Machine learning-based classification of father involvement among adolescents using psychosocial data

Egi Dio Bagus Sudewo[1]*, Annisa Fitri Mulyani[2], Muhammad Kunta Biddinika[3], Reni Yunita[1]

[1]Universitas Royal, Prof. H. M. Yamin St. No.173, Asahan, North Sumatra, 21211, Indonesia

[2]Universitas Sumatera Utara, Dr. T. Mansur St. No.9, Medan, North Sumatra, 20155, Indonesia

[3]Universitas Ahmad Dahlan, Ringroad Selatan St., Bantul, Yogyakarta, Special Region of Yogyakarta, 55191, Indonesia

*Corresponding author, email: egidio8gali@gmail.com

Article history
Submitted: 22 November 2025
Revised: 22 January 2026
Accepted: 27 January 2026
Published: 3 February 2026

Keywords
Adolescents
Decision tree
Father involvement
Logistic regression
Random forest

Abstract
Father involvement (FI) is a key determinant in adolescents' cognitive, emotional, social, and academic development. However, variations in the quality of father child interactions necessitate accurate classification approaches. This study proposes a machine learning (ML) framework to classify adolescents' FI levels using psychosocial data. A validated Likert-scale FI questionnaire was administered to 195 twelfth-grade students. Two labeling schemes were applied: expert-based psychological categories and score-based classification (10-15 = low, 16-20 = high). Three ML algorithms Logistic Regression, Decision Tree, and Random Forest were evaluated using Python. The results indicate that Logistic Regression achieved 100% accuracy across both labeling schemes, while Random Forest and Decision Tree attained accuracy between 92.31% and 97.44%. The perfect accuracy in Logistic Regression highlights a strong linear correlation between the 20-item psychosocial features and the target labels within this specific dataset. This study provides a robust predictive framework that fills the gap in automated father-involvement assessment, offering a more precise tool for psychological evaluation compared to conventional statistical methods.

1. Introduction

The role of fathers (Father Involvement/FI) in adolescent development has gained increasing attention in modern developmental psychology. FI is not only associated with cognitive development but also influences self-esteem, emotional regulation, social skills, and academic achievement in adolescents [1]. Cross-cultural studies have shown that high paternal involvement enhances adolescents' psychological well-being and resilience, while low involvement is often linked to behavioral problems, low academic motivation, and an increased risk of emotional disorders [2]. However, field observations reveal significant variations: some adolescents report low levels of FI, while others report high levels. This indicates a gap in the quality of father–child relationships that warrants further investigation.

Previous studies, such as one conducted at senior high school Muhammadiyah 2 Medan, revealed differences in self-esteem based on FI levels. The findings indicated that students with higher FI reported stronger self-esteem compared to those with lower FI [3]. However, this research relied on conventional statistical methods, which were limited in providing more comprehensive predictive models. This limitation highlights the need for follow-up studies that integrate traditional quantitative approaches with modern computational methods. To date, very few empirical studies have developed FI classification models using Machine Learning, despite its considerable potential to enhance predictive accuracy.

Machine Learning (ML)[4][5], as a branch of modern informatics, offers a wide range of algorithms capable of processing questionnaire data into predictive models [6]. Algorithms such as Logistic Regression [7], Random Forest [8], Decision Tree [9]. are well known for their strong performance in classification tasks, including psychosocial data [10]. The use of two labeling approaches original psychological categories and score-based labels (10–15 for low, 16–20 for high) represents an important strategy for comparing the effectiveness of models in classifying FI. Recent studies emphasize that the integration of ML with adolescent psychology data can enhance the quality of assessment and data-driven intervention recommendations [11]. Thus, applying ML in FI research holds significant potential to enrich both psychological and educational literature.

