Evaluation of Explainable Artificial Intelligence methods in Language Learning Classification of Spanish Tertiary Education Students

Grigorios Tzionis[1], Gerasimos Antzoulatos[1], Periklis Papaioannou[1], Athanasios Mavropoulos[1], Ilias Gialampoukidis[1], Marta González Burgos[2], Stefanos Vrochidis[1], Ioannis Kompatsiaris[1], Maro Vlachopoulou[3]

[1]Centre For Research & Technology Hellas; [2]Metodo Estudios Consultores, [3]University of Macedonia

Abstract. With the increasing prevalence of AI, significant advancements have been made across various domains, such as healthcare, learning, industry, etc. However, challenges persist in terms of trusting and comprehending the outcomes generated by these technologies. Specifically in the language learning domain, teachers face challenges regarding the classification of the students' learning capabilities and build the appropriate learning path for them. To address these challenges, the concept of Explainable Artificial Intelligence (XAI) was adopted, which is a set of processes and methods that allows human users to interpret, understand and trust the results derived from machine learning models. In this study, we adopt two well-known XAI algorithms, PFI and SHAP in a proposed Knowledge Generation Model equipped with ML models to derive hidden knowledge. The whole framework has been applied and evaluated on the Language Learning Classification of Spanish Tertiary Education Students acquired from the CEDEL2 database. The analysis concludes that in terms of explaining the black-box models, the SHAP model-agnostic method is the most comprehensive and dominant for visualizing feature interactions and feature importance and be applicable to any type of data.

Keywords: Language Learning, Machine Learning, Explainable Artificial Intelligence, Interpretability, Comparative Analysis.

1 Introduction

Language Learning (LL) is a fundamental aspect of human development, enabling individuals to communicate, express themselves, and understand the world around them [1]. It is a complex and ongoing process that begins at birth and continues throughout life. Through interactions with others, individuals acquire linguistic skills, comprehend grammar and vocabulary, and develop the ability to communicate effectively [2]. Linguistic scientists have devoted extensive efforts to study the mechanisms and patterns underlying language acquisition, seeking to uncover the cognitive processes involved and the factors influencing language learning outcomes [1].

On the other hand, in recent years, the field of Artificial Intelligence (AI) has experienced remarkable advancements, revolutionizing various domains with its predictive

and analytical capabilities [3]. It has found applications in diverse domains, ranging from natural language processing to computer vision and speech recognition [4] [5]. Simultaneously, the field of Machine Learning (ML) has witnessed tremendous growth and popularity in the realm of Artificial Intelligence (AI) [6]. ML techniques enable the development of predictive and descriptive models that can analyze vast amounts of data, extract patterns, simulate and understand complex systems and human behaviour as well as make informed decisions [4] [5] [7].

However, the black-box nature of many ML models has raised concerns about their transparency and interpretability [7]. As ML algorithms become more powerful and pervasive, the need for Explainable AI (XAI) has become paramount. XAI encompasses a set of processes and methods that enable human users to comprehend and trust the results and outputs generated by ML algorithms [8]. It aims to provide insights into how an AI model operates, its anticipated impact, and potential biases. By promoting model accuracy, fairness, transparency, and interpretability, XAI contributes to more informed AI-powered decision-making [9].

The interwind between Language Learning and Machine Learning has attracted significant attention from researchers and practitioners, as both fields share common goals of understanding and modeling human behavior [10]. ML techniques can stress language learning processes by providing personalized learning experiences, automated language assessment, and intelligent tutoring systems [8] [9].

In this paper, our research objective is to explore the application of XAI methodologies in the context of language learning. Specifically, we employ the proposed interactive and iterative Visual Analytics framework [11] along with the two well-known XAI methods, namely the Permutation Feature Importance (PFI) [12] and SHapley Additive exPlanations (SHAP) [13]. These techniques offer interpretability by assigning importance scores to features and providing explanations for individual predictions. The goal is to classify Spanish Tertiary Education Learners in terms of their performance relied on general characteristics and their learning profile using machine learning techniques and discover knowledge and patterns that are hidden in the data. Moreover, the application of XAI algorithms to the results of the classification problem aims to evaluate them and compare them in terms of which one of them provides better insights to teachers and practitioners.

2 Language Learning: Processes and Challenges

In recent decades, language learning has garnered significant attention among linguistic scientists [14]. It is widely recognized as an active and continuous process that begins at birth and persists throughout life [15]. It is an ongoing endeavor that encompasses various stages of development, from early childhood to adulthood [16]. Additionally, language learning plays a crucial role in establishing relationships with family members and friends, while also aiding in the comprehension and organization of the world [15]. Language learners of all ages actively engage in acquiring new vocabulary, mastering grammatical structures, and refining their communicative skills [16]. Linguistic scientists worldwide are dedicated to deriving valuable conclusions

and correlations from their research in this domain [17]. Researchers in linguistics strive to unravel the complexities of language acquisition, drawing valuable insights from their investigations [18].

