Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Towards Predictive Modeling of High School Students' Graduation Mentions Using Classification Techniques

Tarik Ahajjam
University of Hassan 2
Sara Jbaida
University of Mohammed V, ENSIAS Rabat
Amina Hassani
University of Hassan 2
Mohammed moutib
University of Hassan 2
Haidar Aissa
University of Hassan 2
Yousef Farhaoui
y.farhaoui@fste.umi.ac.ma

umi

Research Article

Keywords: Machine Learning, XGBoost, Student Classification, Graduation Mentions, Educational Data Mining, Academic Orientation, High School Performance, Decision-Support Systems

Posted Date: February 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8795951/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study develops a machine learning–based system to classify high school students' graduation mentions using common-core grades. Using data from 72,020 students in the Guelmim-Oued Noun region to solve the problem of orientation students, the XGBoost algorithm was trained for multi-class classification across literary, scientific, and technical streams. The model achieved high accuracy: **92%** for the literary stream, **89%** for the scientific stream, and **90%** for the technical stream. These results demonstrate the robustness of XGBoost for educational analytics. The proposed classifier can support early orientation decisions and help identify students requiring academic intervention. This work contributes to integrating data-driven tools into Morocco's educational guidance system.

1. Introduction

Academic and vocational orientation is a priority in the Moroccan education system today, given the complexity of training pathways and profound changes have occurred on the labor market over these past years. several alternative educational programs have emerged in part to prevent students from dropping out or leaving school prematurely, and as a result students are choosing their paths at an earlier age than before. From the first year of secondary education (common core), the students have to decide whether they want a scientific, technical or literary option and from there on, the choices become numerous. *has led to significant changes* in recent years: the recent economic crisis not only destroyed numerous jobs, as evidenced by alarming unemployment figures, but also transformed the professional landscape, increasing precariousness, including a reduction in permanent contracts, easier layoffs, more temporary or seasonal contracts, and lower salaries. Furthermore, social and technological progress has given rise to new professions while others have disappeared, placing greater value on skills previously overlooked, such as adaptability to change, communication skills, and proficiency with new technologies. have been deeply affected by these changes, with career expectations very different from those of previous generations. It is therefore not surprising that academic and professional guidance is considered one of the main areas of intervention in secondary education and should continue throughout life, aiming to prepare young people to face the many academic and professional decisions that will shape their future. This guidance should be based on four fundamental pillars: self-knowledge, knowledge of the academic and professional environment, decision-making skills training, and management of personal and professional life projects, while preparing students for the many changes they will experience, such as changing workplaces, periods of unemployment, or even career shifts. Since the launch of reform projects—from the National Commission for Education Reform to Framework Law 17.51, passing through the National Charter for Education and Training and the strategic vision for reform—the reform of the educational system continues to face major obstacles that have limited the implementation of planned educational initiatives and the response to challenges and expectations. National and international reports have highlighted dysfunctions such as low productivity, high dropout rates, and poor learning outcomes, with more than 300,000 students leaving school each year without minimum skills, over half of them in primary education, in addition to pedagogical issues related to curricula, textbooks, and continuing education. As for school and professional guidance, it still does not

fully fulfill its mission as envisaged by the reform projects, making it necessary to review the regulations governing guidance areas, counseling centers, and the role of guidance counselors. In this context, the Moroccan educational system, faced with rapid social, economic, cultural, political, scientific, and technological transformations, must adopt a new logic in its educational guidance policies, as these constitute a cornerstone of the educational process.

The remainder of this paper is organized as follows. Section 2 presents the theoretical framework of the study. Section 3 provides an overview of related work. Section 4 describes the proposed methodology and the obtained results. Finally, Section 5 concludes the paper and outlines future research directions.

2. Theoretical Framework

Morocco, like many countries, faces with great concern the problem of school supervision. Over the past twenty years various national policy documents such as the National Charter for Education and Training in1999, the Strategic Vision in 2015–2030 and Framework Law 51.17 have emphasized the need to support students in managing difficult educational pathways. These reforms have recognized educational support as an essential instrument for promoting equity ensuring informed decision making, and helping students align their academic trajectories with their abilities and goals. Despite these aspirations, guidance techniques in Morocco remain limited in their effectiveness. Many institutions lack sufficient human resources, particularly due to the small number of professional guidance counselors, and the available decision making tools are neither up to date nor data driven. As a result, students often make crucial decisions especially during the Common core without having adequate information or support, which increases the risks of misorientation and academic failure.

