدی است حوزه‌های تحقیقی نیازمند تحقیقات بین رشته‌ای می‌باشند. یکی از موضوعات چند بعدی، چند متغییره چند رویه‌ای درک مطلب خواندن است که حوزه‌های گوناگونی از زبان‌شناسی کاربردی، روانشناسی، علوم‌شناختی، علوم اعصاب، فلسفه ذهن و مدل‌سازی را به طور همزمان دربرمی‌گیرد. از طرف دیگر هوش مصنوعی یک دانش بین رشته‌ای است که دقیقاً با همین رشته‌های دانشگاهی در ارتباط است. ازاین‌رو تکنیک‌های هوش مصنوعی سازگاری بالایی با شرایط مطالعه‌ی درک مطلب خواندن دارند. درک مطلب را می‌توان یکی از دغدغه‌های اصلی زبان آموزان در حوزه‌ی آموزش زبان در نظر گرفت. از طرفی گرسر و سابتینی (۲۰۲۲) یادگیری زبان را یکی از مقولات مهم حوزه‌ی روانشناسی تربیتی شمرده‌اند و در مقالهٔ جدید خود با عنوان "روانشناسی تربیتی به منظور تطبیق با فناوری، رشته های متعدد و مهارت‌های قرن بیست و یکم، در حال تکامل است" بر ضرورت گسترش روانشناسی تربیتی با هدف انطباق با شرایط جدید تأکید دارند. مقاله حاضر نیز فعالیت‌های پژوهشی اخیر در روانشناسی تربیتی (زبان آموزی) را پوشش می‌دهد و ماهیتی میان رشته‌ای دارد. گرسر و سابیتنی (۲۰۲۲) در کار خود یادگیری مهارت‌های قرن بیست و یکم در تکامل روانشناسی تربیتی را ضروری می‌داند و بر فناوری‌های دیجیتال در کنار پایه‌های سنتی سواد، حساب، علم، استدلال (حل مسئله) و موضوعات آکادمیک دیگر تأکید می‌کنند چرا که معتقدند استفاده از فناوری‌های نوین داده‌های یادگیری را با جزئیات غنی ردیابی می‌کنند و به طور قابل اعتماد مداخلاتی را ارائه دهند که برای یادگیرندگان در زمینه‌های اجتماعی-فرهنگی خاص (بومی) طراحی شده‌اند. از طرف دیگر رویکردهای تحقیقات آموزشی سنتی نیز غیرقابل انعطاف‌اند و عموماً در فرآیند تحقیق سابقه دقیق یادگیری و فعالیت‌های آموزشی افراد را نمی‌توانند به درستی مدیریت نمایند ولی طراحی خوب فناوری آموزشی اصول یادگیری علم را دربرمی‌گیرد، انواع اساسی یادگیری مورد نیاز را شناسایی می‌کند، توانایی های فنی مربوطه را اجرا می‌کند و بازخورد یا بازخوردهای ذینفعان مختلف را در نظر می‌گیرد (گرسر و سابیتنی، ۲۰۲۲). هوش مصنوعی که در قله فناوری‌های آموزشی قرار دارد گستره‌ی وسیعی از تکنیک‌ها را در خود جای می‌دهد با این وجود دو تکنیک مهم هوش مصنوعی که تمرکز اصلی این تحقیق می‌باشد، پردازش زبان طبیعی و یادگیری ماشین است. یادگیری ماشین و پردازش زبان طبیعی در زمینه‌ی سنجش و یادگیری[1] زبان راه‌های نوینی را ایجاد کرده‌اند. به عبارتی این تکنیک‌ها مانند ابزاری می‌مانند که برای رانندگی در دل کوه‌ها جاده می سازند. از این تکنیک‌ها به صورت مقدماتی در سنجش روان‌خوانی شفاهی یادگیرندگان (بلک و همکاران، ۲۰۰۸)، در حوزه‌ی سنجش نوشتن و سخنرانی‌های خود به خودی (ایوانی و همکاران، ۲۰۱۵) استفاده کرده‌اند. در ادبیات [2] از تکنیک‌های یادگیری ماشین برای مطالعه‌ی رابطه‌ی بین حرکات چشمی و ریسک‌های نارساخوانی[3] (بنفاتو و سایرین، ۲۰۱۶) استفاده شده است. اما درک خواندن یکی از مهارت‌های اساسی در حوزه یادگیری زبان است که با مهارت‌های مختلفی در ارتباط است به گونه‌ای که محل تلاقی مهارت‌های مختلف به شمار می‌رود مهم‌ترین نگرانی حوزه‌ی یادگیری و آموزش زبان آموزی است. از طرف دیگر با توجه به قدرت بالای تکنولوژی‌های پردازش زبان طبیعی و یادگیری ماشین، اهمیت کاربست این تکنولوژی‌ها در حوزه‌ی درک خواندن بیش از پیش مشخص است (گرسر و سایتینی، ۲۰۲۲).

دو ساختار اصلی زیربنای همه حوزه‌های زبان عبارتند از واژگان (عمق و دامنه واژگانی که فرد می‌تواند بفهمد و تولید کند) و نحو که پیچیدگی و دقت دستور زبانی که فرد می‌تواند بفهمد و تولید کند (تراکمیلر و همکاران، ۲۰۲۱). برای فهم میزان واریانسی از درک مطلب خواندن را که می‌توان در قالب‌های گفتاری و نوشتاری تبیین کرد، مطالعات بسیاری انجام شده است (به عنوان مثال، بریمو و همکاران، ۲۰۱۷). با این حال، بسیاری از مطالعات در این زمینه بر ملاک‌هایی متمرکز شده‌اند که دانش عمیقی را فراهم نمی‌کنند بلکه به سنجش دانش در سطوح پایین می‌پردازند. مهارت‌های زبانی و سوادآموزی از طریق پاسخ‌های زبانی سازنده و ساخته شده (برساخته)، پردازش شناختی عمیق‌تر را می‌طلبد و اطلاعات بیشتری را نسبت به شیوه‌های آزمون‌گیری مبتنی بر پاسخ‌های انتخابی ایجاد می‌کند (پیرسون و هام، ۲۰۰۵). بنابراین، نیاز مبرمی وجود دارد تا مشخص شود که چگونه فن‌آوری‌های پیشرفته می‌توانند ارزیابی زبان سازنده و برساخته را

[1]. assessment and learning
[2]. Literacy
[3]. Dyslexia

تسهیل کنند؟ ML و پردازش زبان طبیعی در پاسخ به این سؤال می‌تواند چگونگی کیفیت کلی تحلیل داده‌ها در تحقیقات آموزشی را تحت تأثیر خود قرار دهد و نقش تعیین‌کننده‌ای در اعتبارسنجی مدل‌های تجربی بازی کند. در یادگیری ماشین مجموعه‌ای از داده‌ها به عنوان مجموعه آموزش[1] شناخته می‌شوند که الگوریتم‌های یادگیری ماشین را قادر می‌سازد تا مشخص کنند کدام متغییرها با درک خواندن به عنوان متغییر خروجی ارتباط دارد. ارتباط نیز دارای دو مشخصه اصلی است یکی جهت ارتباط و دیگری شدت یا قدرت ارتباط است. هر دو وجه ارتباط از طریق این الگوریتم‌ها مشخص می‌شود. بخش دیگری از داده‌ها که به آن‌ها داده‌های مجموعه‌ی "آزمون" یا تست گفته می‌شود، با هدف ارزشیابی یا ارزیابی الگوریتم بکار گرفته شده استفاده می‌شود. به عبارتی از طریق مجموعه‌ی تست ما به این موضوع می‌پردازیم که الگوریتم چقدر و با چه صحتی روابط آموخته شده را در موقعیت‌های جدید و با ورودهای جدید بکار می‌گیرد. این مطالعه قصد فهم مجموعه مهارت‌های زبانی توسعه یافته‌تر در ارتباط با توانایی خواندن را دارد و این که دریابد چطور پردازش زبان طبیعی و یادگیری ماشین در درک عمیق‌تر و فراتر از آنچه شاخص‌ها و تکنیک‌های تحلیل سنتی فراهم می‌کند، کمک‌کننده می‌باشند. علاوه‌بر این پژوهشگران به این موضوع پرداخته‌اند که چه نسبتی از واریانس درک خواندن یادگیرندگان می‌تواند به عنوان تابعی از ویژگی‌های لکسیکال و سینتکتیک استخراج شده از طریق روش پردازش زبان طبیعی، در نظر گرفته شود. به عبارتی چند درصد از واریانس درک خواندن یادگیرندگان قابل تبیین است؟ نتایج بدست آمده از این بخش به پژوهشگران این امکان را می‌دهد تا تعیین کنند آیا مقدار واریانس تبیین شده توسط ویژگی‌های لکسیکال و سینتکتیک مستخرج از روش پردازش زبان طبیعی قابل اعتنا است؟