Language learning presents several challenges that vary across individuals and contexts [19]. Learners may encounter difficulties in pronunciation, grammar, vocabulary acquisition, or cultural adaptation [20]. Factors such as age, motivation, learning environment, and exposure to the target language influence the learning process [2]. Understanding these challenges and developing effective instructional strategies are essential for facilitating successful language learning outcomes [16]. Research in this domain aims to uncover the underlying mechanisms of language acquisition, identify effective instructional methods, and address the specific needs of diverse learner populations [16]. By examining the challenges faced by language learners, researchers can contribute to the development of evidence-based pedagogical approaches, language assessment tools, and interventions to support language learning across different contexts [16].

3 Enhancing Language Learning with Machine Learning

Language learning is a complex and dynamic process that involves acquiring linguistic skills, comprehending grammar and vocabulary, and developing effective communication abilities [21]. One key area where ML has made significant contributions to language learning is in the development of intelligent tutoring systems [22]. These systems utilize ML algorithms to understand the unique learning needs and preferences of individual learners, enabling the delivery of personalized instruction and feedback [22]. By analyzing learner data and performance patterns, ML models can adapt the learning content and pace to optimize learning outcomes [22].

ML algorithms also play a crucial role in automated language assessment, providing objective and efficient evaluation of learners' language proficiency [23]. Natural language processing techniques enable the automatic scoring and analysis of learners' written and spoken responses, providing detailed feedback on grammar, vocabulary usage, and overall language proficiency [23]. This automated assessment process saves time for educators and allows learners to receive immediate feedback, enhancing the learning experience [24].

As ML models become increasingly complex and powerful, the need for interpretability has gained significant attention [25]. Interpretability refers to the ability to understand and explain the decisions and outputs of ML models [26]. It helps build trust and confidence in the predictions made by these models, especially in critical domains such as manufacturing, healthcare and finance [25] [27] [28] [29]. Explainable AI (XAI) has emerged as a field that focuses on developing processes and methods to enhance the interpretability of ML models [25] [30]. XAI enables users to comprehend the inner workings of ML algorithms, understand the reasoning behind their decisions, identify potential biases, and assess the reliability and fairness of the model's outcomes [31]. XAI aims to bridge the gap between the black-box nature of complex ML models and the need for transparency and accountability in AI-powered decision-making systems [25].

In conclusion, the growing complexity of ML models, the need for interpretability and Explainable AI has become crucial [26]. Researchers and practitioners are actively working towards developing methods and techniques to enhance the transparency, interpretability, and trustworthiness of ML models, ultimately leading to more responsible and reliable AI systems [30].

4 Proposed Methodology

In this paper, we employ and extend the proposed Knowledge Generation Model (KGM) (**Fig. 1**) for language learning [11] by consolidating advanced Machine Learning techniques, to deal with challenges in the language learning domain, along with XAI approaches to provide more interpretable and reliable findings and results.

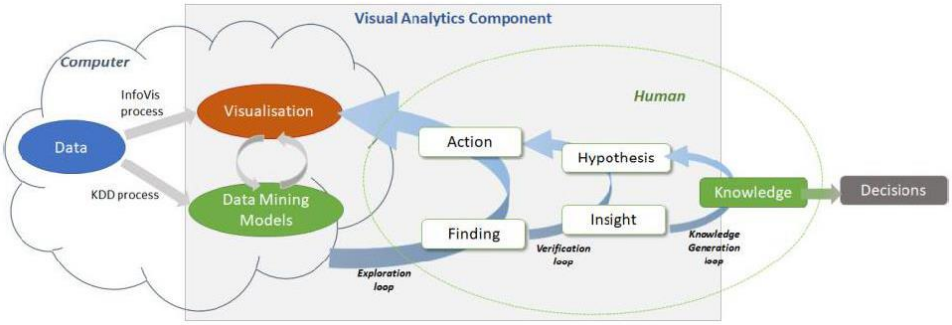


Fig. 1. Knowledge generation model (KGM) for language learning.

Specifically, the interactive and iterative Visual Analytics schema fosters complex decision-making processes by leveraging two main pipelines of processing data, namely from raw Data to Visualisation (InfoVis process) or Data Mining modeling through the Knowledge Discovery in Databases (KDD) processes [32], coupling with machine learning algorithms (**Fig. 1**). In this work, we focus on the process and analysis of obtained learning data as well as the outcomes interpretation by applying XAI algorithms, and particularly the Permutation Feature Importance (PFI) [12] and SHapley Additive exPlanations (SHAP) [13]. Therefore, the importance of the selected features/attributes that utilized in the machine learning models will be discovered providing useful findings to the teachers and other stakeholders. Those findings, which are interesting observations derived from the data mining models and XAIs, lead to further interaction between the human-analyst (in our case the teachers) and the Visual Analytics Component (system) by following the Exploration loop (**Fig. 1**) [33]. Furthermore, the findings can be interpreted by experts (teachers) using their intrinsic previous knowledge in the context of the problem domain (language learning) and hence, new insights are emerged, followed by new hypotheses that should be analysed and evaluated (Verification loop) (**Fig. 1**). Finally, these iteration loops will conclude to the generation of new knowledge by the Knowledge Generation Loop (**Fig. 1**). The knowledge generation concerns the verification of hypotheses and exist-