Empirical evidence supporting the use of machine learning in adolescent psychology has been documented in several cross-cultural studies. Cross-cultural research integrating theoretical approaches and machine

© 2026 The Author(s)
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3110-1577

learning has shown that preadolescent risk factors, particularly internalizing and externalizing behaviors, are the most significant predictors of adolescent mental health problems [12]. Systematic reviews also confirm that algorithms such as logistic regression, Decisiom Tree, Random Forest, and ANN provide reliable predictive capabilities in the context of adolescent psychology [13]. Within the family context, father involvement has been shown to play a significant role in reducing children's problematic behaviors, although further exploration is still needed in cross-cultural aspects [14]. Another study applied the elastic net algorithm to identify factors influencing the need for and utilization of mental health services among at-risk adolescents, highlighting the importance of parental stress and children's clinical problems [15]. In addition, machine learning techniques have been used to predict adolescent involvement in family conflict, finding that factors such as emotional security, responses to conflict, and family dynamics hold relevant predictive capacity [16]. A critical synthesis of previous studies reveals that while algorithms like ANN or Elastic Net provide reliable capabilities, their accuracy in adolescent psychology contexts varies significantly between 60% and 78%. This study fills the gap by specifically integrating two labeling schemes expert-based and score-based to validate the consistency of ML performance, a methodological rigor that was often overlooked in previous descriptive reviews.

While extensive literature has documented the importance of father involvement (FI) in adolescent development, most existing studies rely on descriptive statistics or simple correlational analyses, which lack predictive capabilities for individual assessment. This research fills the gap by implementing a machine learning (ML) framework to provide a more systematic, automated, and high-precision mapping of psychosocial data. The novelty of this study lies in its integration of dual labeling schemes expert-based and score-based to ensure model robustness. By bridging developmental psychology and informatics, this study statistically highlights a new methodological contribution to the field, offering a more scalable assessment tool than previously established in paternal involvement research.

2. Method

This research method is designed to provide a systematic and measurable framework for analyzing the level of Father Involvement (FI) through the application of Machine Learning algorithms. The use of a quantitative approach with a comparative-experimental design is grounded in the need to empirically examine performance differences across classification models. This methodological choice is also supported by previous studies, which highlight the effectiveness of algorithms such as Logistic Regression, Random Forest, and Decision Tree in quantitative psychology and applied social sciences, particularly due to their capability to manage categorical data and generate objective evaluations. By integrating validated psychological instruments with advanced computational techniques, this study aims to produce academically relevant insights while offering practical contributions to the mapping of fathers' involvement in their children's education, the research flow can be seen in Figure 1.

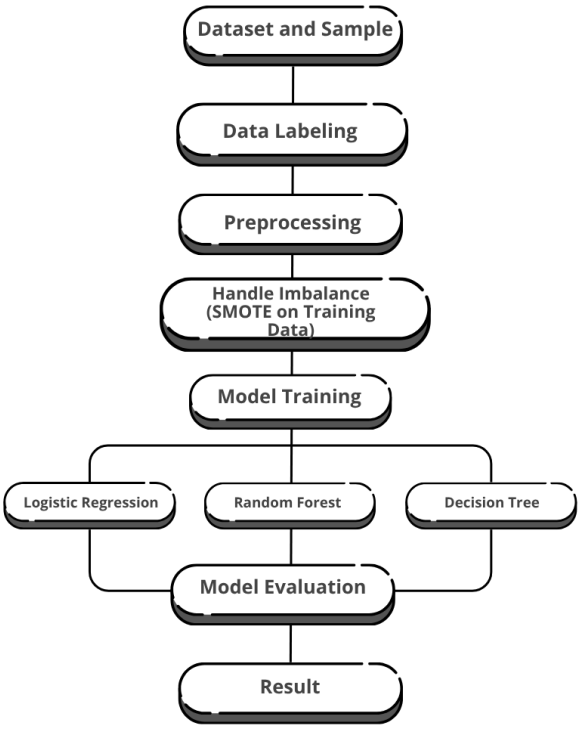

Figure 1. Research flow

© 2026 The Author(s)
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3110-1577

2.1. Research design

This study employs a quantitative research method with a comparative machine learning study. The primary objective is to compare the performance of three Machine Learning algorithms Logistic Regression, Random Forest, and Decision Tree in classifying Father Involvement (FI) data. The psychological variable FI is treated as a case study dataset, allowing the research to focus on algorithmic performance rather than psychological interpretation.