روش

الف) طرح پژوهش و شرکت‌کنندگان: روش تحقیق این مطالعه مدل یابی کمّی است که در قالب یک طرح علّی-مقایسه‌ای اجرا شد. تمرکز این پژوهش بر روش‌شناسی در حوزه یادگیری زبان است و داده‌ها با هدف مدل‌بندی درک خواندن زبان انگلیسی دانش‌آموزان متوسطه دوم، جمع آوری شده‌اند. مبنای مقایسه کاربست تکنیک‌های مختلف یادگیری ماشین که همگی مبتنی بر رابطه هستند پژوهش حاضر از نوع همبستگی محسوب می‌شود. تحقیقات نشان می‌دهد که سطح مهارت زبان آموزان و اتخاذ استراتژی‌های درک مطلب با یکدیگر در ارتباط است. به این منظور در این مطالعه زبان‌آموزان شرکت‌کننده در تحیق به دو دسته زبان‌آموزان با «مهارت بالا» و زبان‌آموزان با «مهارت پایین» تقسیم شدند. مبنای این تقسیم‌بندی نمره‌ای است که مؤسسات آموزش زبان به زبان‌آموزان اختصاص داده‌اند. از طرفی پاسخ شرکت‌کنندگان به سؤالات مهارت خواندن هم به صورت شفاهی و هم به صورت کتبی بوده است.

ویژگی‌های لکسیکال و سینتک از طریق پرازش زبان طبیعی از پاسخ‌های زبان‌آموزان استخراج شده‌اند که به عنوان متغییرهای پیش‌بین درک خواندن به حساب می‌آیند. بر اساس طرح جمع‌آوری داده‌ها چهار مجموعه داده گردآوری شد که عبارتند از مجموعه پاسخ‌های کتبی زبان‌آموزان با مهارت بالا، مجموعه پاسخ‌های شفاهی زبان‌آموزان با مهارت بالا، مجموعه پاسخ‌های کتبی زبان‌آموزان با مهارت پایین و مجموعه پاسخ‌های شفاهی زبان‌آموزان با مهارت پایین. از این چهار مجموعه داده ویژگی‌های لکسیکال و سنتیک مبتنی بر پردازش زبان طبیعی استخراج شده است. برای انجام این پژوهش ۱۸۰ نفر زبان‌آموز با سطح مهارت درک خواندن زبانی بالا و ۱۸۰ دانش‌آموز با سطح مهارت درک خواندن زبانی پایین به صورت در دسترس و داوطلبانه از زبان‌آموزانی که به مراکز آموزش زبان در شهر تهران مراجعه کرده بودند و مقطع تحصیلی آن‌ها متوسطه دوم بود انتخاب شده‌اند. در تحقیقات مشابه حداکثر حجم نمونه ۵۰ نفر در هر گروه در نظر گرفته شده است (یاپ و همکاران، ۲۰۲۱؛ گومز مرینو و همکاران، ۲۰۲۱؛ تروکن میلر و همکاران، ۲۰۲۱). به دلیل امکان افت آزمودنی‌ها، افزایش پایایی پارامترهای آماری برآورد شده و همچنین افزایش قدرت تعمیم نتایج در این تحقیق ۱۸۰ نفر در هر گروه (با مهارت بالا و پایین) انتخاب شد. ملاک ورود به پژوهش مشارکت داوطلبانه، تحصیل در مقطع متوسطه دوم بود و ملاک خروج عدم رضایت برای ادامه همکاری و تکمیل ناقص ابزارهای مطالعه بود. دانش‌آموزانی که در سطوح بالای آموزش زبان قرار دارند به عنوان گروه نمونه‌ی دارای مهارت زبانی بالا و دانش‌آموزان سطوح اولیه آموزش زبان به عنوان گروه نمونه دارای مهارت زبانی پایین در نظر

[1]. training

گرفته شدند. پیش‌فرض پژوهشگر این است که زبان آموزی در سطوح بالا نیازمند مهارت زبانی بیشتری نسبت به سطوح پایین است.

ب) ابزار

ابزار درک خواندن بلا[1] (BALA): برای سنجش وجوه مختلف درک خواندن در این پژوهش از ابزار درک خواندن که یک ابزار جدید سنجش درک خواندن است و دو فرمت مختلف دارد و امکان سنجش را هم به صورت مداد کاغذی و هم به صورت کامپیوتری تحت وب (آنلاین) فراهم می‌کند، استفاده شد. بلا مهارت‌های مختلف درک خواندن را از طریق یک متن روایی[2] و یک متن غیر-داستانی[3] به دست می‌آورد. دو نسخه از بِلا تا کنون گسترش یافته است نسخه معمول و نسخه اصلاح شده[4] آن. نسخه معمول بِلا از ۱۸ سؤال چند گزینه‌ای (با مقدار آلفای کرونباخ ۰/۸) و نسخه اصلاح شده‌ی بلا از ۱۷ سؤال چند گزینه‌ای (با مقدار پایایی ۰/۸۵) تشکیل شده است. نسخه معمول بلا برای زبان آموزانی است که مشکلی جدی در خواندن (مثل مشکلات خواندن) نداشته و مهارت نسبتاً خوبی در زبان آموزی دارند. نسخه اصلاح شده بِلا برای زبان‌آموزانی استفاده می‌شود که مهارت زبانی پایین یا مشکلات جدی در خواندن دارند (که این افراد هم اکثراً به ترم‌های پایین زبان‌آموزی تعلق دارند). در این پژوهش زبان‌آموزان با مهارت بالا از نسخه معمول و زبان‌آموزان با مهارت پایین از نسخه اصلاح شده‌ی بلا استفاده کرده‌اند. از طرف دیگر هر زبان‌آموز باید به سؤالات باز-پاسخ درک خواندن درباره‌ی متونی که از بلا دریافت می‌کرد به صورت تشریحی پاسخ دهند. نسخه معمول بِلا از ۱۸ سؤال چند گزینه‌ای با مقدار آلفای کرونباخ ۰/۸۰ و نسخه اصلاح شده بلا از ۱۷ سؤال چند گزینه ای با مقدار پایایی ۰/۸۵ تشکیل شده است.

تکلیف نوشتاری محقق ساخته: دومین ابزار یک تکلیف نوشتاری جداگانه است. در این تکلیف ویدیویی به زبان‌آموز نشان داده می‌شود که در مورد استفاده دانش‌آموزان از شبکه‌های اجتماعی است. سپس از آن‌ها پرسیده شد که "به نظر شما استفاده از شبکه‌های اجتماعی برای افراد نوجوان خوب است یا بد؟" پاسخ‌های تشریحی زبان‌آموزان به این سؤالات باز-پاسخ (که هم در مورد متون موجود در بِلا است و هم پاسخ آن‌ها به قضاوت در مورد استفاده نوجوان از شبکه های اجتماعی) به عنوان منبع داده‌های خام برای استخراج ویژگی‌های زبان‌شناختی مبتنی بر متن (کتبی) محسوب می‌شود. به منظور دستیابی به داده‌های خام مبتنی بر گفتار (پاسخ‌های شفاهی) زبان آموزان، از یک اپلیکیشن که برای دانش‌آموزان مقطع دوم متوسطه مناسب سازی شده بود، استفاده شد. سه چالش اصلی در استفاده از این اپلیکیشن وجود داشت: الف) اطمینان از بیان درست و دقیق و یکپارچه دستورالعمل به زبان‌آموز ب) ضبط (رکورد) کردن صدای زبان‌آموز و ج) هزینه‌ی ساخت ابزار. در مواجه با این چالش‌ها دستورالعمل‌ها توسط یک معلم زبان خوانده و با کیفیت بالا ضبط شد و این صدا برای تک‌تک شرکت کنندگان پخش شد. امکان پخش مجدد دستورالعمل و متون شفاهی در صورت درخواست زبان‌آموز فراهم بود. دو تکلیف برای زبان‌آموزان محتوای تکلیف‌ها یکی توضیح تصویر است و دیگری تعریف مجدد داستانی که زبان‌آموز می‌شنود. بیان شفاهی زبان‌آموز به این دو محرک مبنای استخراج ویژگی‌های زبانی (لکسیکال و سینتکتیک) قرار گرفت. از طریق مصاحبه با دانش‌آموزان متوسطه دوم دارای روایی با ضریب ۰/۸۵ تشخیص داده شد.

برای دستیابی به هدف پژوهش از الگوریتم‌های موجود در بسته نرم‌افزاری COVFEFE[5] (کمیلی و سایرین، ۲۰۱۹) تحت نرم‌افزار Python استفاده شد. فیچرها متغیرهای پیش‌بین مطالعه می‌باشند و خود عواملی محسوب می‌شوند که به منظور کاهش داده‌ها بدون از دست داده اطلاعات اصلی جایگزین انبوه زیاد داده‌های خام می شوند.