ing assumptions based on the revealed evidence. This evidence-based approach permits teachers and analysts to trust hypotheses leading to gained knowledge and generate new ones [34]. Otherwise, they should discard the hypotheses and return to the exploration of new, undiscovered correlations in the data. Assessing the trustworthiness of new knowledge depends on the collected evidence and requires a critical review of the overall KDD analysis process starting from data gathering. In more depth the steps of the KDD [32] are following:

1. *Selection* - during this step a target data set is created, by focusing on a subset of attributes or data samples that require further exploration and analysis [35].
2. *Pre-processing* - the selected dataset undergoes pre-processing to obtain consistent data. Potential actions include handling missing values, detecting outliers and extreme values, feature scaling and normalization. For example, extreme values out of rational interval for the age attribute are being detected and eliminated from further analysis. Moreover, some data mining methods work well when the attributes are in the same scale. Hence, methods for normalisation such as z-score, min-max scaling have been adopted [36].
3. *Transformation* - by utilising dimension reduction processes the pre-processed dataset is transformed for further analysis. Automated processes to transform the pre-processed data into a compatible format to be further analysed by data mining techniques have been deployed [37].
4. *Modeling* - involves the development of Machine Learning methods and techniques for extracting or discovering previously unknown, interesting patterns or trends in a particular representational form that depends on the Data Mining goals (e.g. prediction or classification). In this work, we have applied machine learning methods to deal with the classification problem in language learning as described in the following section.
5. *Evaluation* - this step is very important as the generalisation capabilities of the trained models are assessed. The developed machine learning models tested in terms of their performance against specific validation measures, such as accuracy, F1-score etc. After the fine tuning of the parameters the best trained ML model is selected [38].
6. *Interpretation* – this step concerns the application of the two well-known explainability algorithms (PFI and SHAP) and interpret the results.

5 Analysis & Results

In this study, seven different machine learning methods were utilized, namely Logistic Regression (LR), K-Nearest Neighbors (K-NN), Linear Discriminant Analysis (LDA), Decision Tree (CART), Support Vector Machine (SVM), Random Forest (RF), and Bagging (BG) to classify Spanish Tertiary Education Learners in terms of their Placement Test Score. Specifically, data from the CEDEL2[1] database was employed.

[1] CEDEL2 stands for Corpus Escrito del Español como L2 (L2 Spanish Written Corpus).

CEDEL2 contains data from learners of Spanish at all proficiency levels (beginner, intermediate, advanced) and different L1 (means learners' mother tongue) and 'L2' (means learners' foreign language) backgrounds. Initially, contains 3034 registrations but after preprocessing the remained registrations decreased to 1473 records. We selected six (6) features from the dataset contained in the database due to the fact that they appear to be most relevant to the language learning problem. These features are *Sex (Gender), Age, Mothers native language, Languages spoken at home, Years studying Spanish, and Additional Foreign Languages*. As target variable can be employed the *Placement Test Score (%)* feature, which has classified into three distinct classes which are the following: *Class '0'* below 48% (142 regs); *Class '1'* from 48% to 81.9% (572 regs) and finally, *Class '2'* from 82% to 100% (759 regs).

To improve the accuracy of each classifier the grid search approach was applied to investigate which are the optimal hyperparameter set for each classifier. We conducted a series of experiments using the aforementioned classifiers over the above dataset and their performance has been assessed in terms of the evaluation metrics (avg. Accuracy). The K-NN and SVM classifiers exhibit best generalization capabilities as they achieved the highest accuracy over the testing set compared to the others, around 67.11% (**Fig. 2**).

