Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 16 No. 6 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijep.v16i6.63255

PAPER

An Explainable Chatbot for Student Orientation in Morocco: SHAP and LIME on a Hybrid Neural Network-SVM Model

Morad Badrani[1] (✉), Adil Marouan[1], Nabil Kannouf[2], Abdelaziz Chetouani[1]

[1]Mohammed First University, Oujda, Morocco

[2]Abdelmalek Essaadi University, Tetouan, Morocco

m.badrani@ump.ac.ma

ABSTRACT

Poor orientation feeds school dropout in Morocco. Leaving the scientific common core (tronc commun), a pupil must pick one of three tracks: Mathematical Sciences (SM), Physical Sciences (SF), or Life and Earth Sciences (SVT). Software can help with that choice. Accuracy alone cannot: a model that will not say why it recommends a track earns no trust from the families and counsellors who must act on it. Our system has three parts. A chatbot gathers the pupil's marks. A sequential hybrid neural network into SVM issues the prediction. An explainability layer built on SHAP and LIME accounts for it, and that layer is the contribution. We trained on some 20,000 anonymised records from the MASSAR system, obtained through the Nador provincial directorate and covering 2018 to 2022. The hybrid reaches 98.87% on the held-out set, against 95.79% for the network alone and 97.52% for the SVM alone, with macro AUC 0.9981 and Cohen's kappa and MCC both 0.983. The national platform holds over seven million pupil records and cannot be extracted at that scale, so we repeated everything on OULAD, public and far larger, after mapping its variables onto our feature schema. Accuracy is 98.73%. SHAP puts the decision on the three science subjects; LIME accounts for pupils one at a time.

KEYWORDS

student orientation, conversational agent, chatbot, hybrid model, support vector machine, neural network, explainable AI, SHAP, LIME, MASSAR, OULAD, big data

1 INTRODUCTION

Dropout is a chronic weakness of Moroccan schooling: 304,545 pupils left without finishing in 2019–2020 alone [1]. Causes abound, but one keeps surfacing in the work of education authorities and researchers alike. Pupils land in tracks that suit neither their abilities nor their interests, and disengagement follows. Guidance, then, bears directly on retention.

Badrani, M., Marouan, A., Kannouf, N., Chetouani, A. (2026). An Explainable Chatbot for Student Orientation in Morocco: SHAP and LIME on a Hybrid Neural Network-SVM Model. *International Journal of Engineering Pedagogy (iJEP)*, 16(6), pp. 164–175. https://doi.org/10.3991/ijep.v16i6.63255

Article submitted 2026-06-04. Revision uploaded 2026-09-05. Final acceptance 2026-09-07.

© 2026 by the authors of this article. Published under CC-BY.

Orientation happens twice here, not once. The first decision closes lower-secondary school. The second closes the scientific common core, and it is the consequential one: SM, SF, or SVT, with everything academic downstream hanging on it. Yet the choice is still made by hand, weighted by family preference, rarely by any systematic reading of the pupil's own record. Aptitude and track drift apart.

Two conditions govern whether such a tool is worth building. Reach is the first: pupils and families must be able to use it untrained, which points to the chatbot idiom they already know. Readability is the second, and the harder one. A recommendation nobody can trace back to a reason will not be acted on, whatever its accuracy, and hybrid models are black boxes by construction.

Scope and contribution. This paper answers both conditions at once, with an explainable conversational system for the SM/SF/SVT decision. It combines a *chatbot interface* that gathers the pupil's marks, a *hybrid Sequential Neural Network-SVM model* that issues the prediction, and an explainability layer built on SHAP and LIME. That layer is the **main contribution** here, since it turns each recommendation into something a family can read and argue with. We validate the model twice over. Real anonymised MASSAR records fix its performance in the setting it is meant for. A large public benchmark, whose features we aligned with the national schema, tells us whether the same architecture would hold up at a national scale. Section 2 reviews related work, Section 3 describes the system and its data, Section 4 covers the model and its results on both datasets, Section 5 the explainability layer, and Section 6 concludes.