2.2. Dataset and sample

Data for this study were gathered using a Father Involvement questionnaire administered to high school students in their 12th grade. The sampling technique employed was purposive sampling, a non-probability method chosen to select a representative sample based on specific criteria. The primary inclusion criterion for this research was students who live with their biological father.

An initial screening was conducted to identify the father involvement levels within the population that met the criterion. This screening resulted in a final sample of 195 students, comprising 38 students classified with low father involvement and 157 students with high father involvement. This class distribution highlights a significant imbalance, which was addressed during the data preprocessing stage. The structured dataset was then compiled into a CSV file for computational analysis. To ensure transparency and reproducibility, the anonymized dataset is publicly available at https://github.com/egidio23/Dataset-Father-involvement.

2.3. Data labeling

The Father Involvement (FI) scale used in this study was developed based on the theoretical aspects proposed by Doherty: engagement, accessibility, and responsibility. The instrument was designed as a Guttman scale, consisting of 20 statements with dichotomous response options: "Ya" (Yes) or "Tidak" (No). The statements included both favorable items (where "Yes" indicates higher involvement) and unfavorable items (where "No" indicates higher involvement).

To comprehensively evaluate the classification models, two distinct labeling schemes were applied:

a. Expert-based psychological categories (Original FI): This labeling scheme corresponds to the results of the initial screening process. Based on their responses to the Guttman scale, a classification was made that resulted in 157 students being labeled as 'High' FI and 38 students as 'Low' FI. This label represents the initial ground truth classification used in the study.

b. Score-based classification (Calculated FI): The second scheme involved a quantitative scoring method. Each respondent's answers were scored, where a response indicating positive involvement (e.g., "Yes" to a favorable item) received 1 point and a response indicating low involvement received 0 points. The total score was calculated by summing the points for all items. Based on the total score, a new set of labels was generated using predefined thresholds as described in the original questionnaire design:

    a) Low: Respondents with a total score between 10 and 15.

    b) High: Respondents with a total score between 16 and 20.

This dual-labeling approach allowed for a comparative analysis of model performance on labels derived from an initial grouping versus labels derived from a strict score-based threshold.

2.4. Research procedure

The research was executed through a structured pipeline encompassing feature engineering, data preprocessing, model development, and evaluation. Each stage was designed to ensure the robustness and reproducibility of the results.

2.4.1. Feature engineering and preprocessing

The features for the machine learning models were derived directly from the binary responses of the Guttman scale questionnaire. Each of the 20 items on the scale was treated as a distinct feature. The "Ya" (Yes) and "Tidak" (No) responses were numerically encoded into 1 and 0, respectively. Consequently, each student was represented by a feature vector of length 20 consisting of binary values.

The preprocessing pipeline began with splitting the dataset into an 80% training set and a 20% testing set. To mitigate potential model bias towards the majority class, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. Crucially, SMOTE was applied only to the training data after the split. This practice ensures that the model is trained on balanced data while the test set remains an authentic, unseen

© 2026 The Author(s)
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3110-1577

representation of the original data distribution, thus preventing data leakage and providing a realistic evaluation of the model's generalization performance.

2.4.2. Model development

Three distinct classification algorithms were implemented and compared in this study:

a. Logistic Regression was chosen for its efficiency and interpretability in modeling linear relationships between independent variables and class probabilities, making it a strong baseline for binary classification tasks [17].

b. Decision Tree was selected for its simplicity and ability to generate transparent, rule-based classification models through recursive data partitioning, making it highly interpretable [18].

c. Random Forest, an ensemble method, was used for its robustness against overfitting and its ability to enhance prediction stability by aggregating the results of multiple decision trees [19].

The models were implemented using the scikit-learn library in Python. The key hyperparameters for each model are detailed in Table 1. Additionally, k-fold cross-validation was used during the development phase to ensure model reliability and prevent overfitting.

Table 1. Parameter comparison

Model	Hyperparameter	Score
Logistic regression	max_iter	1000
	solver	lbfgs
Decision tree	criterion	gini
	max_depth	None
	min_samples_split	2
Random forest	n_estimators	100
	criterion	gini
	max_depth	None
	bootstrap	True

2.5. Model evaluation

The final stage was model evaluation, conducted using Accuracy, Precision, Recall, and F1-score as performance metrics [20]. These measures were employed to assess the ability of each algorithm in classifying FI data and to compare their effectiveness across expert-based.