یافته‌ها

مشارکت‌کنندگان در این مطالعه شامل ۳۶۰ نفر بود که ۵۰ درصد آن‌ها دختر و ۵۰ درصد دیگر پسران بودند. همچنین میانگین سن دختران ۱۷ سال و میانگین سن پسران ۱۷/۵ سال بود. گام اول تحلیل داده‌ها پیش پردازش[6] است. گفتار شفاهی و نوشتار کتبی زبان‌آموزان از طریق نرم‌افزار به متن دیجیتالی تبدیل شد. در مرحله ی پیش پردازش هدف اصلی استخراج

[1]. BAlance Literacy Assessment
[2]. narrative
[3]. Non-fiction
[4]. regular and modified
[5]. Automated linguistic data collection for personal assessment
[6]. per-processing

ویژگی‌ها یا فیچر[1]ها می‌باشد که در اینجا ویژگی‌های لکسیکال و سینتکتیک زبان است. نتایج خروجی پیش پردازش توسط خط لوله (پایپ) lexicosyntactic مربوط به بسته COVFEFE فایلی است که ارزش‌های آن (یا مقادیر بدست آمده از آن) از طریق کاما جدا شده‌اند. این ارزش‌ها ۲۶۰ متغیر لکسیس و سینتکتیک استخراج شده برای هر فایل ورودی (فایل های تهیه شده بر اساس پاسخ‌های کتبی و شفاهی زبان‌آموزان) است. پاسخ‌های شفاهی مربوط است به توضیح تصویر ۱ (n= ۱۶۸)، توضیح تصویر ۲ (n= ۱۶۸)، بازگویی داستان ۱ (n= ۱۶۸) و بازگویی داستان ۲ (n= ۱۴۰) می‌باشد که در آن زبان‌آموز برداشت‌های خود را به صورت شفاهی در مورد دو تصویر و دو داستان بیان کرده است. مجموعه داده‌های مربوط به حوزه ی زبان زایای نوشتاری در برگیرنده‌ی ویژگی‌های مربوط به پاسخ‌های تشریحی باز زبان‌آموزان به سؤالات (n= ۱۵۴) و ویژگی‌های نوشتاری (n= ۱۵۵) آن‌ها است (n نشان‌دهنده‌ی تعداد پاسخ‌های معتبر بعد از تمیز داده‌ها در هر بخش است). برای همه‌ی سؤالات باز پاسخ خروجی NLP دربرگیرنده‌ی نتایج مربوط به پیش‌بینی درک خواندن شرکت کنندگان، توضیح علاقمندی‌شان به موضوع، سؤالاتی که دانش‌آموزان ساخته‌اند و پاسخ‌هایشان به سؤال سطح بالای درک مطلب می‌باشد. در مجموع ۱۵۶۷ واحد تحلیل مربوط به خواندن (که در همه‌ی آن‌ها ۲۶۰ متغیر lexicosyntactic مشترک است) برای ۱۵۴ دانش‌آموز شرکت‌کننده نهایی بدست آمد. از آنجایی که یادگیری ماشین نسبت به مدیریت داده‌های گمشده کارایی مطلوبی ندارد و در عین حال به منظور مدیریت پراکندگی و تنک بودن پاسخ زبان‌آموزان لازم است ابعاد این داده‌ها نیز کاهش یابد. برای انجام چنین کاری، میانگین هر ویژگی در پاسخ‌های باز هر فرد با ویژگی‌های مشترک و هر دو متن خواندن بدست آمد. در عوض، هر ویژگی از این مجموعه ویژگی‌های میانگین‌گیری شده برای هر ویژگی یکسان استخراج شده از خروجی تکلیف نوشتاری (کتبی) میانگین گرفته می‌شود. با این کار ارزش‌های گمشده وارد تحلیل نمی‌شوند و بنابراین از دست دادن یک عنصر برای یک تکلیف اثر منفی روی حوزه تحلیل آن نمی‌گذارد. بعد از پردازش دو مجموعه ویژگی وجود دارد که هر کدام ۲۶۰ متغییر دارد که ۱۶۵ متغییر آن استخراج شده از پاسخ‌های شفاهی است و ۱۶۶ ویژگی استخراج شده از متون کتبی و نوشتاری است (بین متغییرهای استخراج شده همپوشی وجود دارد). بعد از این فرآیند ارزش‌های گمشده به خاطر اینکه همه‌ی شرکت‌کنندگان حداقل یک تکلیف گفتاری یا یک تکلیف نوشتاری را کامل کرده‌اند وجود ندارد. متغییرهایی که واریانس ندارند از طریق فرمان nearZeroVar در R حذف شدند. خروجی درک مطلب (BALA) با مجموعه داده‌های شفاهی و کتبی برای هر شرکت‌کننده ترکیب شد. به‌منظور تحلیل، مجموعه داده‌ها بدست آمده به دو دسته تقسیم شدند داده‌هایی که از طریق ورژن معمول بلا بدست آمده‌اند و داده‌هایی که از نسخه اصلاح شده‌ی بلا جمع‌آوری شده‌اند.

جدول ۱. حجم نمونه و میانگین و انحراف استاندارد نمرات درک مطلب بلا (نسبت درست) به تفکیک هر مجموعه داده‌ها

	شاخص توصیفی	ورژن معمول	ورژن اصلاح شده
مجموعه داده‌های استخراج شده از منابع شفاهی (توضیح عکس و بازگویی داستان)	حجم نمونه	۹۵	۷۰
	میانگین	۰/۷۴	۰/۸۱
	انحراف استاندارد	۰/۱۲	۰/۱۳
مجموعه داده‌های استخراج شده از منابع نوشتاری (پاسخ های باز خواندن و نوشتن)	حجم نمونه	۹۹	۶۷
	میانگین	۰/۷۴	۰/۸۲
	انحراف استاندارد	۰/۱۲	۰/۱۳

به منظور تعیین این که چه میزان از واریانس درک مطلب خواندن از طریق ویژگی‌های لکسیکال و سینتکتیک استخراج شده توسط پردازش زبان طبیعی تبیین می‌شود هشت مدل مختلف نظارتی یادگیری ماشین آموزش داده شد. تحلیل‌های یادگیری ماشین بر مبنای ایجاد یک مدل رگرسیونی در پیش‌بینی یک متغییر خروجی پیوسته که در اینجا نمرات درک خواندن است، تنظیم شده‌اند. تمام این مدل‌ها در جولیا و با استفاده از بسته‌ی MLJ.jl (بلائوم و سایرین، ۲۰۲۰) انجام شده است. نتایج مربوط به هشت مدل ML در جدول زیر ارائه شده است.

[1]. feature

جدول ۲. مدل‌های یادگیری ماشین بکار گرفته شده در پژوهش

طبقه کلی مدل	نوع مدل	توضیح
مجموعه‌ی روش‌های مربوط به درخت تصمیم	درخت تصمیم[1]	روش ناپارامتریک که با کامپایل کردن نتایج تعداد مشخصی از درخت‌های تصمیم مستقل، یک نتیجه را پیش‌بینی می کند.
	تقویت گرادیان[2]	به ترتیب از درخت‌های تصمیم یاد می گیرد به طوری که نتایج درخت‌های اولیه نتایج بعدی را "تقویت" می کند و در نتیجه دقت بالاتری را به همراه دارد. مدل تقویت گرادیان ترکیبی خطی از یک سری مدل های ضعیف است که به صورت تناوبی برای ایجاد یک مدل نهائیِ قوی ساخته شده است. این روش به خانواده الگوریتم‌های یادگیری گروهی تعلق دارد و عملکرد آن همواره از الگوریتم‌های اساسی یا ضعیف (مثلاً درخت تصمیم) یا روش‌های براساس کیسه گذاری (بسته‌سازی) (مانند جنگل تصادفی) بهتر است.
	جنگل تصادفی[3]	نتایج حاصل از درخت‌های تصمیم مستقل را کامپایل می کند، اما تنها کسری از k متغیرهای پیش‌بین (معمولاً $\sqrt{k}$) در هر درخت استفاده می‌شود.
روش نزدیک‌ترین همسایه	نزدیک ترین همسایه -k[4]	نتیجه یک نقطه داده ناشناخته را بر اساس میانگین نزدیک‌ترین همسایگان در فضای ویژگی داده یاد گرفته و پیش‌بینی می کند.
روش بردار حمایتی	بردار حمایتی[5]	صفحه‌ی خطی یا غیرخطی را در فضای ویژگی شناسایی می کند و بر اساس آن نتیجه گرفته و پیش‌بینی می کند.
	بردار حمایتی خطی[6]	تمرکز این بردار بر نقاط داده‌ای است که درون یا بیرون یک فاصله‌ی مشخص شده از صفحه قرار دارد. مانند بردار حمایتی است اما از آنجایی که این بردار کرنل‌های خطی را اجرا می کند سریع‌تر است
روش شبکه عصبی	پرسپترون چند لایه[7]	الگوهای خطی یا غیرخطی را با فعال کردن یک سری لایه‌های پنهان (نرون‌ها) که به‌طور متوالی ویژگی‌های ورودی را از طریق یک تابع وزنی تغییر می‌دهند، یاد می گیرد. در لایه خروجی از دست دادن (خطا) اندازه‌گیری می‌شود و با هدف اصلاح در یادگیری بعدی استفاده می‌شود.
روش خطی منظم شده[8]	عمل‌گر گزینش و انقباض[9] کمترین قدرمطلق ۹ یا lasso	ضرایب غیر مهم را با استفاده از پارامتر معمول‌سازی λ (که ضرایب رگرسیون را جریمه می کند) به صفر می‌رساند. هرچه λ بیشتر باشد تعداد بیشتری از ضرایب به صفر تبدیل می‌شود.