2 RELATED WORK

2.1 Machine learning for student orientation

Track prediction from school records has attracted steady work, Moroccan work included. Ouatik et al. [6] paired supervised classifiers with big data processing for roughly 92% accuracy. El Mrabet and Ait Moussa [2] built a supervised framework for orientation prediction, later extended to data from connected objects. Mimis et al. [3] grounded a guidance framework for Moroccan preparatory classes in educational data mining. Zahour et al. [4] took the conversational route, with a chatbot for educational and vocational guidance: evidence that the channel works here. Our own earlier papers compared classical classifiers for lower-secondary guidance [5], then turned to ensemble and hybrid designs on Big Data infrastructure [6].

A neighbouring literature abroad predicts performance rather than track, on much the same features. Deep networks over course grades [7]. Final marks forecast early in gamified courses [8]. Supervised techniques were compared on high-school data [9]. School records clearly carry signal; Bayesian generative modelling of results across course networks [22] is included; raising accuracy is where the attention goes.

2.2 Hybrid and ensemble architectures

On educational data no single classifier wins consistently, a finding repeated often enough to have pushed the field toward ensembles and hybrids. Sokkhey and Okazaki [10] confirmed the hybrid advantage for academic performance. Our two

families fail differently. Support vector machines draw maximal-margin boundaries and generalise well on moderate, well-prepared features. Neural networks learn non-linear representations from the data but overfit sooner and resist reading. Pair them and robustness improves, whether in parallel by vote or sequentially, one feeding the other. We take the second route.

2.3 Explainable AI and its use in education

Performance bought that way costs transparency, and explainable artificial intelligence grew out of the bargain. Adadi and Berrada [11] and Arrieta et al. [12] map the families of methods. Miller [13] adds a condition: an explanation is useful only if it resembles how people explain things to each other. Rudin [14] goes further and holds that high-stakes decisions should use models interpretable by construction; Doshi-Velez and Kim [15] want interpretability measured properly. Two post-hoc model-agnostic methods dominate in practice. LIME [16] fits a locally faithful surrogate around a single prediction, and SHAP [23] derives attributions from cooperative game theory and keeps the local and global views consistent. Counterfactual explanations [24] name the smallest input change that would flip the decision, which is generally what the person affected wants to know.

Education raises the stakes, the recommendations going to minors and their families through a counsellor. Holstein et al. [17] found tools designed for teacher-AI complementarity adopted more readily than tools that hand down decisions, and Sokol and Flach [18] argue that explanations delivered in conversation are easier to use than static charts. Orientation systems, however, rarely integrate XAI at all. Accuracy is reported; how a recommendation might be justified to the pupil who receives it is left open.

3 THE CONVERSATIONAL ORIENTATION SYSTEM

A pupil finishing the scientific common core passes through three stages. The conversational interface collects the profile, the hybrid model predicts a track, and the recommendation is then explained, which is the subject of Section 5. Below we cover the interface and the two datasets; Section 4 takes up the model and its results.

3.1 Chatbot interface

The chatbot used here was purpose-built for this study. It carries the whole system to the user, asking for marks subject by subject and answering with a track. Nothing must be learned first, which is the point: the exchange runs as a guided sequence, in the idiom pupils already use elsewhere. Conversational guidance is not new to the Moroccan context, Zahour et al. [4] having shown the channel works here. The interface reported below is a later development of the systems described in [19] and rests on the personalised guidance approach of [20], extended in this paper so that every recommendation now travels with an explanation. Figure 1 shows a session end-to-end.

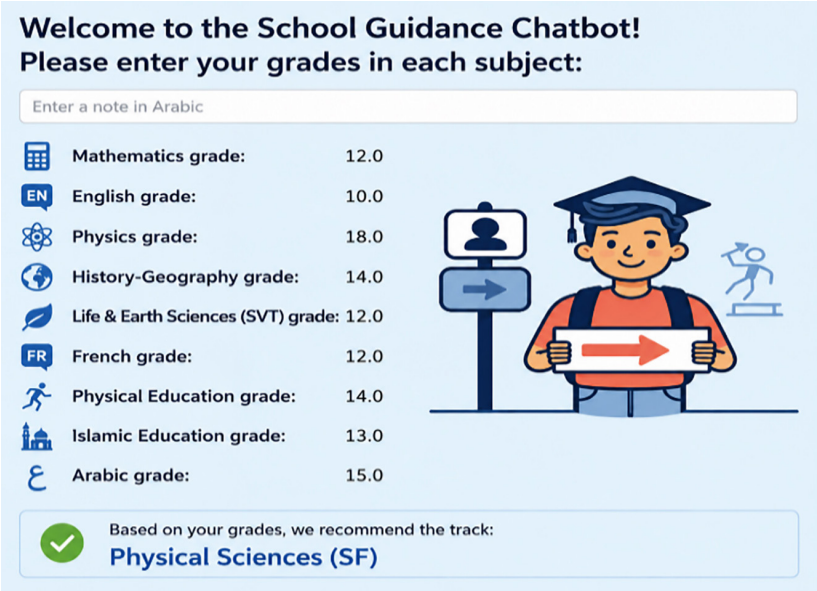


Fig. 1. The conversational interface: marks entered subject by subject, a track returned