3. Results and Discussion

3.1. Result

The analysis of father involvement was carried out using the Python programming language with the support of several popular libraries in the field of data science and machine learning. Data processing was performed with pandas and numpy, while visualization was supported by matplotlib and seaborn. Data splitting, model training, and evaluation were conducted using the scikit-learn library, which provides various classification algorithms such as Logistic Regression, Decision Tree, and Random Forest. In addition, class imbalance handling was addressed using SMOTE (Synthetic Minority Oversampling Technique) available in the imbalanced-learn library. All experiments were Based on the evaluation of classification models for father involvement (Father Involvement), the results are presented in Table 2. In general, all models demonstrated excellent performance, with test accuracy above 92% in both scenarios, whether using the original label (Original FI) or the recalculated score label (Calculated FI). executed in the interactive programming environment of Google Colaboratory.

Table 2. Result

Scenario	Model	Train Accuracy	Test Accuracy	Precision	Recall	F1
Original FI	Logistic Regression	1.0	1.0	1.0	1.0	1.0
	Decision Tree	1.0	0.923	0.938	0.967742	0.952
	Random Forest	1.0	0.949	0.939	1.0	0.969
Calculated FI	Logistic Regression	1.0	1.0	1.0	1.0	1.0
	Decision Tree	1.0	0.949	0.939	1.0	0.969
	Random Forest	1.0	0.974	0.968	1.0	0.984

© 2026 The Author(s)
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3110-1577

In the Original FI scenario, Logistic Regression achieved a perfect test accuracy of 100% with equally flawless precision, recall, and f1-score. Random Forest attained a test accuracy of 94.87% with an f1-score of 0.969, while Decision Tree recorded a test accuracy of 92.31% with an f1-score of 0.952. Meanwhile, in the Calculated FI scenario, Logistic Regression once again reached perfect performance with 100% accuracy. Random Forest showed an improvement in accuracy to 97.44% with an f1-score of 0.984, and Decision Tree remained stable with a test accuracy of 94.87% and an f1-score of 0.969. Original FI confusion matrix can be seen in Figure 2.

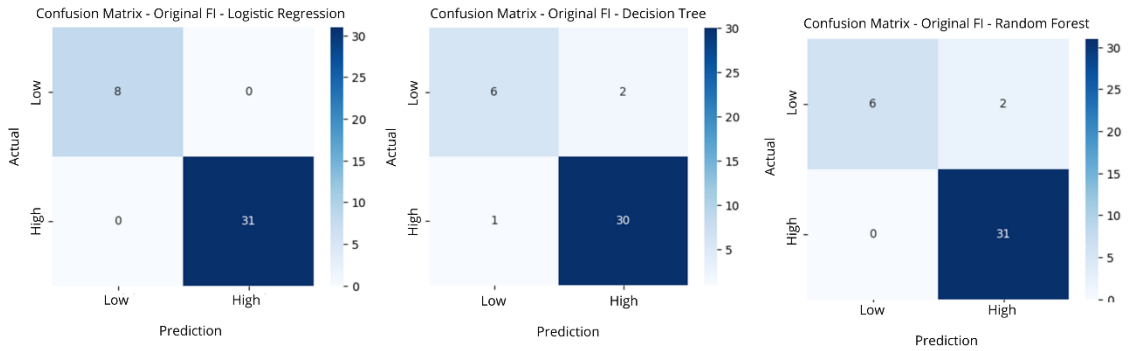

Figure 2. Original FI Confusion Matrix

The Figure 2 presents three confusion matrices comparing the performance of Logistic Regression, Decision Tree, and Random Forest models in classifying adolescents' Father Involvement (FI) levels into Low and High categories using the original labeling scheme. Logistic Regression achieved perfect classification, correctly identifying all eight Low and thirty-one High cases without any errors. The Decision Tree model misclassified two Low cases as High and one High case as Low, resulting in a slight accuracy drop to about 92%. Random Forest improved on this, with only two Low cases misclassified as High and no errors for the High category, yielding roughly 95% accuracy. The dominance of High-FI samples and the flawless performance of Logistic Regression suggest that the relationship between the questionnaire features and FI levels is largely linear, while Random Forest provides a robust alternative with better generalization than a single Decision Tree when class distributions or data complexity increase. Calculated FI confusion matrix can be seen in Figure 3.