از روش Scikit-Learn's ShuffleSplit که یک روش اعتباریابی متقاطع است با هدف ایجاد ۲۰۰ بخش تصادفی در داده‌های ورودی استفاده شد که از این تعداد ۸۰ درصد برای یادگیری یا ترینینگ و بیست درصد برای آزمودن یا تستینگ می‌باشد. میانگین خط پایه‌ی الگوریتم همان میانگین استراتژی در رگرسور ساختگی[10] است.

جدول ۳. ویژگی‌های زبانی استخراج شده از طریق پردازش زبان طبیعی

ویژگی‌های زبانی استخراج شده از طریق پردازش زبان طبیعی				
اجزای گرامری تشکیل‌دهنده در سطح کلمه	شفاهی/معمول	شفاهی/اصلاح شده	کتبی/معمول	کتبی/اصلاح شده
تعداد کلمات Number of words	۰/۰۹	۰/۱۷	۰/۳۴	-۰/۰۶
صفت‌ها Adjectives	۰/۰۱	۰/۳۵	۰/۰۱	۰/۰۹
قیدها Adverbs	۰/۳۲	-۰/۰۲	-۰/۰۶	-۰/۱۳

1. Decision tree
2. Gradient boosting
3. random forest
4. k nearest neighbor
5. Support vector
6. linear support vector
7. Multi-layer perceptron
8. Regularized linear method
9. Least absolute shrinkage & selection operator (Lasso)
10. Dummy Regressor

ویژگی‌های زبانی استخراج شده از طریق پردازش زبان طبیعی				
اجزای گرامری تشکیل‌دهنده در سطح کلمه	شفاهی/معمول	شفاهی/اصلاح شده	کتبی/معمول	کتبی/اصلاح شده
هماهنگ کننده‌ها Coordinates	۰/۰۶	-۰/۱۷	۰/۰۴	-۰/۱۸
اشاره‌ها Demonstratives	۰/۰۳	۰/۰۳	۰/۲۸	-۰/۱
تعیین‌کننده‌ها Determiners	-۰/۰۹	-۰/۱۷	۰/۱۳	۰/۱۵
افعال عطفی Inflected verbs	۰/۲	۰/۰۴	۰/۰۱	۰/۱۸
افعال سبک Light verbs (be, have, come, go, give, take, make, do, get, move, put)	-۰/۰۱	۰/۰۱	۰/۰۴	۰/۰۲
اسامی Nouns	-۰/۱۴	-۰/۰۴	-۰/۰۳	۰/۰۵
کلمات تابعی Function words	۰/۱	-۰/۰۶	۰/۰۶	-۰/۰۷
حروف اضافه Prepositions	۰/۰۷	۰/۱۲	۰/۰۹	۰/۰۳
ضمایر شخصی Personal pronouns	۰/۱۸	۰/۰۶	-۰/۱۴	-۰/۱۳
حروف ربط فرعی (تابعی) Subordinating conjunctions	۰/۲۳	۰/۱	۰/۱۸	۰/۱۳
افعال Verbs	۰/۲۵	۰/۰۸	-۰/۱۶	۰/۰۶
نسبت اسم به اسم و فعل Ratio of nouns to nouns and verbs	-۰/۲۴	-۰/۰۷	۰/۱۲	۰/۰۱
نسبت اسامی به افعال Ratio of nouns to verbs	-۰/۱۵	۰/۱۴	۰/۰۵	-۰/۰۶
نسبت ضمایر شخصی به ضمایر شخصی و اسامی Ratio of Personal pronouns to Personal pronouns and nouns	۰/۱۶	۰/۰۷	-۰/۰۹	-۰/۱۱
عبارت قیدی متشکل از یک قید Adverbial phrase consisting of an adverb	۰/۲۲	۰/۰۹	۰/۰۱	-۰/۰۲

نتایج حاصل از ساخت و آزمون مدل ML در جدول زیر ارائه شده است. حذف بازگشتی ویژگی با اعتبار سنجی متقابل (RFECV) در مرحله‌ی پیش پردازش در همه‌ی مدل‌ها به جز مدل بردار حمایتی یا پشتیبانی (چون این مدل برای بردار حمایتی تعریف نشده است) استفاده شد.

نتایج تحلیل نشان می‌دهد که مدل‌های اصلاح شده در همه‌ی طبقات بر اساس پارامترهای پیش‌فرض عمل می‌کنند. رگرسیون جنگل تصادفی بهترین مدل در هر دو مجموعه داده‌های متنی استخراج شده بود. بهترین مدل برای داده‌های شفاهی/معمول رگرسیون بردار پشتیبان بود و رگرسیون تقویت‌گرادیان به بهترین حالت داده‌های شفاهی/اصلاح شده را پیش‌بینی کرده است.

کاهش خطای نسبی از طریق پیدا کردن تفاوت بین میانگین خط پایه MAE و بهترین مدل MAE اصلاح شده تقسیم بر میانگین خط پایه‌ی MAE بدست می‌آید. مدلی که بهترین اصلاح نسبی را مبتنی بر خط پایه داشته است مربوط به داده‌هایی است که از طریق تکالیف شفاهی (یا گفتاری) استخراج شده بودند (۷/۲۱ درصد صلاح نسبت به خط پایه). در مقام دوم بیشترین اصلاح (نسبت به خط پایه) مربوط به شیوه ی کتبی/معمول با ۷/۱۲ درصد اصلاح قرار دارد و بعد از آن داده‌های شفاهی/معمول با ۶/۰۷ درصد اصلاح و داده‌های کتبی/اصلاح شده با ۲/۴ می باشند. تفاوت‌ها در کاهش خطای نسبی برای هر چهار مدل را می‌توان از طریق تغییر استحکام رابطهٔ پیش‌بینی بین درک مطلب خواندن و مجموعه ویژگی‌های سینتکتیک و لکسیکال تبیین کرد. از آنجایی که داده‌های پرت می‌توانند روی مدل خط پایه تأثیر بگذارند و از طرف دیگر داده‌های پرت رابطه ی قوی با متغییرهای وابسته ندارند ولی در تمام مدل‌ها حتی بهترین مدل‌ها تأثیر می‌گذارند از میانگین متغییر خروجی مدل ترینینگ به جای نمرات واقعی به عنوان خط پایه استفاده شد.

مقادیر R^2 بدست آمده را می‌توان به عنوان واریانسی از درک مطلب خواندن در نظر گرفت که توسط ویژگی‌های سینتکتیک و لکسیکال در تمام مجموعه داده‌ها توضیح داده می‌شود. به عبارتی این مقادیر دقیقاً بر اساس همان تعریف موجود در آمار استنباطی توضیح داده می‌شوند. در داده‌های تستینگ مقادیر R^2 مجذور اختلاف بین نتایج پیش‌بینی شده ($\hat{y}$) از نتایج برآمده (y) برای هر نقطه داده در مجموعه داده‌های تستینگ است (باقیمانده‌ها). این مقدار از طریق تقسیم بر واریانس نتایج خروجی یا برآمده به باقیمانده‌ی استاندارد شده تبدیل می‌شوند. بنابراین از آنجا که هدف ما واریانسی است که توسط ویژگی‌ها تبیین می‌شود کافی است مقدار یک را از این مقدار کم کنیم.

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

تفاوت بین R^2 سنتی که در مدل های رگرسیون سنتی و R^2 یادگیری ماشین در این است که در مدل‌های رگرسیونی سنتی کل واریانس متغیر خروجی (که در آنجا به آن متغییر ملاک یا وابسته گفته می‌شود) از طریق یک مجموعه متغیر مستقل توضیح داده می‌شود. به عبارتی شاخصی است که توضیح می‌دهد یک تابع از متغیرهای ورودی چقدر (به چه خوبی) می‌تواند یک متغییر خروجی را توضیح دهد. از طرفی R^2 در یادگیری ماشین این موضوع را مورد ارزیابی قرار می‌دهد که یک تابع چگونه می‌تواند برای یک مجموعه داده‌ی نادیده (unseen) بکار گرفته شود. این موضوع بیانگر این است که الگوریتم چقدر توانسته است رابطه‌ی بین متغییرهای مستقل (ویژگی‌های زبانی) و متغییر وابسته (نمرات درک مطلب خواندن) را بفهمد (یادبگیرد) و ازاین‌رو چقدر (به چه خوبی) می‌تواند در مجموعه‌ی داده‌های جدید که همان داده‌های نادیده هستند بکار گرفته شود. در جدول زیر خلاصه‌ی بهترین مدل‌ها برای مجموعه داده‌های حاضر در پژوهش (شفاهی/کتبی و منظم/اصلاح شده) ارائه شده است.