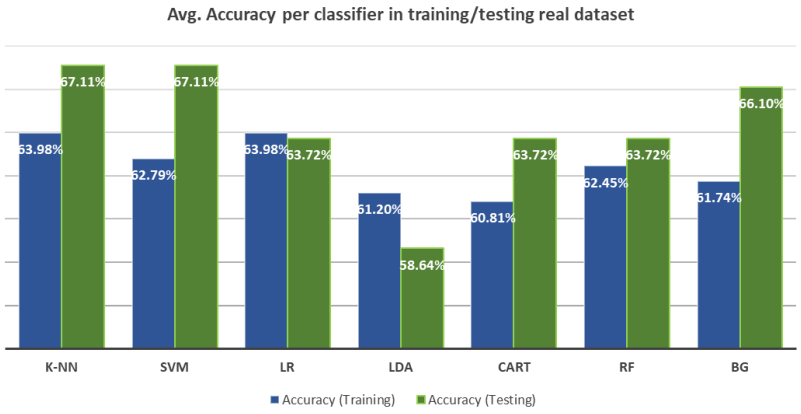


Fig. 2. Performance (avg. Accuracy) per classifier over the training/testing CEDEL2 dataset

5.1 Interpretation

Feature importance refers to techniques that assign a score to input features based on how useful they are at predicting a target variable. Additionally, feature importance scores provide insight into the data and the models, improving the efficiency and effectiveness of the predictive models.

Permutation feature importance (PFI) is a technique for calculating relative importance scores that is independent of the model used [12]. First, a model is fit on the dataset, such as a model that does not support native feature importance scores. Then the model is used to make predictions on a dataset, although the values of a feature (column) in the dataset are scrambled. This is repeated for each feature in the dataset. Then, this whole process is repeated a number of times. The result is calculated as a

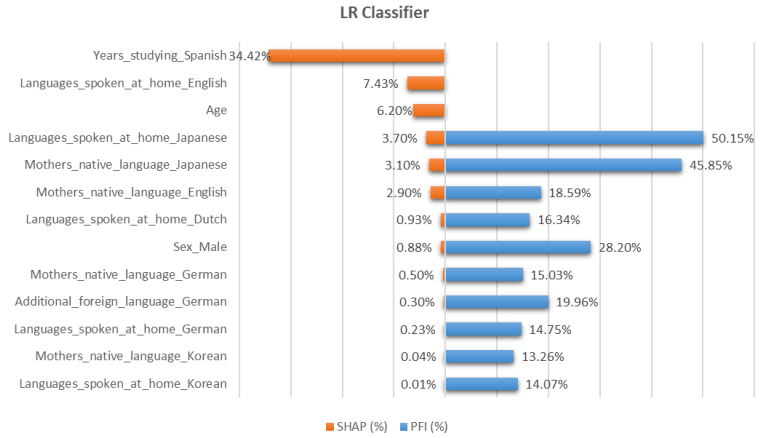
mean of importance score for each input feature (and distribution of scores given the repeats). This approach can be used for regression or classification and requires that a performance metric be chosen as the basis of the importance score, such as the mean squared error for regression and accuracy for classification.

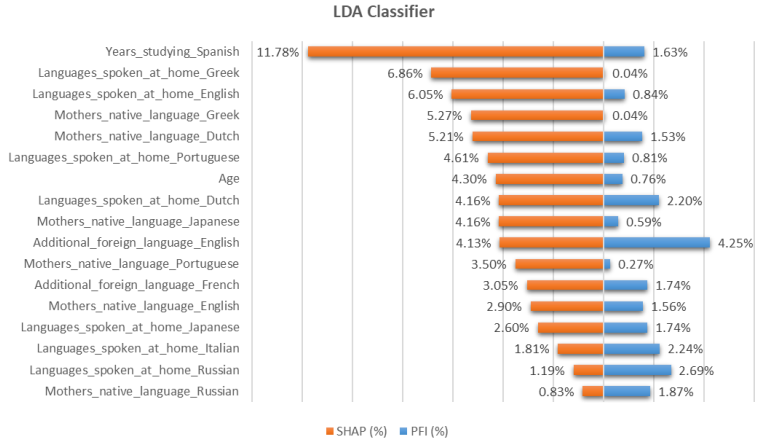
In terms of explaining any black-box model, the ***SHapley Additive exPlanations (SHAP)*** method is, by far, the most comprehensive and dominant across the literature methods for visualizing feature interactions and feature importance [13]. The SHAP methods are not only model-agnostic, but they have been demonstrated to be applicable to any type of data. The SHAP values represent the contribution of each feature to the prediction of each individual instance. To compute global feature importance based on SHAP values, the mean absolute SHAP value for each feature across all instances should be estimated.