School dropout is a complex phenomenon influenced by individual, family, institutional, and regional factors. National-level reports show that hundreds of thousands of Moroccan students exit the school system every year without mastering even a minimal skill set and this is particularly easy to observe from the first year of secondary education. The Tronc Commun is a very vulnerable time when students must begin making important choices meeting up with the demand for other that will follow. So that is the age at which they need to make their initial decisions about what kind of a future they want in terms of studying science, engineering or humanities. There is a wealth of research about how well early academic performance predicts later -students grades in the *Tronc Commun* can be good indicators of future performance against those who are in or could drop down to it. The appearance of substantial discrepancies between rural and urban students at this stage make the quest to deliver equitable, suitable support more difficult.

These problems are heavily felt in the Guelmim-Oued Noun province. With large rural areas, economic challenges and a lack of infrastructure, educating is often a struggle in the region. wide performance disparities among schools are often observed, and A large number of students do not have consistent access to digital devices or conducive learning settings. Counseling services also are restricted in availability, as few counselors must cover a large number of students spread out in rural areas.

Therefore, many students make educational decision without sufficient assistance, which results in a significant number who become become lost, discouraged, and eventually drop out, disenfranchised and drop out. Thus, the particular conditions of the region make it necessary that tools that are innovative, scalable and fair must be developed to support decision making for guidance and detection of students at risk for failure.

In this regard, machine learning offers a potential solution to enhance decision-making in educational institutions. It involves the use of data with methods from machine learning uncover patterns existing in student data that would be difficult for human experts and teachers to detect, predict academic outcomes and classify students according to their level of performance. These technologies have become increasingly utilized within international educational research to find ways of supporting early diagnosis, personalized learning, and pathway guidance for schools. If we have the categorized outcome of the tasks such as school {grades, mentions} then classification algorithm would be better to use such as Random Forest, Support Vector Machines, Neural Network and Gradient Boosting. They rely on quantifiable data and generate robust and repeatableresults that can enhance the fairness, consistency, as well as accuracy of guidance decisions. Their inclusion in the Moroccan educational system, in places like Guelmim-Oued Noun for example, can help compensate for the structural deficiencies that plague guidance services, ultimately serving as prophylactic support to school failure with a view to identifying student needs earlier and better.

3. Related Work

Several studies have been conducted to confront the problem of school orientation and school dropout. Thus, H. Sun et al. [1] proposed a study to achieve good orientation of engineering students using the XGBoost algorithm. Also, T. Han et al. [2] conducted a study to encourage first-cycle students to follow online learning. P. B. Kikunda et al. [3] develop a model that predicts students who require academic support and guidance. M. Fu et al. [4] carried out a hybrid approach using Bayesian networks and decision trees in order to improve the accuracy and efficiency of educational data analysis. H. Yang et al. [5] conduct a study revealing, thanks to XGBoost and SHAP, that technical skills strongly influence the preparation and teaching phases, while ethics and pedagogical decision-making are essential in evaluation. Moreover, D. E. Kaouache et al. [6] conducted a statistical study to strengthen orientation and entrepreneurial intention among university students. Furthermore, I. Uysal et al. [7] evaluate the effect of classroom pedagogical intervention on study habits and academic performance in order to predict the risk of failure based on survey responses. B. Rončević Zubković [8] conducts a study on the importance of orientation for students who master ICT tools. However, Iatrellis et al. [9] proposed two learning methods: an unsupervised one with K-means and a supervised one with Random Forest to predict students' results. In addition, A. Tarik et al. [10] conducted a comparative study to predict students' performance during Covid-19. Moreover, Tomasevic et al. [11] compared several machine learning algorithms to predict students' performance in exams, notably in order to identify those presenting a "high-risk" of dropping the course. Then, T. Ahajjam et al. [12] develop a system that predicts the baccalaureate average using neural networks. Furthermore, Abu Saa et al. [13] used decision trees, the