جدول ٤. خلاصه‌ی بهترین مدل‌های مربوط به مجموعه داده‌ها (شفاهی/کتبی و منظم/اصلاح شده)

مجموعه داده‌ها	حجم نمونه (حجم ترین، حجم تست)	RFE برای نتایج اعتباریابی متقابل	نوع مدل یادگیری ماشین	خط پایه اعتباریابی متقابل MAE (SD)	اعتباریابی متقابل میزان شده MAE (SD)	کاهش خطای نسبی (CV)	R^2 برای آموزش فردی/جداسازی تست
گفتاری/ منظم	۹۵ (۱۹/۷۶)	n/a	SVR	۱۰/۲۱ (۱/۳۳)	۹/۵۹ (۱/۳۱)	٪۶/۰۷	۲۰/۴
گفتاری/ اصلاح‌شده	۷۰ (۱۴/۵۶)	۴۳ متغیر از ۲۶۰ متغیر	GBR	۱۰/۳۰ (۱/۸۷)	۸/۱۲ (۱/۸۸)	٪۲۱/۱۷	۲۲/۴
کتبی/ منظم	۹۹ (۲۱/۷۹)	۲۱۱ متغیر از ۲۶۰ متغیر	RF	۹/۹۶ (۱/۴۷)	۸/۶۹ (۱/۴۱)	٪۱۲/۷۵	۳۶/۶
کتبی/ اصلاح‌شده	۶۷ (۱۳/۵۴)	۸۷ متغیر از ۲۶۰ متغیر	RF	۱۰/۴۱ (۱/۸۶)	۱۰/۱۶ (۱/۶۶)	٪۲/۴۰	۱۸/۵

در جدول بالا REE مخفف feature elimination recursive به معنی حذف بازگشتی ویژگی و GBR مخفف کلمات boosting regression gradient و به معنی رگرسیون تقویت گرادیانت است. اعتباریابی متقاطع به منظور پیدا کردن بهترین مدل برای هر چهار مجموعه داده انجام شد که نتایج مربوط به آن‌ها در جدول بالا مشاهده می‌شود. از این نتایج می‌توان به منظور مقایسه‌ی مقادیر مجذور همبستگی بین مجموعه داده‌های ترینینگ و داده‌های تستینگ که نسبت ۸۰به ۲۰ استفاده کرد. در این مطالعه مدل میزان شده اجرا شده و سپس پیش‌بینی‌هایی روی داده‌های تست در بخش تستینگ انجام شد. مقادیر مجذور همبستگی R^2 که در آخرین ستون جدول بالا گزارش شده است به اندازه‌ی اطلاعات مدل‌های اعتباریابی متقابل که تحت عنوان MAE در ستون‌های ۵ و ۶ ارائه شده‌اند، پایا نمی‌باشند چرا که این نتایج میانگین ۲۰۰ بار فرآیند ترینینگ-تستینگ روی داده‌ها است.

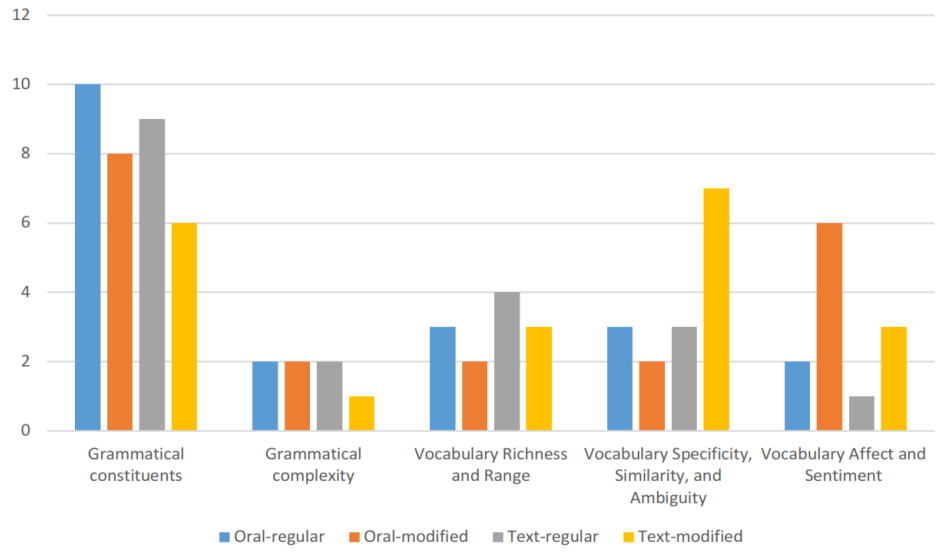
برای مثال، بهترین مقدار R^2 برای مجموعه داده‌ی متنی/معمول است در حالی که این مجموعه داده‌ها بهترین کاهش خطای نسبی اعتبایابی متقابل را نشان نمی‌دهد. زمانی که زیر بخش‌های ترینینگ/تستینگ شکل می گیرد، نقطه داده‌ها به صورت تصادفی به هر کدام از این بخش‌ها تعلق می‌گیرد تا تأثیر انتخاب بر نتیجه ی اعتباریابی کاهش یابد. همانطور که در بالا اشاره شد اگر مشاهدات دارای قدرت نفوذ-بالا[1] در مجموعه داده‌های ترینینگ باشد می‌تواند روی مطلوبیت یا عدم مطلوبیت مدل در پیش‌بینی مجموعه داده‌های تستینگ (که مشاهده نشده اند) تأثیر بگذارد. این موضوع برای زمانی که داده‌های پرت با قدرت نفوذ بالا در مجموعه داده‌های تستینگ هم باشد صادق است. با وجود این وقتی که داده یا داده‌های دارای قدرت نفوذ بالا در بیست درصد داده‌ها یعنی در مجموعه داده‌های تستینگ قرار می‌گیرد بسیار بیشتر نتایج را به تحریف می‌کشاند تا زمانی که در

[1]. high-leverage

مجموعه ۸۰ درصدی داده است. به بیان دیگر، مقادیر R^2 بدست آمده از یک بار بخشی کردن داده‌ها به ترینینگ-تست ناپایدار است. زمانی که مجموعه داده‌ها وسیع‌تر می‌شود مغایرت های بیشتری بین این دو شاخص آشکار می‌شود. به منظور شناسایی نقطه داده‌های دارای قدرت نفوذ بالا این داده‌ها تحلیل شدند و داده‌هایی که میزان صحت یا درستی پاسخ (در BALA) زیر ۰/۵ داشتند حذف شدند. با این وجود، ویژگی‌های زبانی ممکن است داده‌های با قدرت نفوذ بالا باشند (این موضوع قابل انتظار است) و بنابراین بسیاری از آن‌ها بر اساس تعداد یا فراوانی‌شان نرمال یا استاندارد شدند به جای اینکه بر اساس توزیع نرمال استاندارد شوند که می‌تواند ناپایی ایجاد کنند.

در نمودارهای زیر توزیع پراکندگی نمرات y و ŷ (که مقادیر مجذور همبستگی آن‌ها در جدول بالا ذکر شد) ارائه شده است. همانطور که نمودار نشان می‌دهد الگوریتم در پیش‌بینی مقدار واقعی y در تمام طول دامنه‌ی خود به خوبی عمل نمی‌کند زیرا توزیع در امتدادمحور عمودی (ŷ) کوتاه شده است. مواردی که واریانس نسبتاً بیشتری را توضیح می‌دهند (یعنی کتبی/منظم و شفاهی/ اصلاح شده) دارای کمترین محورهای y کوتاه‌شده، مقادیر R2 بالاتر و شیب خط برازش تندتر می‌باشند.