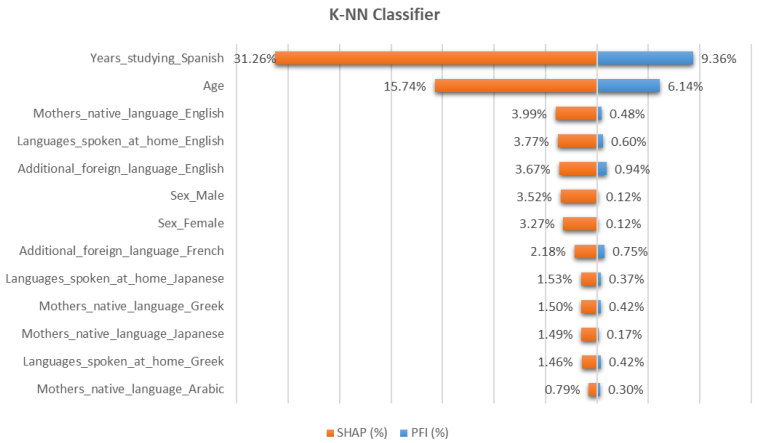
In **Fig. 3** comparisons between PFI and SHAP values (in %) are illustrated per classifier over the CEDEL2 dataset. It should be noted that the attributes have been posed in descending order according to their SHAP values. The findings reveal that the *'Years studying Spanish'* and *'Age'* consistently emerge as the most influential features across Decision Trees (CART), Random Forest, Bagging and kernel (SVM) classifiers, as their SHAP and PFI values are significantly higher compared to the values of other attributes. For those classifiers, the ordering of importance of the attributes is quite similar comparing the SHAP and PFI values. However, the magnitude of the contribution of these features varies depending on the model used, indicating the intrinsic differences between the machine learning algorithms. Slight differences were exhibited for the categories of the attributes that are few representatives in the dataset such as Languages spoken at home Japanese or Mothers native language Portuguese etc.

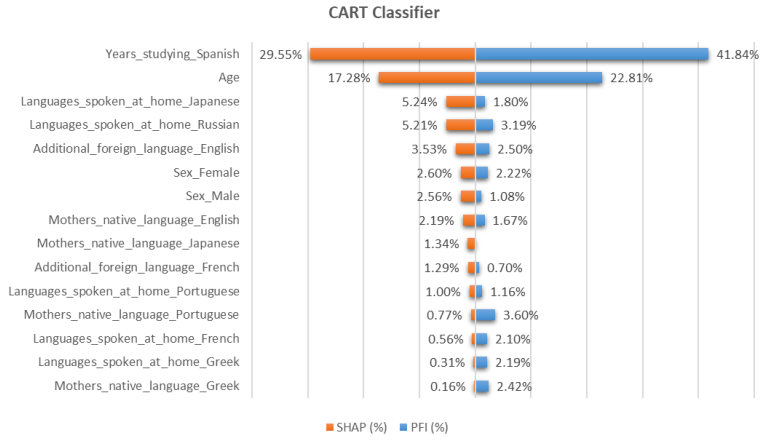
However, the behavior of XAIs algorithms using the linear classifiers LR and LDA is quite different. Features such as *'Languages spoken at home: Japanese'* and *'Mothers native language: Japanese'* also show considerable importance, especially in the Logistic Regression model, suggesting potential context-specific relevance of these features. Nevertheless, these features don't appear to be universally significant across all models. Another potential explanation for this could be the fact that non-linear nature of the problem. Hence, those classifiers could not fit adequately into the dataset and their performance is low. In conclusion, this analysis underscores the importance of understanding and interpreting feature importance in model development and highlights the variability that can exist depending on the algorithm and interpretation method utilized.

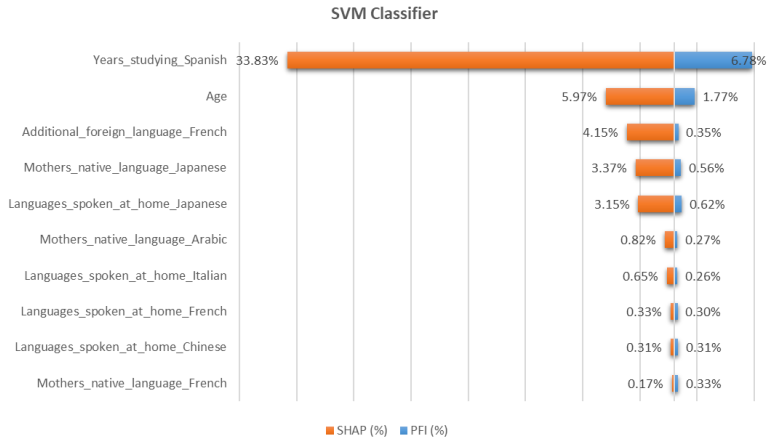
On the other hand, tree and kernel-based classifiers, such as CART, SVM, RF, etc. can adequately capture the non-linear nature of the data, hence the PFI and SHAP can better estimate and hierarchy of the impact of the features and also exhibit a constant behavior. Hence, for those models, the attributes '*Years studying Spanish'* and '*Age*' play a significant role in correctly classifying the students.