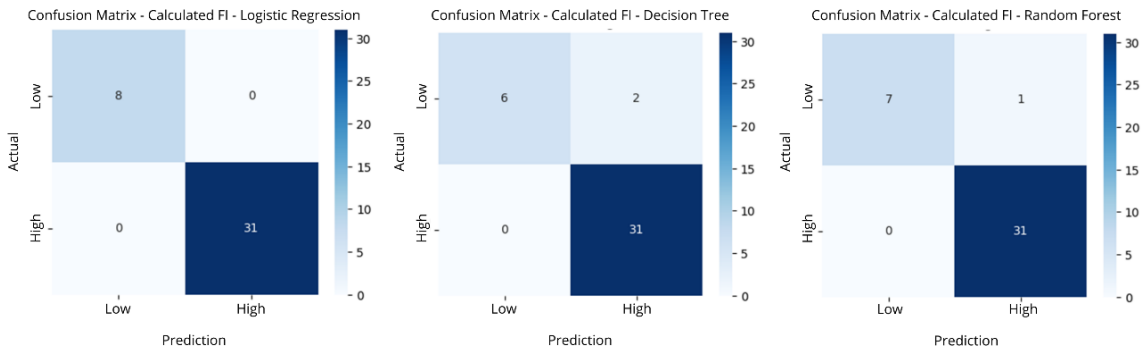

Figure 3. Calculated FI Confusion Matrix

The Figure 3 shows three confusion matrices illustrating the performance of Logistic Regression, Decision Tree, and Random Forest classifiers when predicting adolescents' Father Involvement (FI) levels using the calculated FI labels (Low vs. High). In the first matrix, Logistic Regression achieved perfect predictions: all eight Low and thirty-one High cases were classified correctly, yielding 100 % accuracy. The second matrix, for the Decision Tree, reveals two Low cases incorrectly predicted as High, while all High cases were correctly classified, giving roughly 95 % accuracy. The third matrix shows Random Forest with one Low case misclassified as High and perfect classification of all High cases, resulting in an accuracy close to 97 %. Overall, these results confirm that all models perform strongly under the calculated labeling scheme, with Logistic Regression again providing flawless separation, while Random Forest offers slightly better generalization than a single Decision Tree when minor non-linear patterns are present. Next, model accuracy comparison and comparison of precision, recall, and F1 can be seen in Figure 4.

© 2026 The Author(s)
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3110-1577