در نمودار زیر توزیع ۵ ویژگی زبانی در بین ۲۰ پیش‌بینی‌کننده برتر برای هر یک از چهار مجموعه داده (شفاهی/معمول، شفاهی/تغییریافته، متن/معمول، و متن/تغییریافته) نشان داده شده است (مواد تشکیل دهنده دستوری، پیچیدگی دستوری، غنا و دامنه واژگان، ویژگی، شباهت، و ابهام، و تأثیر و احساس واژگان). مؤلفه‌های دستوری رایج‌ترین پیش‌بینی کننده‌های برتر بودند. این موضوع برای ۲۰ پیش‌بینی‌کننده برتر برای مجموعه داده‌های معمولی نسبت به مجموعه داده‌های اصلاح‌شده بیشتر صادق است. ویژگی‌های پیچیدگی گرامری جز موارد کم تکرار در بین ۲۰ ویژگی پیش‌بینی کننده ی برتر بودند. غنا و دامنه واژگان نیز صرفاً اندکی بیشتر از پیچیدگی گرامری در ۲۰ پیش‌بینی‌کننده برتر قرار گرفتند و بیشتر در ۲۰ ویژگی برتر مدل های معمولی نسبت به مدل‌های تغییر یافته بودند. ویژگی واژگان، شباهت و ابهام نیز کمتر در ۲۰ مورد ویژگی برتر قرار داشتند. البته استثنائاً در مجموعه داده اصلاح شده کتبی یا متنی، ۷ ویژگی از این نوع به عنوان پیش‌بینی‌کننده‌های برتر شناسایی شدند. عواطف و احساسات درک خواندن را در فرمت تغییر یافته با شدت بیشتری نسبت به ورژن معمولی پیش‌بینی کردند.



نمودار ۱. توزیع بیست ویژگی سینتکتیک و لکسیکال (پیش‌بینی‌کننده‌ها) در چهار مجموعه داده به تفکیک طبقه ویژگی‌ها

بحث و نتیجه‌گیری

هدف پژوهش حاضر شناسایی مهم‌ترین ویژگی‌های نحوی و واژگانی مؤثر بر درک مطلب زبان‌آموزان انگلیسی با تکنیک‌های یادگیری ماشین نظارت شده بود. مطالعه حاضر ممکن است اولین مطالعه‌ای باشد که از ویژگی‌های کاملا جزئی NLP استفاده می‌کند که در پاسخ‌های گرفته شده از گفتار و نوشتار زبان‌آموزان یکسان اند، ازاین‌رو امکان مقایسه‌ی مدل‌سازی بین

شیوه‌های مختلف پاسخ را فراهم می‌کند. روش‌های ML نظارت‌شده برای بررسی این موضوع که آیا ۲۶۰ ویژگی گرامری و واژگانی استخراج شده از NLP می‌توانند درک مطلب را پیش‌بینی کنند، استفاده شد. به منظور بررسی عملکرد بهترین مدل از بین هشت مدل گوناگون، نتایج آن‌ها با میانگین پایه مقایسه شدند و مدل‌های با کمترین خطا در هر یک از چهار مجموعه داده انتخاب شدند. برای هر دو مجموعه داده‌های مبتنی بر شیوه‌های نوشتاری (معمولی و اصلاح‌شده)، جنگل تصادفی بهترین الگوریتم بود. با استفاده از اعتبارسنجی متقاطع، کاهش خطای نسبی برای کتبی/معمولی ۷۵/۱۲ درصد بود اما برای شیوه‌ی گردآوری داده به صورت کتبی/اصلاح‌شده تنها ۴/۲ درصد کاهش خطا وجود داشت. در شیوه‌ی شفاهی/معمولی، مدل ماشین بردار پشتیبان بهترین عملکرد را داشت (کاهش نسبی خطا ۶/۰۷ درصد)، درحالی‌که الگوریتم تقویت‌گرادیان بهترین عملکرد را برای داده‌های شفاهی/اصلاح شده نشان داد و خطا را در حدود ۱۷/۲۱ درصد کاهش داد. الگوی واضح و مشخصی که بتوان بر اساس آن تعیین کرد که کدام نسخه خواندن درک مطلب یا کدام حوزه زبان در بین چهار مدل از نظر کاهش خطای نسبی بهترین بوده است، وجود نداشت، با وجود این در سطح توصیف نتایج می‌توان گفت که شیوه‌های شفاهی/اصلاح‌شده، کتبی/معمولی، شفاهی/معمولی و در نهایت کتبی/معمولی به ترتیب بیشترین کاهش خطا را داشتند. به بیان دیگر، الگوریتم‌های ML در برخی مجموعه داده‌ها موفق‌تر عمل کرده بودند اما روش قابل تفسیری برای موفق عمل کردن آن‌ها وجود ندارد. با این وجود، نتایج تحقیق پایه‌های رابطه‌ی بین زبان شفاهی و درک مطلب خواندن را گسترش داد. به این صورت که واریانس قابل‌توجهی از درک مطلب خواندن را می‌توان از طریق ویژگی‌های واژگانی و گرامری سازنده (و نه ویژگی‌های دریافتی زبان) که از طریق پردازش زبان طبیعی از گفتار و نوشتار دانش‌آموزان استخراج می‌شود، مدل‌سازی کرد.

در مدل‌های نظارت‌شده، کاهش خطای نسبی، مبتنی بر رویکرد اعتبارسنجی متقابل سنجیده شد. همان‌طورکه ذکر شد در این روش داده‌ها به طور مستقل، به دو بخش تقسیم شدند. این عمل به در نظر گرفتن نسبت داده های بخش اول (ترینینگ) به بخش دوم (تستینگ) (۸۰/۲۰) ۲۰۰ بار مستقل از یکدیگر انجام شد. الگوریتم داده‌های آموزشی یعنی رابطه بین درک خواندن و ویژگی‌های گفتاری واژگانی و نحوی در داده‌های آموزشی با ۲۰۰ بار تکرار مستقل فراگرفت و هر بار از این یادگیری برای پیش‌بینی امتیاز درک مطلب داده‌های آزمون استفاده کرد. سپس، نتایج برای دستیابی به میانگین خطای مطلق بین مقادیر پیش‌بینی‌شده و واقعی در داده‌های آزمون میانگین‌گیری شدند. بنابراین، مقادیر پیش‌بینی‌شده در اینجا در موقعیت واقعی پیش‌بینی قرار گرفته بودند- به این معنا که بازده الگوریتم برای «پیش‌بینی» نمرات خواندن داده‌های جدید و دیده نشده سنجیده می‌شد. بنابراین، هنگام مقایسه نتایج با مطالعات رگرسیون سنتی، باید توجه داشت که در اینجا صرفاً «تبیین واریانس» یعنی میزان واریانس در نمره درک مطلب که با ویژگی‌های زبانی توضیح داده می‌شود، گزارش نشده است بلکه چیزی فراتر از آن است. این روش اگرچه اکتشافی است، اما هدف تجزیه و تحلیل تعمیم‌پذیری است و منظور از تعمیم‌پذیری در اینجا یعنی آیا ما می‌توانیم به طور معناداری نمرات خواندن در بخش تستینگ را بر اساس داده‌های ویژگی زبانی بخش ترینینگ پیش‌بینی کنیم. عملکرد هر الگوریتم در تقسیم ترینینگ-تستینگ می‌تواند بیشتر یا کمتر از اعتباریابی متقاطع باشد. این موضوع نشان می‌دهد که حجم نمونه کوچک در بی‌ثباتی (ناپایایی) مدل‌سازی نقش دارد: مشاهدات با قدرت نفوذ بالا و دور از دسترس می‌توانند به شدت بر قدرت پیش‌بینی مدل تأثیر بگذارند. در مطالعات آینده با مجموعه داده‌های بزرگتر، ممکن است مدل ها موفق‌تر عمل نمایند (پیشنهاد می‌شود در مطالعات آینده از حجم نمونه بالاتر استفاده شود). با این وجود، حتی با نمونه کوچک، قدرت پیش‌بینی و مقدار واریانس توضیح داده شده توسط این مدل‌های ML با شیوه‌های موجود رقابت می‌کند: محدوده واریانس توضیح داده شده در چهار مدل مختلف در نمونه‌های ترینینگ-تستینگ (R^2) بین ۳/۳۶-۵/۱۸ درصد بود. ازاین‌رو در بحث تحقیق بر آن شدیم تا نتایج ادبیات و پیشینه‌ی منتشر شده در مورد مدل‌سازی رابطه بین درک مطلب و نحو یا واژگان زبان یا هر دو را ارائه دهیم. در تمام این تحقیقات متغیر درک مطلب یک متغیر پیوسته (مانند تحقیق حاضر) بوده است و مطالعاتی که در آن‌ها متغیر درک مطلب طبقه‌ای درنظر گرفته شده بود و از تکنیک‌های خانواده‌ی ANOVA یا طرح فاکتوریل برای مقایسه نتایج شان استفاده کرده بودند (به عنوان مثال، نیشن و اسنولینگ، ۲۰۰۴) ذکر نشده‌اند. مطالعات بررسی شده از ارزش های پیش‌بینی مختلفی برای ویژگی‌های واژگانی و نحوی در پیش‌بینی درک مطلب استفاده کرده‌اند. هیچ مطالعه‌ی که دقیقاً با مطالعه‌ی حاضر