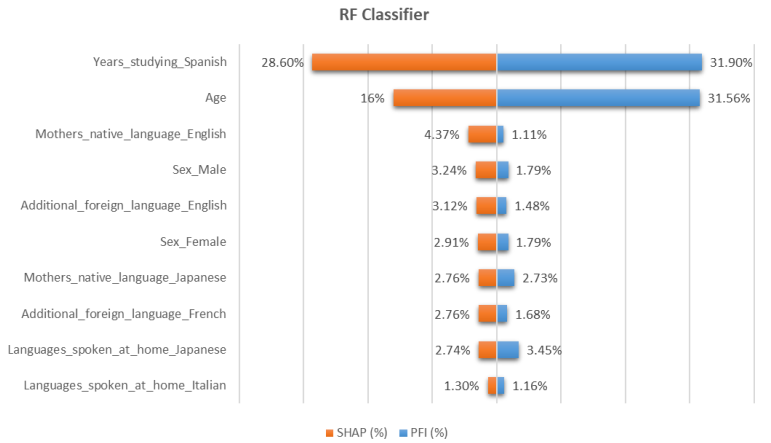

LR Classifier
Years_studying_Spanish 34.42%
Languages_spoken_at_home_English 7.43%
Age 6.20%
Languages_spoken_at_home_Japanese 3.70% 50.15%
Mothers_native_language_Japanese 3.10% 45.85%
Mothers_native_language_English 2.90% 18.59%
Languages_spoken_at_home_Dutch 0.93% 16.34%
Sex_Male 0.88% 28.20%
Mothers_native_language_German 0.50% 15.03%
Additional_foreign_language_German 0.30% 19.96%
Languages_spoken_at_home_German 0.23% 14.75%
Mothers_native_language_Korean 0.04% 13.26%
Languages_spoken_at_home_Korean 0.01% 14.07%
SHAP (%)
PFI (%)


LDA Classifier
Years_studying_Spanish 11.78% 1.63%
Languages_spoken_at_home_Greek 6.86% 0.04%
Languages_spoken_at_home_English 6.05% 0.84%
Mothers_native_language_Greek 5.27% 0.04%
Mothers_native_language_Dutch 5.21% 1.53%
Languages_spoken_at_home_Portuguese 4.61% 0.81%
Age 4.30% 0.76%
Languages_spoken_at_home_Dutch 4.16% 2.20%
Mothers_native_language_Japanese 4.16% 0.59%
Additional_foreign_language_English 4.13% 4.25%
Mothers_native_language_Portuguese 3.50% 0.27%
Additional_foreign_language_French 3.05% 1.74%
Mothers_native_language_English 2.90% 1.56%
Languages_spoken_at_home_Japanese 2.60% 1.74%
Languages_spoken_at_home_Italian 1.81% 2.24%
Languages_spoken_at_home_Russian 1.19% 2.69%
Mothers_native_language_Russian 0.83% 1.87%
SHAP (%)
PFI (%)


K-NN Classifier
Years_studying_Spanish 31.26% 9.36%
Age 15.74% 6.14%
Mothers_native_language_English 3.99% 0.48%
Languages_spoken_at_home_English 3.77% 0.60%
Additional_foreign_language_English 3.67% 0.94%
Sex_Male 3.52% 0.12%
Sex_Female 3.27% 0.12%
Additional_foreign_language_French 2.18% 0.75%
Languages_spoken_at_home_Japanese 1.53% 0.37%
Mothers_native_language_Greek 1.50% 0.42%
Mothers_native_language_Japanese 1.49% 0.17%
Languages_spoken_at_home_Greek 1.46% 0.42%
Mothers_native_language_Arabic 0.79% 0.30%
SHAP (%)
PFI (%)


CART Classifier
Years_studying_Spanish 29.55% 41.84%
Age 17.28% 22.81%
Languages_spoken_at_home_Japanese 5.24% 1.80%
Languages_spoken_at_home_Russian 5.21% 3.19%
Additional_foreign_language_English 3.53% 2.50%
Sex_Female 2.60% 2.22%
Sex_Male 2.56% 1.08%
Mothers_native_language_English 2.19% 1.67%
Mothers_native_language_Japanese 1.34%
Additional_foreign_language_French 1.29% 0.70%
Languages_spoken_at_home_Portuguese 1.00% 1.16%
Mothers_native_language_Portuguese 0.77% 3.60%
Languages_spoken_at_home_French 0.56% 2.10%
Languages_spoken_at_home_Greek 0.31% 2.19%
Mothers_native_language_Greek 0.16% 2.42%
SHAP (%)
PFI (%)


SVM Classifier
Years_studying_Spanish 33.83% 6.78%
Age 5.97% 1.77%
Additional_foreign_language_French 4.15% 0.35%
Mothers_native_language_Japanese 3.37% 0.56%
Languages_spoken_at_home_Japanese 3.15% 0.62%
Mothers_native_language_Arabic 0.82% 0.27%
Languages_spoken_at_home_Italian 0.65% 0.26%
Languages_spoken_at_home_French 0.33% 0.30%
Languages_spoken_at_home_Chinese 0.31% 0.31%
Mothers_native_language_French 0.17% 0.33%
SHAP (%)
PFI (%)