3.2 Primary dataset: real MASSAR records

The records are real, obtained from the Direction Provinciale de l'Education de Nador under Morocco's right of access to information. They cover pupils enrolled in the scientific common core in the province from 2018–2019 to 2021–2022, and they come from MASSAR, the ministerial platform where every Moroccan school registers enrolments, marks, and orientation decisions. Marks and subjects are all the file contains. No pupil name, no pupil identifier, no school identifier, so nothing is traceable to an individual or an establishment, and no confidentiality issue for research use. The same extract served the earlier study [20], and it is reused deliberately here: a setting already characterised is where a new architecture and an explainability layer can be judged. Some 20,000 pupils are covered, each described as a counsellors rather than a proxy.

Preprocessing stayed light, the records arriving structured. A stray index column dropped, the few missing marks handled, and the nine variables standardised. The records were split 80/20 into a training and a test set, stratified so that the tracks keep their provincial proportions. The test set holds 3,982 pupils, 1,408 of them in SF, 1,376 in SM, and 1,198 in SVT; every figure reported in Section 4 is measured on those 3,982 and on no others. Nothing identifying sits in the file, but it is not ours to redistribute, having been released for this research. Aggregated results are available from the corresponding author.

3.3 Toward national deployment: a large-scale benchmark

What a provincial dataset settles is performance in the real setting. What it leaves open is behaviour at another order of magnitude, which national deployment would bring. MASSAR centralises over seven million pupil records and 240,000 teachers, generating a large daily volume. No extract on that scale is available for research: the records belong to minors, and ministerial confidentiality rules out large-scale sharing of individual academic data.

We turned to the Open University Learning Analytics Dataset (OULAD) [21] instead. Public, and vastly larger in interaction volume than our provincial file: 32,593 students,

22 courses, assessment results, and some 10.6 million logged interactions with the virtual learning environment. At that size the architecture and the infrastructure beneath it can be exercised nearer to national conditions.

Feature regrouping and normalisation. Three steps bring OULAD onto our national feature schema. The first is *competency regrouping*. The 22 courses fall into three competency families that match the Moroccan scientific tracks, which turns the label into the same three-class decision: mathematics courses go to SM, physics and technology courses to SF, and natural science and biology to SVT. Next comes *feature alignment*, where assessment and interaction variables are aggregated into per-competency indicators; the resulting vector has the same shape as the nine subject marks of the MASSAR data. Last is *scale normalisation*: every indicator is rescaled onto the Moroccan 0–20 range, then standardised exactly as the MASSAR marks were, so both datasets reach the model through one pipeline. Without that alignment, the two sets of results in Section 4 would not be comparable. One caveat stated plainly: OULAD comes from British distance learning and serves as a **large-scale benchmark rather than a substitute** for Moroccan records; the real-data results of Section 4.2 remain the reference for the Moroccan setting.

4 HYBRID MODEL AND RESULTS

4.1 Sequential hybrid neural network-SVM

The design is sequential: features from a neural network and classification by an SVM (Figure 2). Three hidden layers of 64, 32, and 16 neurones with ReLU activation train on the standardised inputs, with dropout between them against overfitting and Adam for optimisation. The last hidden layer's activations form the learned representation and feed an SVM with an RBF kernel, tuned by 5-fold cross-validation to C = 10 and gamma = 0.01, which returns a membership probability per track. The network is thus free to catch non-linear structure, while the SVM sets robust boundaries in the transformed space. The pair also exposes one predict-probability function, which is precisely what the explainability layer of Section 5 needs to attribute a recommendation back to subjects. Scikit-learn and TensorFlow/Keras throughout.