یکسان باشد، وجود نداشت. زمانی که از اعتباریابی متقابل استفاده کردیم، اصلاح نسبت به میانگین پایه بین ۴۰/۲ تا ۱۷/۲۱٪ متغییر بود و هنگامی که از بخش‌بندی ترینینگ-تستینگ در داده‌های استفاده شد (برای هر یک از چهار مدل)، واریانس توضیح داده شده (R^2) بین ۳۱-۱۹٪ بود که نتایج بدست آمده‌ی تحقیق حاضر در هر دو مورد در محدوده مطالعات بررسی شده بود. دانش نحوی و واژگانی را مستقل نمی‌توان مستقل از یکدیگر فرض کرد (اگر آن‌ها را مستقل فرض کنیم مجموع واریانسی که از درک مطلب تبیین می‌شود ۲۵ درصد خواهد بود) ولی در عین حال ۲۱ درصد واریانس درک مطلب خواندن را تبیین می‌کنند. البته نیاز است که تأکید شود که ملاحظات اندازه‌ی نمونه در این برداشت باید مورد نظر قرار گیرد. با این حال، تفاوت مهم بین مطالعه حاضر و مطالعات دیگر در این است که مطالعه حاضر به طور صریح واژگان یا دستور زبان شرکت‌کنندگان را با استفاده از معیارهای امتیازدهی شده ارزیابی نکرده است، بلکه از مجموعه داده متنی مربوط به گفتار و نوشتار دانش‌آموزان ویژگی‌های زبانی را استخراج کرده است. به بیان دیگر الگوریتم‌ها در اینجا خود به کدگذاری و استخراج نتایج پرداخته‌اند ولی در شیوه‌های سنتی معیارها توسط پژوهشگران ایجاد و برای آن‌ها داده جمع‌آوری می‌شود. در بیشتر این شیوه‌ها داده‌ها از طریق پرسشنامه استخراج شده است و دانش‌آموزان یا شرکت‌کنندگان لازم است از بین پاسخ‌ها یک پاسخ را انتخاب نمایند، به عبارتی شیوه‌ی آزمون‌گیری چند گزینه‌ای بوده است.

علاوه بر این در پژوهش حاضر بر خلاف مطالعاتی مانند گوتاردو و همکاران (۱۹۹۶) و تونمر و چپمن (۲۰۱۲)، هیچ عامل زبانی دیگری در مدل‌سازی ثابت در نظر گرفته نشده است (در سایر تحقیقات عامل‌هایی مانند درک زبان، حافظه‌کاری کلامی ثابت درنظر گرفته شده‌اند). پژوهشگران همچنین به دنبال شناسایی مهم‌ترین ویژگی‌های واژگانی و نحوی پیش‌بینی‌کننده‌های درک خواندن در چهار مدل گوناگون ML بودند. جایگشت تصادفی با استفاده از رویکرد اعتبارسنجی متقابل از طریق کتابخانه‌ی MLJ (بلائوم و همکاران، ۲۰۲۰) در جولیا انجام شد.

مدل‌ها (همان‌طورکه در بالا توضیح داده شد) با استفاده از ۲۰۰ بار بخش بندی ترینینگ/تستینگ با هدف اعتبارسنجی متقابل اجرا شدند. در هر اجرا اعتبارسنجی متقابل، یک متغیر به طور تصادفی جایگشتی حذف می‌شد تا اثر حذف آن بر مدل مشخص شود. این شیوه از منطقی مانند روش حذف-سؤال در روش‌های کلاسیک اندازه‌گیری بهره می‌برد. این فرآیند برای تمام متغیرها در مجموعه داده‌ها تکرار شد. از میانگین افزایش میانگین خطای مطلق به عنوان شاخصی برای اندازه‌گیری سهم هر متغیر در مدل استفاده شد. بیست ویژگی پیش‌بینی‌کننده برتر برای هر یک از چهار مدل (شفاهی/متن با معمولی/اصلاح شده) مشخص شد. به طور کلی، در مجموعه داده‌های معمولی در مقایسه با مجموعه داده‌های اصلاح‌شده اجزای گرامری معمولاً جزء پیش‌بینی‌کننده‌های برتر بودند. همچنین مشخص شد ویژگی‌های مربوط به دستور زبان می‌تواند مهارت‌های درک مطلب را برای خوانندگان ماهرتر بهتر از خوانندگان کم مهارت پیش‌بینی کند. با این حال، پیچیدگی دستوری (نسبت به اجزای سازنده‌ی گرامری) در هیچ یک از مدل‌های نظارت شده متغییر پیش‌بینی‌کننده مهمی نبود. بر خلاف نتایج بدست آمده توسط واز، تیکسرا، گنکالوز (۲۰۲۲) مبنی بر این که گرامر آن حوزه زبانی است که بیشترین ارتباط را با نتایج درک مطلب در دانش‌آموزان پایه های ۲ و ۴ دارد، در این تحقیق هیچ تفاوتی بین نمرات پایه‌های مختلف وجود نداشت. یافته‌ها نشان دادند که ویژگی‌های دستوری در مدل‌های معمولی نسبت به مدل‌های اصلاح‌شده اهمیت بیشتری دارند. این نتیجه با یافته‌های سینکلایر (۲۰۲۰) که بر اساس تحقیق آن‌ها آگاهی نحوی پیش‌بینی‌کنندگی قابل‌توجهی برای درک مطلب دارد هماهنگ است. اجزای گرامری معمولاً جزء پیش‌بینی‌کننده‌های برتر مجموعه داده‌های شفاهی نسبت به مجموعه داده‌های استخراج‌شده متنی برای BALA اصلاح‌شده و معمولی بودند. این یافته نشان می‌دهد ماهیت گرامر در گفتار ممکن است پیش‌بینی‌کننده قوی‌تری نسبت به دستور زبان در نوشتار باشد. مطالعاتی که در بالا ذکر شد دستور زبان را با استفاده از تصحیح دستوری، ترتیب کلمات، دانش نحوی و یا قضاوت دستوری اندازه‌گیری کرده بودند و در این سنجش عمدتاً از معیارهای شفاهی (به جای معیارهای نوشتاری) استفاده کرده بودند (به عنوان مثال، اوکلی، کین و البرو، ۲۰۱۴؛ الیس و روور، ۲۰۲۱). همچنین این از طریق پاسخ‌های از قبل تعیین شده (مثل آزمون‌های چند گزینه‌ای) انتخاب شده‌اند. ازاین‌رو، مقایسه‌ای که ما بین مطالعه‌ی حاضر و سایر تحقیقات انجام می‌دهیم یک مقایسه‌ی مستقیم نیست. ازاین‌رو اهمیت نسبتاً بالاتر ویژگی‌های دستوری در داده‌های شفاهی در مقایسه با داده‌های نوشتاری به طریق اولی قابل مقایسه نیست. بنابراین ماهیت گرامر در گفتار و نوشتار و ارتباط آن با سایر

مهارت‌های سوادآموزی، می‌تواند پیشنهادی برای تحقیقات آتی باشد. الگوی غنای واژگان و متغیرهای دامنه مانند ویژگی‌های دستوری، بیشتر جزء ویژگی‌های برتر داده‌های معمولی نسبت به داده‌های اصلاح‌شده بودند که می‌تواند مربوط به ماهیت اندازه‌های ذهنی یا سوبژکتیو کلمات باشد. با این حال، بر خلاف ویژگی‌های دستوری، در داده‌های استخراج‌شده از متن، غنای واژگان و دامنه کلمات نسبت به داده‌های استخراج‌شده شفاهی با فراوانی بیشتری جزء ویژگی‌های برتر بودند. این موضوع نشان می‌دهد که پیچیدگی واژگان در نوشتار ممکن است به شدت با درک مطلب مرتبط باشد تا پیچیدگی واژگان در گفتار. با این حال، این یافته می‌تواند به ماهیت تکلیف نیز مربوط باشد، زیرا واژگان در تکالیفی که مبتنی بر استنباط شفاهی است ممکن است فرصت کافی برای استفاده از واژگان پیچیده را فراهم نکرده باشد.

احساسات و عواطف واژگانی در مجموعه داده‌های اصلاح‌شده (مخصوصاً اصلاح‌شده شفاهی) بیشتر از مجموعه داده‌های معمولی جزء ویژگی‌های برتر پیش‌بینی درک مطلب به شمار می‌روند. اختصاصیت، شباهت و ابهام واژگان در پیش‌بینی‌های برتر مدل کتبی/اصلاح‌شده دو برابر بیشتر از سایر مدل‌ها بود. به طور کلی، این الگوها نشان می‌دهند که برای خوانندگان با مهارت کمتر که عموماً خوانندگان جوان‌ترند عواطف و احساسات در زبان سازنده‌شان بیشتر از واژگان یا اجزای دستوری که استفاده می‌کنند، پیش‌بینی‌کننده موفقیت آن‌ها در خواندن است. با این حال، این یافته ممکن است قابل‌تعمیم نباشد زیرا ماهیت احساسات و تأثیر بر پیش‌بینی درک خواندن در این داده‌ها ممکن است با چیزی که در سایر کارها پیش‌بینی می‌شود، معادل نباشد. این یافته مستلزم بررسی بیشتر است.