naïve Bayes classifier, and artificial neural networks to predict students' academic performance from a new dataset from a private university in the United Arab Emirates. Moreover, A. Guilherme [14] demonstrates the importance of the teacher and the use of artificial intelligence techniques in the educational system. Additionally, Sekeroglu et al. [15] proposed six experiments covering two areas of machine learning — classification and regression — in order to evaluate students' performance. The results show that the radial basis function neural network outperforms all other models examined in this study. Y. K. Ever et al. [16] conduct a comparative study on four machine learning techniques and show that neural networks offer higher and more stable predictions. J. Y. Chung and S. Lee [17] use the Random Forest algorithm to predict the risk of school dropout among 165,715 Korean high-school students from the national NEIS system. Moreover, in other words, Fernandes et al. [18] presented a study to predict the performance of Bronze Capital students using the Gradient Boosting Machine (GBM) algorithm. Abu Saa et al. [19] conducted a study aimed at identifying the factors influencing student performance — notably previous grades, class results, online learning activity, demographic data, and social information — using data mining algorithms such as decision trees, the naïve Bayes classifier, and artificial neural networks. Conversely, Sekeroglu et al. [20] proposed a prediction and classification system for student performance using five machine learning algorithms, showing that the social and family environment exerts a significant influence on students' results. Moreover, Chavarriaga et al. [21] proposed a recommendation system intended to help students improve their learning skills, showing that their knowledge and social abilities constitute excellent sources for formulating relevant recommendations in online and hybrid courses. Thus, Manouselis et al. [22] highlighted the importance of recommendation systems in technology-enhanced learning (TEL) as well as in other application domains. Thus, Thai-Nghe et al. [23] proposed a student performance prediction system using data mining algorithms such as decision trees, Bayesian networks, and support vector machines, in order to address the class imbalance problem. Furthermore, Thai-Nghe et al. [24] also proposed a comparative study between recommendation-system techniques and traditional regression methods such as logistic regression applied to educational data. The experimental results show that the proposed approach improves predictive performance. Moreover, Romero et al. [25] compared different data mining methods to classify students based on their use of Moodle, and applied discretization and data-rebalancing techniques to assess whether these preprocessing steps allow the development of better classification models. Thus, Nghe et al. [26] used data mining techniques to compare the accuracy of decision-tree algorithms and Bayesian networks in predicting student academic performance. Finally, Bekele and Menzel [27] used data mining techniques to predict student performance, showing that the Bayesian approach applied to the educational domain — and in particular the Bayesian classifier — constitutes a promising tool for predicting students' results.

4. Methodology

In order to solve the aforementioned problem, this paper suggests the construction of an intelligent system using machine learning methods for first-grade secondary students classification based on their grades. Our present approach is different from our previous work [9, 14], where we were concerned with

predicting the Baccalaureate average; recently, in fact, we focus on supervised classification and seek to classify each student into a category of his/her average based on his/her grades obtained in core subjects—fail, passable (suffisant), fairly good (assez bien), good (bien), very good (très bien) or excellent. This is a decision-making support system in order to enhance academic guidance and curtail misplacements.

The approach is based on actual data of students who did the Baccalaureate in Guelmim-Oued Noun. First, the data is organised in a MySQL database as it supports high read performance and concurrent access. Then a Python script is responsible for data pre-processing, model training and performance checking. In fact, due to its readability, modularity and the wealth of dedicated libraries for it, Python is a great choice for artificial intelligence programming.

The model learns from a dataset partitioned into both training/testing/validation sets to obtain stable estimating for the generalization capability. Once the model is trained, it is used in order to classify new students by automatically assigning grade category to them using their Tronc Commun grades. This pipeline mirrors the rigorous methodology demanded in machine learning projects and offers a transparent framework which is reproducible to support educational decision-making.

1. **Data search and preprocessing**

Data pre-processing is the most crucial step for any Data Scientist when starting a Machine Learning project. The dataset for this study is that of the Regional Academy of Education and Training in Guelmim-Oued Noun from southern Morocco. Similar when working with educational data, the raw records had a number of inconsistensies,missing entries and unnecessary details that require preprocessing before models can be developed. The noisy or incomplete data is not carefully cleannes which can introduce unreliable patterns, and will finally reduce the performance of classification model.