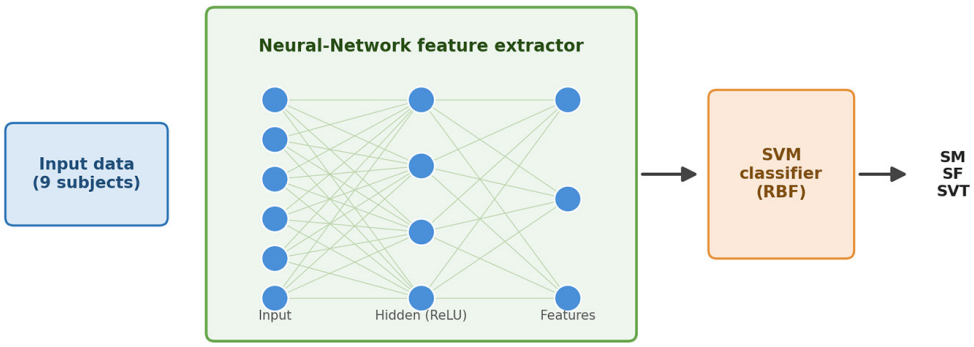


Fig. 2. Sequential hybrid architecture

Note: The network extracts features; the SVM classifies them into SM, SF or SVT.

4.2 Results on real MASSAR data

Table 1 sets the hybrid against the neural network alone and the SVM alone, all three trained on the same data and measured on the held-out MASSAR test set.

Table 1. Performance on the real MASSAR dataset

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Neural Network only	95.80	96.79	96.30	95.79
SVM only	97.20	97.50	97.35	97.52
Hybrid NN-SVM	**98.87**	**98.87**	**98.87**	**98.87**

At 98.87%, the hybrid clears the network alone (95.79%) and the SVM alone (97.52%), the expected return on a sequential arrangement whose two halves are good at different things. Per-class figures come out level, near 0.99 on precision, recall, and F1 for SF, SM, and SVT alike. Not a detail, that evenness: a model performing only on the majority track would quietly penalise pupils belonging to the smaller ones [25].

4.3 Results on the large-scale benchmark

Table 2 reports the performance of the same architecture, with the same preprocessing pipeline, on the aligned OULAD data.

Table 2. Performance on the OULAD large-scale benchmark after feature regrouping and normalisation

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Neural Network only	95.62	96.51	96.06	95.74
SVM only	97.08	97.31	97.19	97.28
Hybrid NN-SVM	**98.71**	**98.74**	**98.72**	**98.73**

4.4 Consistency across the two datasets

The two evaluations land almost together (see Figure 3). 98.87% on real MASSAR records, 98.73% on the aligned benchmark, 0.14 points apart, and the three models ordered identically both times (network < SVM < hybrid). Architecture, hyperparameters, and preprocessing were held fixed, and only the data changed, so the agreement carries information: the hybrid owes its accuracy to neither the provincial sample nor its size and behaves the same on a far larger volume. That is the property national deployment would need.

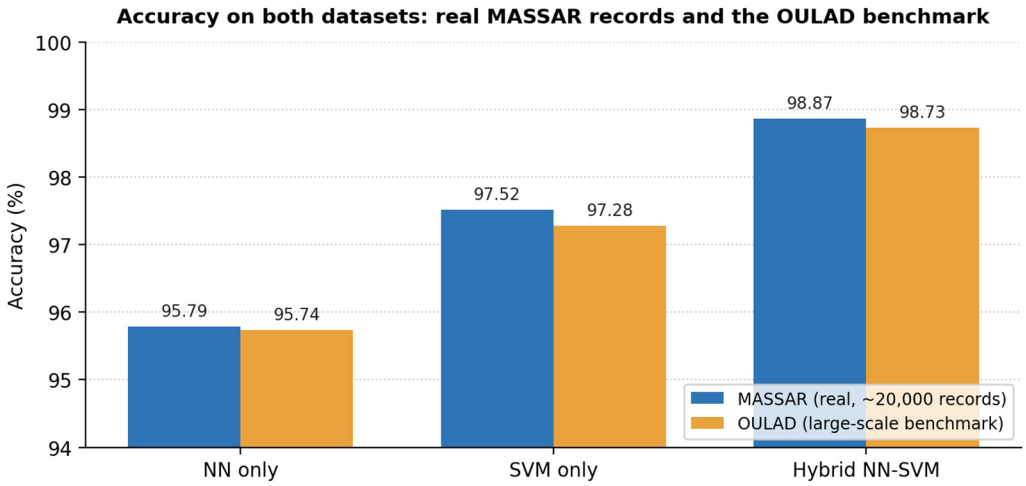


Fig. 3. Accuracy of the three models on MASSAR and on the aligned OULAD benchmark