درخصوص محدودیت‌های پیش‌روی مطالعه، این احتمال وجود دارد که سؤالات باز پاسخ درک مطلب که در این تحقیق استفاده شده است، کاملاً مستقل از معیارهای مربوط به نمره‌دهی درک مطلب خواندن نباشند. با این حال تلاش شد که عوامل متعددی را با هم ترکیب گردد تا از نگرانی‌ها در این مورد کاسته شود. اول اینکه، هیچ یک از پاسخ‌های باز از نظر محتوا یا «صحت» ارزیابی نشدند. در واقع، سه نوع از چهار نوع پاسخ خواندن باز مورد استفاده در اینجا – پیش‌بینی‌های باز درباره آنچه خوانده می‌شود، توصیف علاقه به متن، ایجاد سه سؤال در مورد آنچه خوانده شده است – پاسخ‌های صحیح یا نادرست ندارند. سؤال آخر یعنی سؤال مربوط به سطح بالای درک مطلب کاملاً باز است اما به شواهد متنی نیاز دارد. با این حال، مانند سه سؤال قبلی صحت محتوا در مطالعه حاضر به طور خاص ارزیابی نشده است. درحالی‌که درک کمتر متن ممکن است منجر به پاسخ‌هایی با پیچیده کمتر به این چهار سؤال منجر شود، عدم درک کامل متن پس از خواندن مانع ایجاد پاسخ پیچیده واژگانی و نحوی نمی‌شود. صرف نظر از آنچه گفته شد چشم‌انداز پژوهشی بنا به رسوب دانشی که از این انجام این مطالعه حاصل شده است این است که ارزیابی به‌موقع با استفاده از این ابزارها می‌تواند به معلمان و دانش‌آموزان کمک کند تا آن‌ها به صورت مشخص بدانند روی کدام وجه زبان‌آموزی و مهارت‌های زبانی تمرکز کنند تا از یادگیری کامل زبان و سواد حداکثر استفاده را داشته باشند. می توان سناریوی آموزشی را در آینده نزدیک متصور شد که در آن این فناوری‌ها ارزیابی تشخیصی معتبر، جامع، کارآمد و به موقع از مهارت‌های زبانی و سوادآموزی دانش‌آموزان را در اندازه مطلوب و مناسب هم برای معلمان و هم برای دانش‌آموزان امکان‌پذیر می‌کند. پیشنهاد می‌شود در مطالعات آتی الگوریتم‌های معتبر ML را برای ارائه بازخورد به زبان‌آموزان و معلمان‌شان مورد ارتقاء زبان گفتاری و نوشتاری توسعه داده شود، همچنین پیشنهاد می‌شود تمرکز مطالعات آتی توسعه یک سیستم سنجش و ارزیابی کم هزینه، مؤثر و کم خطر باشد تا جایگزین سیستم‌های آموزشی نسبتاً ناکارآمد فعلی در حوزه زبان‌آموزی شود.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش: این مقاله برگرفته از رساله دکتری نویسنده اول در رشته روانشناسی تربیتی در دانشکده روانشناسی دانشگاه علوم و تحقیقات است. به جهت حفظ رعایت اصول اخلاقی در این پژوهش سعی شد تا جمع‌آوری اطلاعات پس از جلب رضایت شرکت‌کنندگان انجام شود. همچنین به شرکت‌کنندگان درباره رازداری در حفظ اطلاعات شخصی و ارائه نتایج بدون قید نام و مشخصات شناسنامه افراد، اطمینان داده شد.

حامی مالی: این پژوهش در قالب رساله دکتری و بدون حمایت مالی می‌باشد.

نقش هر یک از نویسندگان: این مقاله از رساله دکتری نویسنده اول و به راهنمایی نویسنده دوم و مشاوره نویسندگان سوم و چهارم استخراج شده است.

تضاد منافع: نویسندگان همچنین اعلام می‌دارند که در نتایج این پژوهش هیچ‌گونه تضاد منافعی وجود ندارد.

تشکر و قدردانی: بدین‌وسیله از اساتید راهنما و مشاوران این تحقیق که در این پژوهش شرکت کردند، تشکر و قدردانی می‌گردد.

References

Benfatto, M. N., Seimyr, G. Ö., Ygge, J., Pansell, T., Rydberg, A., & Jacobson, C. (2016). Screening for dyslexia using eye tracking during reading. *PloS One*, 11(12), 1-16. https://doi.org/10.6084/m9.figshare.c.3521379.v1

Black, M., Tepperman, J., Lee, S., & Narayanan, S. S. (2008). Estimation of children's reading ability by fusion of automatic pronunciation verification and fluency detection. In *Ninth Annual Conference of the International Speech Communication Association*. https://sail.usc.edu/publications/files/i08_2779

Blaom et al., (2020). MLJ: A Julia package for composable machine learning. Journal of Open Source Software, 5(55), 2704, https://doi.org/10.21105/joss.02704

Brimo, D., Apel, K., & Fountain, T. (2017). Examining the contributions of syntactic awareness and syntactic knowledge to reading comprehension. *Journal of Research in Reading*, 40(1), 57-74. https://doi.org/10.1111/1467-9817.12050

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. AI Magazine, 25(3), 27-27. https://onlinelibrary.wiley.com/doi/pdf/10.1609/aimag.v25i3.1774

Century skills. *Annual review of psychology*, *73*, 547-574. https://DOI:10.1111/bjet.13342

Ellis, R., & Roever, C. (2021). The measurement of implicit and explicit knowledge. *The Language Learning Journal*, *49*(2), 160-175. DOI:10.1080/09571736.2018.1504229

Evanini, K., Heilman, M., Wang, X., & Blanchard, D. (2015). Automated scoring for the TOEFL Junior® comprehensive writing and speaking test. ETS Research Report Series, 2015(1), 1-11. https://doi.org/10.1002/ets2.12052

Gómez-Merino, N., Fajardo, I., & Ferrer, A. (2021). Did the three little pigs frighten the wolf? How deaf readers use lexical and syntactic cues to comprehend sentences. *Research in Developmental Disabilities*, *112*, 103908. https://doi.org/10.1016/j.ridd.2021.103908

Graesser, A. C., Sabatini, J. P., & Li, H. (2022). Educational psychology is evolving to accommodate technology, multiple disciplines, and twenty-first-century skills. *Annual Review of Psychology*, *73*, 547–574. https://doi.org/10.1146/annurev-psych-020821-113042

Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing: An Interdisciplinary Journal*, *16*(6), 505–539. https://doi.org/10.1023/A:1025521722900

Komeili, M., Pou-Prom, C., Liaqat, D., Fraser, K. C., Yancheva, M., & Rudzicz, F. (2019). Talk2Me: Automated linguistic data collection for personal assessment. *PLoS ONE*, *14*(3), Article e0212342. https://doi.org/10.1371/journal.pone.0212342

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of research in reading*, *27*(4), 342-356. https://doi.org/10.1111/j.1467-9817.2004.00238.x

Oakhill, J., Cain, K., & Elbro, C. (2014). *Understanding and teaching reading comprehension: A handbook*. Routledge.

Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices-past, present, and future. *Children's reading comprehension and assessment*, *2*. https://doi.org/10.1177/1534508417728685

Poulsen, M., & Gravgaard, A. K. (2016). Who did what to whom? The relationship between syntactic aspects of sentence comprehension and text comprehension. *Scientific Studies of Reading*, *20*(4), 325-338. https://core.ac.uk/download/pdf/269272114.pdf

Sinclair, J. (2020). *Using machine learning to predict children's reading comprehension from lexical and syntactic features extracted from spoken and written language*. University of Toronto (Canada).

Truckenmiller, A., Shen, M., & Sweet, L. E. (2021). The Role of Vocabulary and Syntax in Informational Written Composition in Middle School. *Reading and Writing*, *34*(4), 911-943. https://doi.org/10.1007/s11145-020-10099-1

Vaz, P. M. F., Teixeira, C., & Gonçalves, V. (2022). Identification of students at risk in reading, writing and grammar: a study with students from 3rd and 4th grades in northern portugal. In *16th annual International Technology, Education and Development Conference* (Vol. 1, pp. 7943-7951). IATED http://dx.doi.org/10.21125/inted.2022.2001

Yapp, D., de Graaff, R., & van den Bergh, H. (2023). Effects of reading strategy instruction in English as a second language on students' academic reading comprehension. Language Teaching Research, 27(6), 1456-1479. https://doi.org/10.1177/1362168820985236

Young, S. (2022). *Fidelity of Implementation of a Balanced Literacy Program in the Elementary Classroom* (Doctoral dissertation, Walden University).