After data collection, a considerable pre-analytic effort was performed to get rid of wrong entries, face missing values and filter out records not eligible for the study. After applying this first filter, the final data set contained 72,020 valid student records with course information between 2000 and 2015.

The data were organised according to the students' academic tracks for the current classification analysis. The final cleaned dataset includes:

- **52,595 students** enrolled in literary tracks,
- **19,009 students** enrolled in scientific tracks,
- **416 students** enrolled in technical tracks.

This balanced and structured dataset forms the foundation for building an intelligent system capable of classifying Tronc Commun students into predefined performance categories (fail, passable, fairly good, good, very good, excellent). The ultimate objective of the research is not to predict a numerical

baccalaureate score, but rather to develop a reliable **classification model** that can support academic decision-making and improve student orientation in the region.

1. **Presentation of the Data**

The GestionNotes database holds grades for high school students across three levels: common core, first-year baccalaureate, and second-year baccalaureate.

Table 1
lists the different study paths available, including the specific options students pick in their first year. This setup helps us track student progress over time and check if their early grades match the difficulty of the specialty they chose.

Table 1 High school career.

	Technical	Scientific	Literature
Level I	Technical Core	core science	core Literature
Level II	MST	MS	LSS
	EST	ES	
	ESM		
	MS		
Level III	MST	MSA	L
	EST	MSB	SS
	ES	PS	
	AMS	LES	
	MSB		

Table 1 High school career

Note

MST: Branch of Mechanical Science and Technology; MS: Branch of Mathematics Sciences; LSS: Branch of literature and Social Sciences; EST: Branch of Electrical Science and Technology; ESM: Branch of Economic Science and Management; MSA: Branch of Mathematical Science A; L: Branch of Literature; MSB: Branch of Mathematical Science B; SS: Branch of Social Sciences; ES: Branch of Economics Sciences; PS: Branch of Physical Sciences; AMS: Branch of Accounting Management Sciences; LES: Branch of Life and Earth Sciences.

1. **Select Model**

For this study we decided to stick with the XGBoost algorithm for the classification model. It's known for working well with big varied datasets and has done well in educational data mining before.

As you can see in Fig. 1 the XGBoost is a cleaned-up version of gradient boosting. People choose this algorithm because it can model hard with nonlinear patterns without taking forever to compute.

A big plus is that it has regularization built in, which lowers the chance of overfitting. This is crucial when dealing with real educational data, which often has noise and varies a lot between student groups. That's why we picked XGBoost as the only model to train and see how well it could sort students into different performance groups. especially since we've already worked with this data in the regression section.

By concentrating on one proven algorithm, we keep our method consistent. It also lets us look closer at how XGBoost predicts things, especially for academic choices in the Guelmim-Oued Noun area.

1. **Result**

We trained the XGBoost model on three different areas branch technical, branch literary, and branch scientific and we could compare them fairly and keep things steady. The model worked very well across all areas however the results changed a bit depending on the stream. This difference is normal since each set of info is different when it comes to size, grade distribution, and how much things change inside it.

In the branches of technical, the model got 90% for the accuracy. Even though this is a small group 416 people as shown in Table 2, the algorithm easily saw the main patterns in the results. Classifiers usually struggle with small sets of info.

The branches of literary had the best result, with 92% accuracy. It's not too shocking since this area has the most students 52,595 in total as shown in Table 2. Since the info is quite big and balanced, the model learned clearer and steadier differences between the performance levels.

For the branches of scientific the accuracy was 89%, it's a bit lower but still good. Science subjects usually have scores that change more, and the set of info with 19,009 students as shown in Table 2 has lots of different student types. All this variety makes it harder to classify, but the model still does well.

In general, the accuracy values from 89% to 92% tell us that XGBoost works well across all three areas. These results show that the model can give dependable predictions and might be a helpful way to assess how well students are doing and to help improve guidance in the Guelmim-Oued Noun area.

Table 2
Statistics for branchs

Branch	Number of students	accuracy
Literature	52 595	92%
Science	19 009	89%
Technical	416	90%
Total students	72 020	

Figure 2 shows the performance of the model, with an accuracy of 92% for the literary branch, 89% for the scientific branch, and 90% for the technical branch.