RF Classifier
Years_studying_Spanish 28.60% 31.90%
Age 16% 31.56%
Mothers_native_language_English 4.37% 1.11%
Sex_Male 3.24% 1.79%
Additional_foreign_language_English 3.12% 1.48%
Sex_Female 2.91% 1.79%
Mothers_native_language_Japanese 2.76% 2.73%
Additional_foreign_language_French 2.76% 1.68%
Languages_spoken_at_home_Japanese 2.74% 3.45%
Languages_spoken_at_home_Italian 1.30% 1.16%
SHAP (%)
PFI (%)

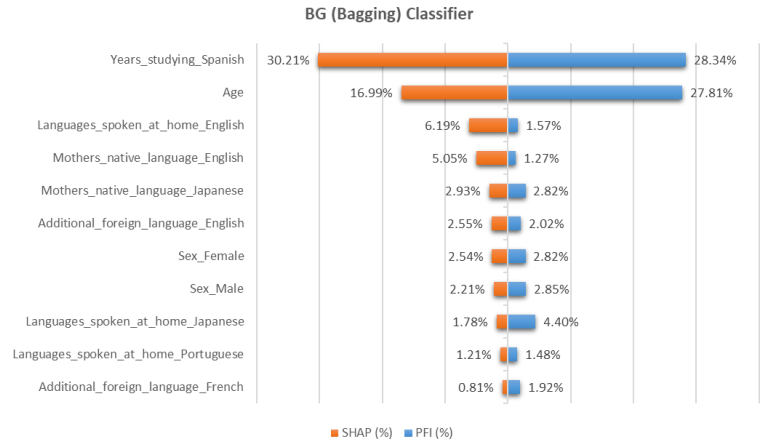


Fig. 3. Experimental evaluations of PFI and SHAP values (in %) over classifiers

6 Conclusion and Future Directions

In this work, an experimental comparison between two well-known XAI approaches (PFI and SHAP) in the framework of Language Learning classification problem has been proposed. The proposed Knowledge Generation Model (KGM) for language learning [11] extended by consolidating XAI approaches (PFI and SHAP) to enhance our understanding of language learning problems and advance the interpretability of ML models through the integration of XAI principles. By conducting a comparative analysis of these XAI approaches, we aim to deduce valuable insights into their features and draw conclusions that contribute to the field's knowledge.

The application and evaluation of our approach on the Language Learning Classification of Spanish Tertiary Education Students from the CEDEL2 database revealed noteworthy insights. In line with the principle of XAI, the primary objective of our study was to unveil and comprehend the 'black-box' that is often associated with AI models. The findings suggest that while both PFI and SHAP offer valuable insights, the SHAP model-agnostic method emerges as superior in terms of providing comprehensive visualization of feature interactions and feature importance.

As we look to the future, we believe there are several avenues that warrant further exploration. Firstly, it would be beneficial to apply this approach to other datasets and language learning contexts to assess the generalizability of our findings. Secondly, while SHAP has emerged as the more comprehensive method in this study, further exploration of other XAI techniques could yield valuable insights and provide more robust explanations. Thirdly, incorporating more granular features pertaining to student learning patterns, curriculum intricacies, and individual learner preferences could further enrich the model and enhance the accuracy and interpretability of the predictions. This study provides a significant starting point, demonstrating the potential of XAI in the language learning domain. It is hoped that future research will continue to extend the boundaries of this approach in other sectors like industry by applying the KGM and XAIs to relative problems.

Acknowledgements

This research was funded by EU Horizon 2020 Research and Innovation program WELCOME, under Grant Agreement No 870930, and by HE Research and Innovation program AIDEAS under Grand Agreement No 101057294.

References

[1] S. Pinker, The Language Instinct, New York: NY Harper Perennial Modern Classics, 1994.

[2] T. Rietveld and R. Van Hout, Statistical techniques for the study of language behaviour, Berlin: De Gruyter Mouton, 1993.

[3] Y. LeCun, Y. Bengio and G. Hinton, Deep learning, Nature, 2015.

[4] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th international conference on Machine learning*, 2008.

[5] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. R. Mohamed, N. Jaitly and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82-97, 2012.

[6] C. M. Bishop and N. M. Nasrabadi, Pattern recognition and machine learning, vol. 4, Springer New York, 2006, p. 738.

[7] I. Goodfellow, Y. Bengio and A. Courville, Deep learning, London, England: MIT press, 2016.

[8] G. P. Papanastasiou, A. S. Drigas and C. Skianis, "Serious games in preschool and primary education: Benefits and impacts on curriculum course syllabus," *International Journal of Emerging Technologies in Learning*, vol. 12, no. 1, 2017.

[9] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever, "Language models are unsupervised multitask learners," OpenAI blog, 2019.

[10] M. Thoma, "A survey of semantic segmentation," arXiv preprint, 2016.