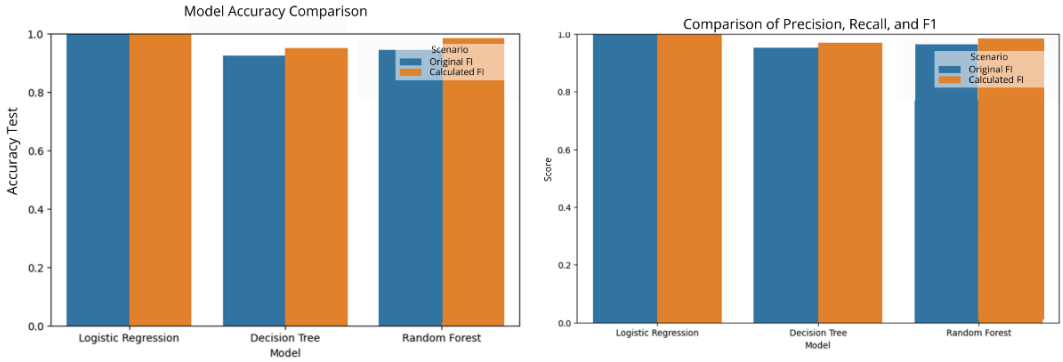

Figure 4. Model Accuracy Comparison and Comparison of Precision, Recall, and F1

The Figure 4 presents two side-by-side bar charts comparing model performance under the Original FI and Calculated FI labeling schemes. The left panel, Model Accuracy Comparison, shows that Logistic Regression achieves perfect accuracy in both scenarios, while the Decision Tree improves slightly with Calculated FI (about 0.92 to 0.95) and Random Forest also gains a modest increase (approximately 0.95 to 0.97). The right panel, Comparison of Precision, Recall, and F1, illustrates that all models maintain consistently high scores close to 1.0 regardless of labeling method. Logistic Regression remains flawless across all metrics, whereas Decision Tree and Random Forest display only minor variations. Collectively, the charts emphasize that labeling choice does not significantly influence model performance, confirming Logistic Regression's perfect classification and the strong, stable performance of tree-based methods. Next, comparison of train and test can be seen in Figure 5.

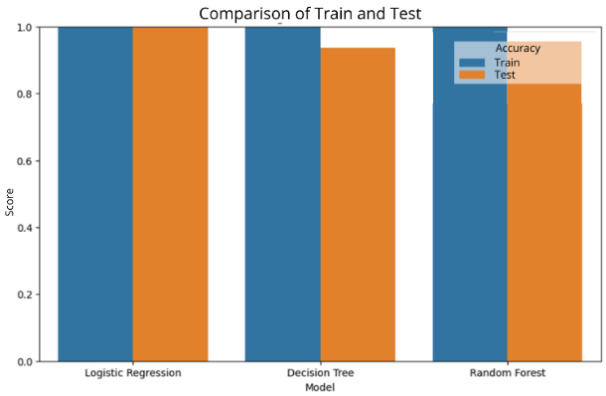

Figure 5. Comparison of Train and Test

Figure 5 compares model accuracy on training and testing data for Logistic Regression, Decision Tree, and Random Forest. Each model shows two bars: blue for training accuracy and orange for testing accuracy. Logistic Regression achieves a perfect score of 1.0 on both training and test sets, demonstrating excellent generalization with no signs of overfitting. Decision Tree also reaches 1.0 on the training set but drops slightly to around 0.93 on the test set, indicating mild overfitting as it fits the training data more closely than the unseen data. Random Forest maintains perfect accuracy on the training data while its test accuracy is somewhat lower, near 0.95, reflecting a more balanced performance and better generalization than the single Decision Tree. Overall, the figure highlights that all three algorithms perform strongly, with Logistic Regression remaining the most stable. The perfect accuracy achieved by Logistic Regression suggests that the dataset may possess high linear separability, where questionnaire features clearly distinguish between FI levels. However, utilizing ML in psychological assessment carries inherent risks, such as over-reliance on automated scores without clinical intuition. Furthermore, this study is limited by its sample size and potential self-report bias, necessitating future research in broader cultural contexts to enhance model generalizability. Next, Figure 6 shown a ROC curve original and calculated F1.

© 2026 The Author(s)
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3110-1577

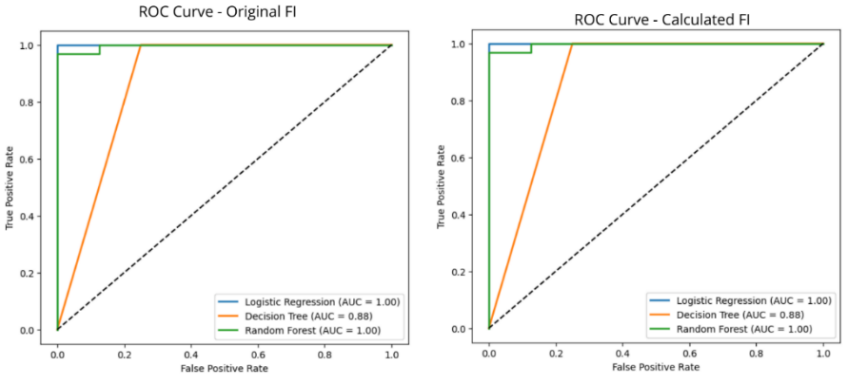

Figure 6. ROC Curve Original and Calculated F1

Figure 6 compares model performance in predicting Father Involvement (FI) levels under two labeling schemes, Original FI (left) and Calculated FI (right). Both Logistic Regression (blue) and Random Forest (green) produce nearly perfect curves hugging the top-left corner with an AUC of 1.00, indicating excellent discrimination between Low and High FI classes. In contrast, the Decision Tree (orange) yields a lower AUC of 0.88, though it still performs as a strong classifier. The consistency across both panels confirms that the choice of labeling has no impact on model ranking or predictive power, reinforcing Logistic Regression and Random Forest as the most reliable methods for this classification task.