5. Conclusion

The aim of the current research was to create an intelligent classification system for transferring students from common core into their corresponding performance category using machine learning methods. The proposed model, using real educational data from the region of Guelmim-Oued Noun and based on XGBoost algorithm, had good predictive performance for all the streams. We achieve accuracy equal to 89% in the scientific stream, 90% in the technical stream, and 92% in the literary stream demonstrating the strength of our model on heterogeneous datasets.

These results validate the assertion that XGBoost is a very effective algorithm in classifying student performance from common core grades. The model's sensitivity to multifaceted internal relationships enables the model to serve as a useful guide for educational decision-making. Such a form can also be used to detect students at risk of needing extra pedagogical support, lessen the degree of misalignment and increase target efficiency for guidance services in the school area.

School, career, and university guidance in the age of digital transformation is a key lever for supporting learners in building their academic and professional paths. This requires investing in the technological capabilities available within and outside institutions, while ensuring the establishment of a responsible digital culture that balances the exploitation of opportunities with the avoidance of risks. The digital revolution is an effective pedagogical and guidance tool, provided that it is directed in the right direction and achieves the strategic objectives of the educational process in its holistic dimension, with the conscious and responsible involvement of all stakeholders in the process.

In summary, the paper shows that the application of machine learning is a promising and dependable way of assisting academic orientation and handling school performance in the region of Guelmim-Oued Noun. The implementation of such intelligent systems could have an important impact on school failure reduction and in the improvement of alignment between students' performance and their educational traits.

Required Declarations for the Manuscript

1. Corresponding Author

The corresponding author for this manuscript is:

Yousef Farhaoui (corresponding author)

Email: y.farhaoui@fste.umi.ac.ma

2. Funding Declaration

Declarations

Funding:

The authors declare that this research received no external funding.

3. Clinical Trial Number

Clinical trial number: not applicable.

4. Consent to Publish Declaration

Consent to Publish declaration: not applicable.

(This study does not involve human participants, identifiable data, or materials requiring publication consent.)

5. Consent to Participate Declaration

Consent to Participate declaration: not applicable.

(This study uses anonymized educational data provided by the regional academic authority, with no direct involvement of human subjects.)

6. Ethics Declaration

Ethics declaration: not applicable.

(No experiments involving human or animal subjects were performed, and all data used were anonymized administrative records.)

Author Contribution

Tarik Ahajjam: Conceptualization, Methodology, Data curation, Formal analysis, Software, Writing – original draft.Sara Jbaida: Data curation, Validation, Investigation, Writing – review & editing.Amina Hassani: Methodology, Formal analysis, Visualization, Writing – review & editing.Mohammed Moutaib: Supervision, Resources, Validation, Writing – review & editing.Haidar Aissa: Investigation, Data curation, Visualization, Writing – review & editing.Yousef Farhaoui: Conceptualization, Supervision, Project administration, Methodology, Writing – review & editing.

Data Availability

2. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Sun H, Jiang Y, Tang X. Artificial Intelligence-Driven Academic Performance Early Warning System for Engineering Universities: Application of XGBoost-SHAP Algorithm. Int J High Speed Electron Syst. 2025;2540536:1–31. 10.1142/s0129156425405364.
2. Han T, Xu G, Lu W. Examining the Effects of Different Types of Achievement Goal Orientation on Undergraduate Students' Engagement in Distance Learning: The Mediating Effect of Self-Efficacy. Behav Sci. Jan. 2025;15(1):39. 10.3390/bs15010039.
3. Kikunda PB et al. Sep., Predicting First-Year Student Performance with SMOTE-Enhanced Stacking Ensemble and Association Rule Mining for University Success Profiling, *Journal of Computing Theories and Applications*, vol. 3, no. 2, pp. 132–144, 2025, 10.62411/jcta.14043
4. Fu M, Xiao J, Deng T, He X. A hybrid approach integrating Bayesian networks and decision trees for data-driven statistical learning with political education orientation. accessed Nov 28, 2025.
5. Yang H, Lee W, Kim J. Identification of Key Factors Influencing Teachers' Self-Perceived AI Literacy: An XGBoost and SHAP-Based Approach. Appl Sci. 2025;15:4433. https://doi.org/10.3390/app15084433.
6. Kaouache DE et al. The Effect of Individual Entrepreneurial Orientation on Entrepreneurial Intention Among University Students. accessed Nov. 28, 2025.
7. Uysal I et al. Jun., Tracking and Predicting Student Performance Across Different Semesters with Matched Action-State Orientation Surveys and Interventions, 2024, accessed Nov. 28, 2025.
8. Rončević Zubković B, Kolić-Vehovec S, Pahljina-Reinić R. Achievement Goal Orientation Profiles and the Experiences of Gymnasium Students with Digital Technologies in Education, *Psihologijske Teme*, vol. 33, no. 1, pp. 67–89, Apr. 2024, 10.31820/pt.33.1.4
9. Iatrellis O, Savvas IK, Fitsilis P, Gerogiannis VC. A two-phase machine learning approach for predicting student outcomes. Educ Inform Technol. 2021;26(1):69–88.