[11] G. Antzoulatos, A. Mavropoulos, G. Tzionis, A. Karakostas, M. G. Burgos , A. G. Costas, S. Vrochidis and I. Kompatsiaris, "A Case of Advanced Visual Analytics in Complex Language Learning, Immigration-Related Scenarios," in *Interactive Mobile Communication, Technologies and Learning (IMCL 2021)*, 2021.

[12] A. Altmann, L. Toloşi, O. Sander and T. & Lengauer, Permutation importance: a corrected feature importance measure, Bioinformatics , 26(10), 2010.

[13] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 2017.

[14] M. H. Long, "Methodological principles for language teaching," in *The handbook of language teaching*, Wiley, 2009, pp. 371-394.

[15] M. Tomasello, Constructing a language: A usage-based theory of language acquisition, Harvard university press, 2005.

[16] S. M. Gass, J. Behney and L. Plonsky, Second Language Acquisition: An Introductory Course, 5th ed., New York: Routledge, 2020.

[17] N. Chomsky, A review of BF Skinner's Verbal behavior, vol. 35, Language, 1959, p. 26–58.

[18] R. Ellis, Task-based language learning and teaching, Oxford university press, 2003.

[19] P. M. Lightbown and N. Spada, How languages are learned 4th edition-Oxford Handbooks for Language Teachers, Oxford university press, 2013.

[20] Z. Dörnyei, The psychology of the language learner: Individual differences in second language acquisition, Routledge, 2014.

[21] R. C. Gardner and P. D. & MacIntyre, "A student's contributions to second language learning. Part I: Cognitive variables.," *Language teaching*, vol. 25, no. 4, pp. 211-220, 1992.

[22] T. Heift and M. Schulze, "Errors and intelligence in computer-assisted language learning: Parsers and pedagogues," *Literary and Linguistic Computing*, vol. 24, no. 2, p. 245–247, 2009.

[23] L. März, D. Trautmann and B. Roth, "Domain adaptation for part-of-speech tagging of noisy user-generated text," arXiv preprint, arXiv:1905.08920, 2019.

[24] R. Lyster and L. Ranta, "Corrective feedback and learner uptake: Negotiation of form in communicative classrooms," *Studies in second language acquisition*, vol. 19, no. 1, pp. 37-66, 1997.

[25] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti and D. Pedreschi, "A survey of methods for explaining black box models," 2018.

[26] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv, 2017.

[27] M. Carletti, C. Masiero, A. Beghi and G. A. Susto, "Explainable Machine Learning in Industry 4.0: Evaluating Feature Importance in Anomaly Detection to Enable Root Cause Analysis," in *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy, 2019.

[28] T. C. Lyn, "A survey of credit and behavioural scoring: forecasting financial risk of," *International journal of forecasting*, vol. 16, no. 2, pp. 149-172, 2000.

[29] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 2015.

[30] M. T. Ribeiro, S. Singh and C. Guestrin, ""Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016.

[31] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery," *Queue*, vol. 16, no. 3, pp. 31-57, 2018.

[32] U. Fayyad, G. Piatetsky-Shapiro and P. & Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, pp. 37-54, 1996.

[33] D. Keim, J. Kohlhammer, G. Ellis and F. Mansmann, Mastering The Information Age – Solving Problems with Visual Analytics, Eurographics Association, 2010.

[34] D. M. Rousseau and S. McCarthy, "Educating managers from an evidence-based perspective," *Academy of Management Learning & Education*, vol. 6, no. 1, pp. 84-101, 2017.

[35] M. Savić, V. Kurbalija, M. Ivanović and Z. Bosnić, "A feature selection method based on feature correlation networks," in *7th International Conference in Model and Data Engineering (MEDI), October 4–6*, Barcelona, Spain, 2017.

[36] T. Chen and Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016.

[37] P. Nskh, M. N. Varma and R. R. Naik, "Principle component analysis based intrusion detection system using support vector machine," in *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2016.

[38] M. L. Seghier, "Ten simple rules for reporting machine learning methods implementation and evaluation on biomedical data," *International Journal of Imaging Systems and Technology*, vol. 32, no. 1, pp. 5-11, 2022.

[39] I. Lage, A. Ross, S. J. Gershman, B. Kim and F. Doshi-Velez, "Human-in-the-loop interpretability prior," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018.

[40] R. Pugliese, S. Regondi and R. Marini, "Machine learning-based approach: Global trends, research directions, and regulatory standpoints," *Data Science and Management*, vol. 4, pp. 19-29, 2021.

[41] N. Stiennon, L. Ouyang, J. Wu, D. Ziegler, R. Lowe, C. Voss and P. F. ... Christiano, "Learning to summarize with human feedback," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3008-3021, 2020.

[42] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," in *In 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, 2018.

[43] T. Imielinski and H. Mannila, "A database perspective on knowledge discovery.," *Communications of the ACM*, vol. 39, pp. 58-64, 1996.