4.5 Discrimination and agreement metrics

Where classes are not perfectly balanced, accuracy misleads, so we add measures on the real MASSAR data that survive imbalance and examine discrimination across thresholds. Macro one-vs-rest AUC reaches 0.9981, near-clean separation of the three tracks. Cohen's kappa gives 0.983, and the Matthews correlation coefficient is the same; the latter is recommended precisely for resisting differing class frequencies. Both sit far above chance or a majority-class rule, so the agreement between model and actual orientations is real and evenly spread over SF, SM, and SVT rather than an artefact of the class distribution. Table 3 collects these alongside the per-class results.

Table 3. Detailed evaluation of the hybrid model on the real MASSAR test set

Track/Metric	Precision	Recall	F1-Score
SF (Physical Sciences)	0.99	0.99	0.99
SM (Mathematical Sciences)	0.99	0.99	0.99
SVT (Life & Earth Sciences)	0.99	0.98	0.99
Overall accuracy	–	–	**98.87%**
Macro AUC (OvR)	0.9981	**Cohen's kappa**	0.983
MCC	0.983	–	–

4.6 Big data infrastructure and scalability

Benchmark interaction volume, and the millions of records of national deployment imply a call for distributed infrastructure. The configuration is the one described in [20]: a Hadoop cluster of one NameNode and three DataNodes, HDFS for storage, YARN for resources, and Spark for processing. Training spreads across the cluster, interaction records are batched so memory stops being the constraint, and extracted features sit in a distributed cache to accelerate the SVM stage. Execution time drops about 40%, from 33 seconds on a single machine to 19.8 (see Figure 4), as reported in [20]. The figure is recalled rather than claimed anew; it belongs to the inherited infrastructure, and it matters here only because that infrastructure is what keeps the approach workable.

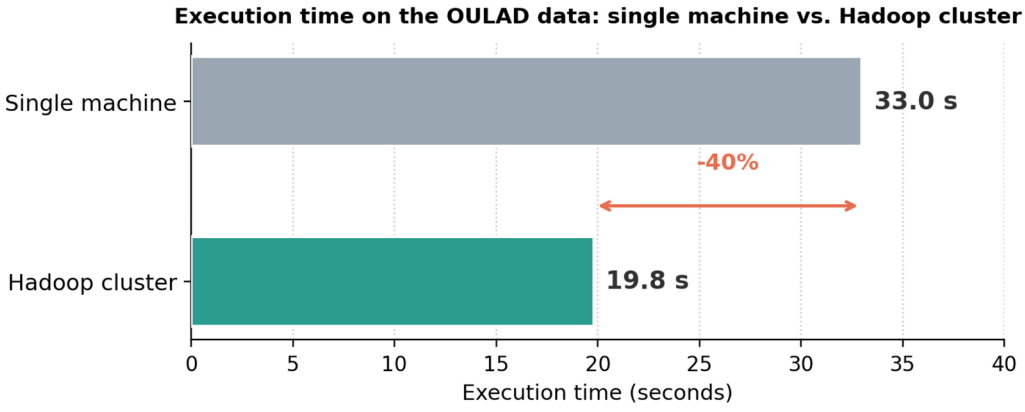


Fig. 4. Execution time, single machine versus Hadoop cluster

5 EXPLAINABILITY LAYER (MAIN CONTRIBUTION)

Accurate and usable, perhaps; transparent, no. A hybrid is a black box, and in orientation that has consequences, since pupils, parents, teachers, and counsellors all have grounds to ask why this track rather than another [14], [17]. So we add an explainability layer inside the conversational system, and no recommendation arrives alone. Below are the two methods it rests on, then the global and the individual explanations obtained from the model trained on real MASSAR records.