10. Tarik A, Aissa H, Yousef F. Artificial intelligence and machine learning to predict student performance during the COVID-19. Procedia Comput Sci. 2021;184:835–40.
11. Tomasevic N, Gvozdenovic N, Vranes S. An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers Educ. 2020;143:103676.
12. Ahajjam T, Moutaib M, Aissa H, Azrour M, Farhaoui Y, Fattah M. Predicting Students' Final Performance Using Artificial Neural Networks, in Big Data Mining and Analytics, vol. 5, no. 4, pp. 294–301, December 2022, 10.26599/BDMA.2021.9020030
13. Abu Saa A, Al-Emran M, Shaalan K. Mining student information system records to predict students' academic performance, in *Proc. Int. Conf. Advanced Machine Learning Technologies and Applications*, Cairo, Egypt, 2020, pp. 229–239.
14. Guilherme A. AI and education: The importance of teacher and student relations. AI Soc. 2019;34(1):47–54.
15. Sekeroglu B, Dimililer K, Tuncal K. Artificial intelligence in education: Application in student performance evaluation. Dilemas Contemporáneos: Educación. 2019;7(1):1–21. Política y Valores.
16. Ever YK, Dimililer K, Sekeroglu B. Comparison of machine learning techniques for prediction problems, in *Proc. 33rd Int. Conf. Advanced Information Networking and Applications*, Matsue, Japan, 2019, pp. 713–723.
17. Chung JY, Lee S. Dropout early warning systems for high school students using machine learning. Child Youth Serv Rev. 2019;96:346–53.
18. Fernandes E, et al. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. J Bus Res. 2019;94:335–43.
19. Abu Saa A, Al-Emran M, Shaalan K. Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. Technol Knowl Learn. 2019;24(4):567–98.
20. Sekeroglu B, Dimililer K, Tuncal K. Student performance prediction and classification using machine learning algorithms, in *Proc. 8th Int. Conf. Educational and Information Technology*, Cambridge, U.K., 2019, pp. 7–11.
21. Chavarriaga O, Florian-Gaviria B, Solarte O. A recommender system for students based on social knowledge and assessment data of competences, in *Proc. 9th European Conf. Open Learning and Teaching in Educational Communities*, Graz, Austria, 2014, pp. 56–69.
22. Manouselis N, Drachsler H, Vuorikari R, Hummel H, Koper R. Recommender systems in technology enhanced learning. In: Ricci F, Rokach L, Shapira B, Kantor PB, editors. in Recommender Systems Handbook. Boston, MA: Springer; 2011. pp. 387–415.
23. Thai-Nghe N, Drumond L, Krohn-Grimberghe A, Schmidt-Thieme L. Recommender system for predicting student performance. Procedia Comput Sci. 2010;1(2):2811–9.
24. Thai-Nghe N, Busche A, Schmidt-Thieme L. Improving academic performance prediction by dealing with class imbalance, in *Proc. 9th IEEE Int. Conf. Intelligent Systems Design and Applications*, Pisa, Italy, 2009, pp. 878–883.

25. Romero C, Ventura S, Espejo PG, Hervas C. Data mining algorithms to classify students, in *Proc. 1st Int. Conf. Educational Data Mining*, Montréal, Canada, 2008, pp. 8–17.
26. Nghe NT, Janecek P, Haddawy P. A comparative analysis of techniques for predicting academic performance, in *Proc. 37th IEEE Frontiers in Education Conf.*, Milwaukee, WI, USA, 2007, pp. T2G-7–T2G-12.
27. Bekele R, Menzel W. A Bayesian approach to predict performance of a student (BAPPS): A case with Ethiopian students, in *Proc. 23rd Int. Conf. Artificial Intelligence and Applications*, Innsbruck, Austria, 2005, pp. 189–194.

Figures

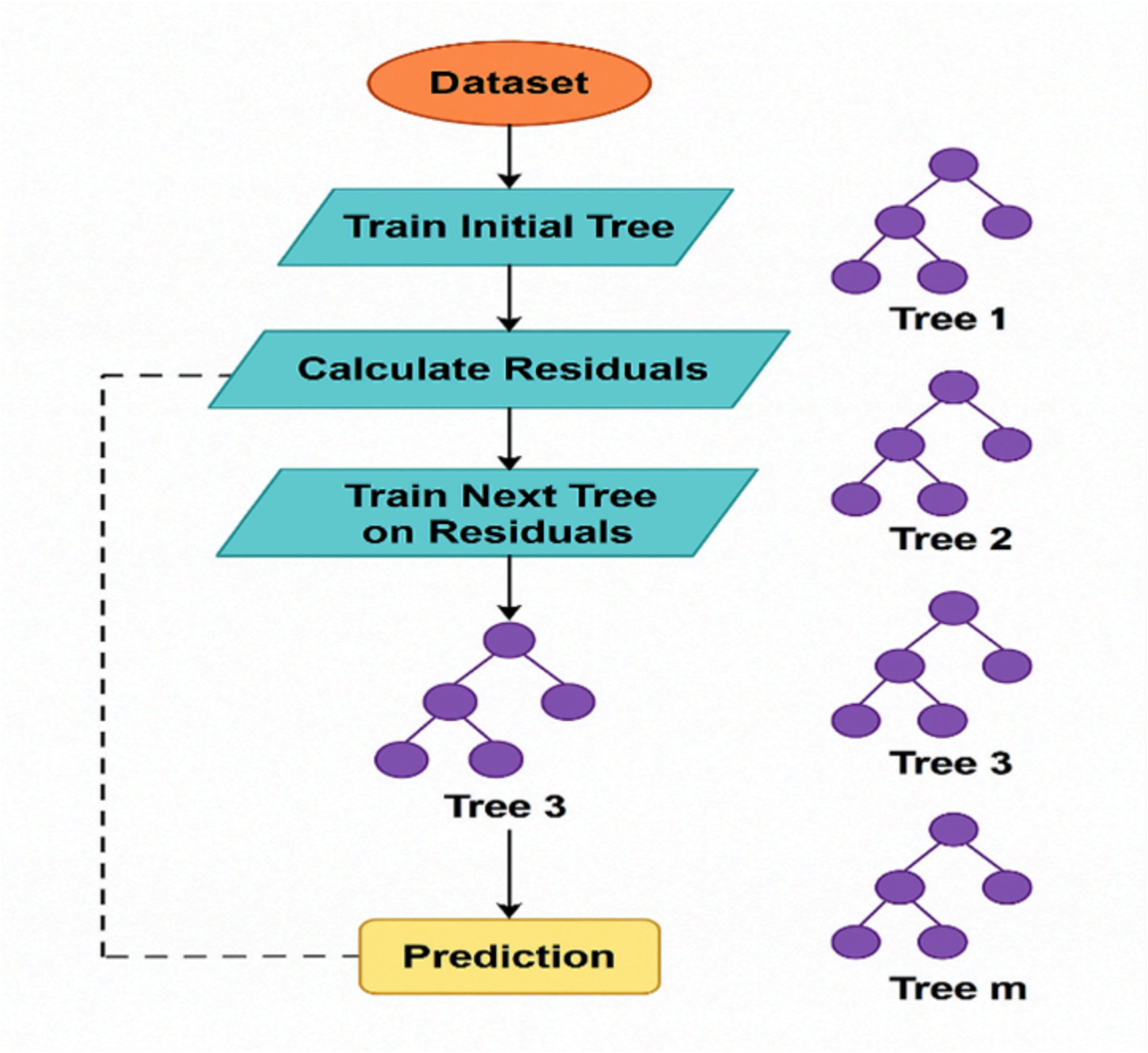


Figure 1

Schéma général du fonctionnement de XGBoost basé sur l'entraînement séquentiel d'arbres via le gradient boosting.