5.1 SHAP and LIME

The layer rests on two complementary methods. **LIME** (Local Interpretable Model-agnostic Explanations) [16] perturbs the instance, watches how the predictions move, and fits a weighted linear model that approximates the complex one in that neighbourhood. Here it tells us how much each subject pushed toward the recommended track. **SHAP** (SHapley Additive exPlanations) borrows from cooperative game theory, giving each feature an importance value that satisfies efficiency, symmetry, and nullity. In a sensitive setting, those guarantees are worth having. The pair divides the work: LIME reads more naturally to a non-specialist, and SHAP stands on firmer theoretical ground and keeps the local and global pictures consistent.

5.2 Global explanation (SHAP)

Across the whole test set, SHAP reports how much each subject weighs on average. The three sciences take nearly everything (see Figure 5): Mathematics first at 0.078 mean absolute value, followed by Physics-Chemistry 0.071, Life and Earth Sciences 0.064, and the six others trailing. Nothing startling there. SM, SF and SVT are scientific specialisations, so a model leaning on the matching marks follows the shape of the problem it was handed. The interesting reading is negative: no spurious correlation with a non-scientific subject got picked up on the way, worth knowing before an explanation is shown to a family.

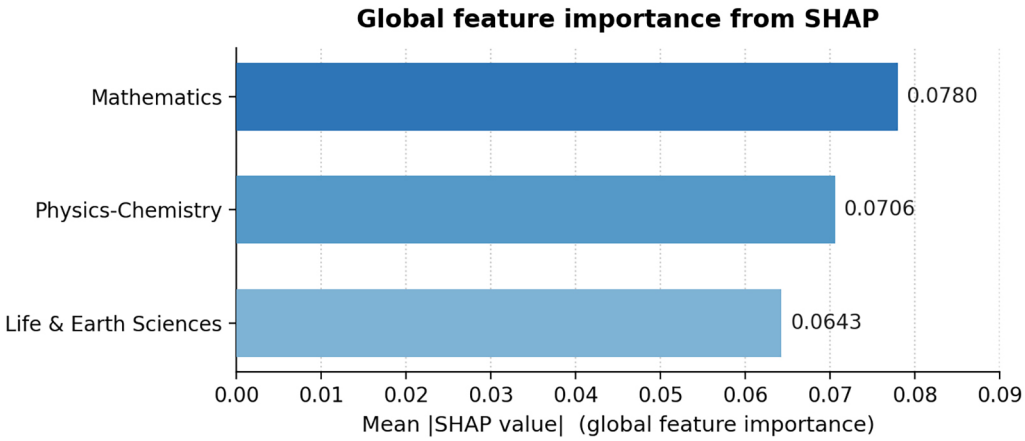


Fig. 5. Mean absolute SHAP value per subject on the MASSAR data

5.3 Individual explanation (LIME)

SHAP describes the model. A pupil and a counselor face one recommendation, which is LIME's ground. Figure 6 takes a test-set pupil oriented toward mathematical sciences and gives the contributions with respect to that class. Mathematics leads at +0.312, physics-chemistry follows at +0.214, and between them they account for most of the decision. Life and Earth Sciences adds little, +0.039, as one would expect of a pupil whose strength among the sciences lies elsewhere. French pulls the other way at –0.081, then History-Geography, Islamic Education, and Arabic, none of them weighing much. A teacher reads this without difficulty: strong mathematics and strong physics point to SM, weaker literary marks do not contradict it, and the recommendation follows from the pupil's own profile rather than from anything opaque. Two contributions deserve a word rather than a silence. English appears at +0.107, ahead of Life and Earth Sciences and Physical Education at +0.056, neither belonging to the scientific core of the decision. The model learned from orientation decisions actually taken between 2018 and 2022, and those weigh the whole record rather than the science marks alone; SM being the most selective of the three tracks, a globally solid profile counts in its favor even where a subject carries no scientific content. Their weight stays well under mathematics and physics, and Figure 5 confirms that the sciences carry the model overall. What the local view adds is a reminder that one pupil's explanation need not reproduce the global ranking. Through the chatbot, an explanation of this kind gives the pupil and the family something to read and to bring to a counselor [18], not a verdict with nothing behind it.