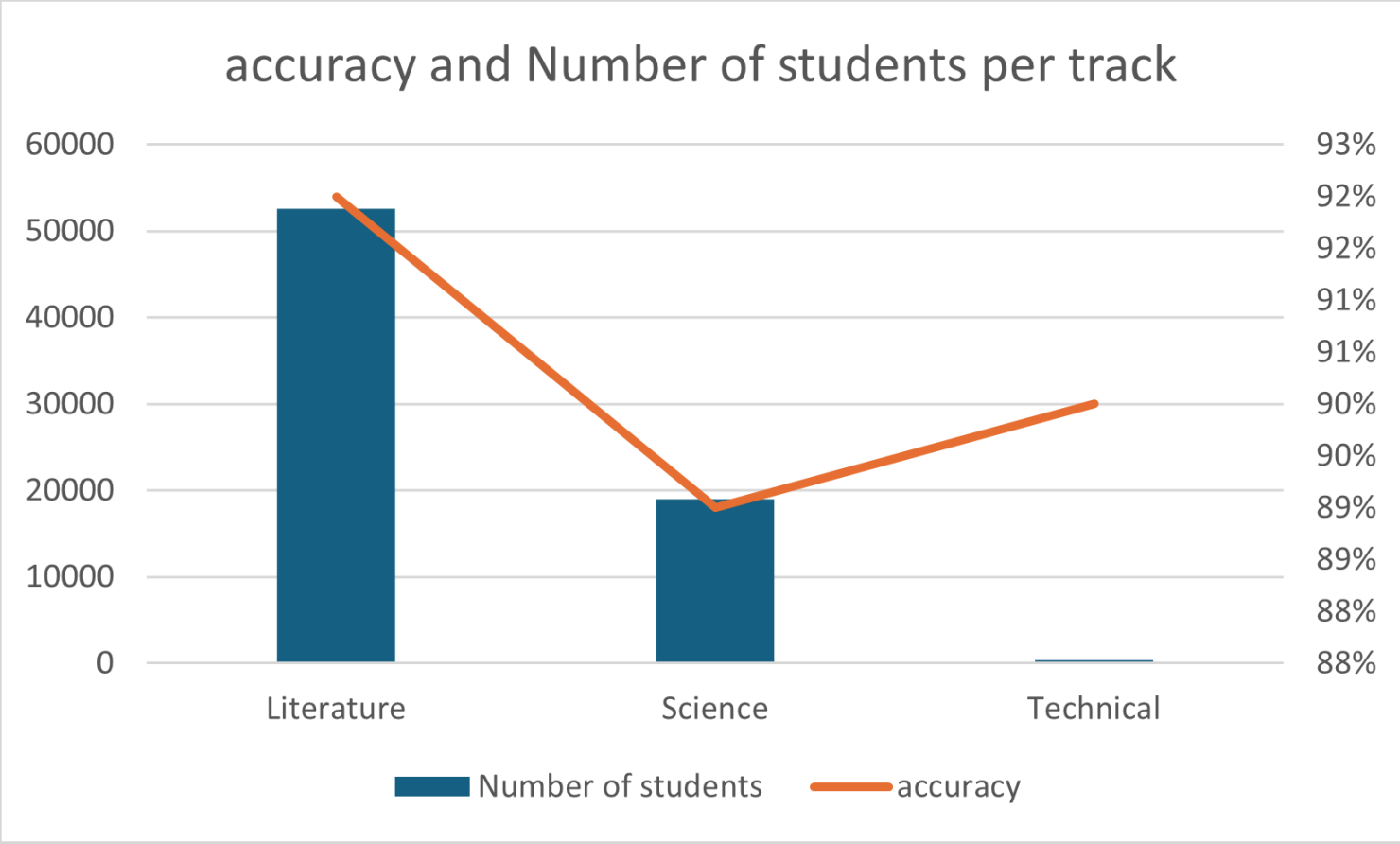


Figure 2

Model Accuracy Across Academic Streams.