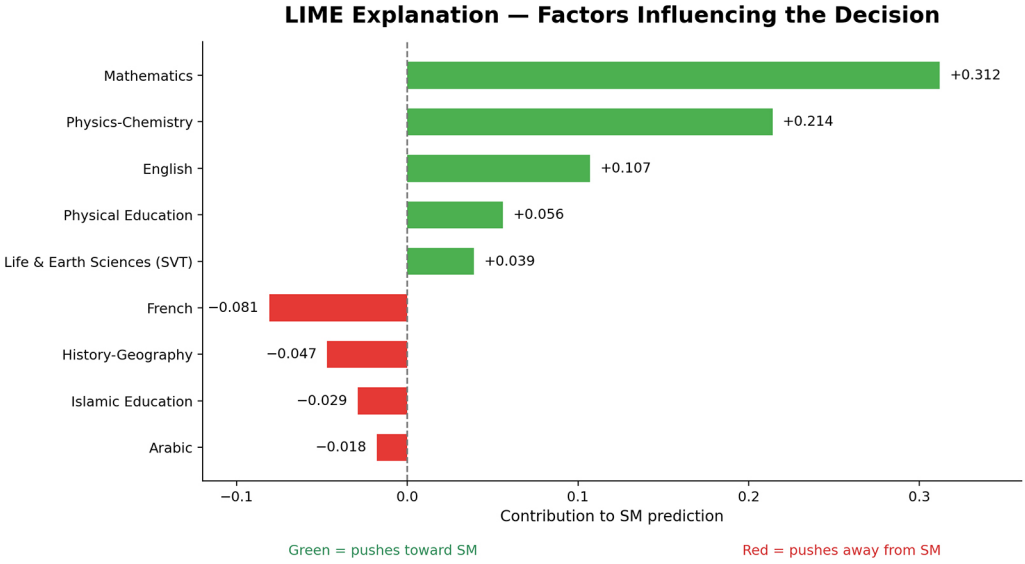


Fig. 6. LIME explanation for a pupil orientated toward SM, computed with respect to that class

6 CONCLUSION

We have described an explainable conversational system for the scientific-track decision closing the Moroccan common core: A chatbot interface, a hybrid Sequential Neural Network-SVM model, and explainability layer on SHAP and LIME. Across some 20,000 anonymised MASSAR records from the Nador directorate, the hybrid reaches 98.87%, where the network alone gives 95.79% and the SVM alone 97.52%,

with a macro AUC of 0.9981, Cohen's kappa and MCC of 0.983, and per-class figures level across SF, SM, and SVT. Unchanged, the architecture reaches 98.73% on a large public benchmark aligned to the same feature schema, reproducing level and ordering alike, which is what supports going beyond one directorate. The explainability layer is the contribution, and it is what makes a recommendation interpretable: SHAP puts the decision on the three science subjects, as the problem's structure would predict, and LIME accounts for individual cases in terms a teacher would recognise.

7 CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

8 FUNDING

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

9 AVAILABILITY OF DATA

The datasets collected and analysed during the current study are available from the corresponding author on request. The corresponding author had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

10 REFERENCES

[1] A. Ibourk and S. Raoui, "Territorial obstacles causing early school dropout in Morocco: Multivariate spatial analysis," *Heliyon*, vol. 11, no. 1, Art. no. e41386, 2025. https://doi.org/10.1016/j.heliyon.2024.e41386

[2] H. El Mrabet and A. Ait Moussa, "A framework for predicting academic orientation using supervised machine learning," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 12, pp. 16539–16549, 2023. https://doi.org/10.1007/s12652-022-03909-7

[3] M. Mimis, M. El Hajji, Y. Es-saady, A. Oueld Guejdi, H. Douzi, and D. Mammass, "A framework for smart academic guidance using educational data mining," *Educ. Inf. Technol. (Dordr).*, vol. 24, pp. 1379–1393, 2019. https://doi.org/10.1007/s10639-018-9838-8

[4] O. Zahour, E. H. Benlahmar, A. Eddaoui, H. Ouchra, and O. Hourrane, "A system for educational and vocational guidance in Morocco: Chatbot E-orientation," *Procedia Comput. Sci.*, vol. 175, pp. 554–559, 2020. https://doi.org/10.1016/j.procs.2020.07.079

[5] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Smart techniques for Moroccan students' orientation," in *International Conference on Emerging Technologies and Intelligent Systems*, Springer, 2022, pp. 361–366. https://doi.org/10.1007/978-3-031-20429-6_33

[6] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Big data and ensemble learning for effective student orientation in Morocco," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 3, pp. 1904–1910, 2024. https://doi.org/10.11591/ijeecs.v36.i3.pp1904-1910

[7] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of students' academic performance based on courses' grades using deep neural networks," *IEEE Access*, vol. 9, pp. 140731–140746, 2021. https://doi.org/10.1109/ACCESS.2021.3119596

[8] A. H. Nabizadeh, D. Goncalves, S. Gama, and J. Jorge, "Early prediction of students' final grades in a gamified course," *IEEE Transactions on Learning Technologies*, vol. 15, no. 3, pp. 311–325, 2022. https://doi.org/10.1109/TLT.2022.3170494

[9] N. N. Sanchez-Pozo, J. S. Mejia-Ordonez, D. C. Chamorro, D. Mayorca-Torres, and D. H. Peluffo-Ordonez, "Predicting high school students' academic performance: A comparative study of supervised machine learning techniques," in *2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop*, IEEE, 2021, pp. 1–6. https://doi.org/10.1109/IEEECONF53024.2021.9733756

[10] P. Sokkhey and T. Okazaki, "Hybrid machine learning algorithms for predicting academic performance," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 1, pp. 32–41, 2020. https://doi.org/10.14569/IJACSA.2020.0110104

[11] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018. https://doi.org/10.1109/ACCESS.2018.2870052

[12] A. Barredo Arrieta *et al.*, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020. https://doi.org/10.1016/j.inffus.2019.12.012

[13] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artif. Intell.*, vol. 267, pp. 1–38, 2019. https://doi.org/10.1016/j.artint.2018.07.007

[14] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019. https://doi.org/10.1038/s42256-019-0048-x

[15] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608,* 2017. https://doi.org/10.48550/arXiv.1702.08608

[16] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, 2016, pp. 1135–1144. https://doi.org/10.1145/2939672.2939778

[17] K. Holstein, B. M. McLaren, and V. Aleven, "Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity," *Journal of Learning Analytics*, vol. 6, no. 2, pp. 27–52, 2019. https://doi.org/10.18608/jla.2019.62.3

[18] K. Sokol and P. Flach, "Conversational explanations of machine learning predictions through class-contrastive counterfactual statements," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, California, 2018, pp. 5785–5786. https://doi.org/10.24963/ijcai.2018/836

[19] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Pathways to success: A study of classification algorithms for guiding middle school students in Morocco," in *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security*, 2023, pp. 1–5. https://doi.org/10.1145/3607720.3607748

[20] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Personalized guidance for Moroccan students: An approach based on machine learning and big data," *International Journal of Engineering Pedagogy (iJEP)*, vol. 15, no. 1, pp. 125–136, 2025. https://doi.org/10.3991/ijep.v15i1.51985

[21] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open university learning analytics dataset," *Sci. Data*, vol. 4, no. 1, Art. no. 170171, 2017. https://doi.org/10.1038/sdata.2017.171

[22] M. R. Haas, C. Caprani, and B. Van Beurden, "Bayesian generative modelling of student results in course networks," *Journal of Learning Analytics*, vol. 10, no. 3, pp. 135–152, 2023. https://doi.org/10.18608/jla.2023.7957

[23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.
[24] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box: Automated decisions and the GDPR," *Harvard Journal of Law and Technology*, vol. 31, no. 2, pp. 841–887, 2018. https://doi.org/10.2139/ssrn.3063289
[25] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. 45, no. 4, pp. 427–437, 2009. https://doi.org/10.1016/j.ipm.2009.03.002

11 AUTHORS

Morad Badrani is currently a Professor at the Faculty of Sciences and Techniques of Taourirt, LIRMO-MSI Laboratory, MASI Team, ENCG, Mohammed First University, Oujda, Morocco (E-mail: m.badrani@ump.ac.ma).

Adil Marouan is a Ph.D. student within the MASI Team at the LIRMO-MSI Laboratory, MASI Team, ENCG, Mohammed First University, Oujda, Morocco (E-mail: adil.marouan@ump.ac.ma).

Nabil Kannouf is currently working as Professor of computer science at ENSA, Abdelmalek Essaidi University, Tetouan, Morocco (E-mail: n.kannouf@uae.ac.ma).

Abdelaziz Chetouani is currently working as Professor in the Department of CTAM at LIRMO-MSI Laboratory, MASI Team, ENCG, Mohamemd First University, Oujda, Morocco (E-mail: a.chetouani@ump.ac.ma).