3.2. Discussion

These findings show that the Father Involvement (FI) dataset is relatively easy to classify, as indicated by the strong performance across all models. The flawless results achieved by Logistic Regression highlight a clear linear relationship between the psychosocial variables and FI categories, suggesting high linear separability within the current dataset. However, these results must be interpreted with caution. This study is constrained by several limitations, including a relatively small sample size (n=195) which may affect the model's broader applicability. Furthermore, the data relies on self-report questionnaires, which are susceptible to social desirability or self-report bias from the adolescent respondents. The findings are also grounded in a specific cultural context in Indonesia, which may limit their generalizability to different demographic or cultural settings. Future research should aim to validate these models using larger and more diverse datasets to enhance the reliability of ML-based father involvement assessments.

However, these results must be interpreted with caution. The high accuracy may reflect the specific characteristics of the dataset, and the application of ML in psychological assessment carries inherent risks, such as over-reliance on automated scores without clinical intuition. Furthermore, this study is constrained by a relatively small sample size and potential self-report bias, necessitating future research in broader cultural contexts to enhance model generalizability. In conclusion, while all three models are suitable, Logistic Regression is recommended as the primary method due to its simplicity, stability, and consistently optimal outcomes in classifying father involvement levels. Table 3 shown a comparison with previous research.

Table 3. Comparison with Previous Research

Researcher & Year	Title	Method	Sample	Findings
Rothenberg et al. (2023)	Predicting Adolescent Mental Health Outcomes Across Cultures: A Machine Learning Approach	Integration of theory (bioecological) and machine learning (79 variables, classification)	1176 families (9 countries)	Internalizing & externalizing behaviors at age 10 were the strongest predictors of mental health problems at ages 13 & 17; accuracy 60.6–78%
Mena et al. (2022)	Machine Learning Techniques Used in Psychological Studies with Adolescents: A Systematic Review	Systematic review (PRISMA)	5 studies from international databases	ML (logistic regression, SVM, RF, ANN, XGB) proved comparable or superior to traditional statistical methods
Nafisah, (2023)	The Impact of Father Involvement in Early Childhood Problematic Behavior	Systematic Literature Review (SLR)	10 articles (2013–2023)	Father involvement negatively correlated with children's problematic behavior; cross-cultural studies are needed
Vázquez et al. (2021)	High-Value Correlates of Caregiver-Reported Counseling Service Need and Utilization for At-Risk Adolescents	Machine learning (elastic net)	878 adolescent–caregiver dyads	Need and utilization of MHS predicted with 71–77.6% accuracy; main factors: parental stress & child clinical problems

© 2026 The Author(s)
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3110-1577

Researcher & Year	Title	Method	Sample	Findings
Lopez et al. (2022)	Using Machine Learning Techniques to Predict Adolescents' Involvement in Family Conflict	Machine learning (family conflict prediction classification)	251 Spanish adolescents (residential care & family)	Accuracy 65%; main predictors: emotional security, conflict responses, triangulation, and family dynamics
This research	Machine Learning Approach for Classifying Father Involvement Levels among Adolescents Using Psychosocial Data	Machine learning (Logistic Regression, Decision Tree, Random Forest; two labeling schemes)	197 twelfth-grade students (Likert-scale FI questionnaire)	Logistic Regression achieved 100% accuracy; Random Forest 94.87–97.44% (f1 ≈ 0.98); Decision Tree 92.31–94.87%. ML provides accurate FI classification with practical implications for counselors, teachers, and psychologists

Previous studies in Table 3 has shown that the application of Machine Learning (ML) in adolescent psychology is increasingly developing, both through cross-cultural predictive approaches (Rothenberg et al., 2023) and systematic reviews that emphasize the superiority of ML algorithms over conventional statistical methods (Mena et al., 2022). Research on father involvement has largely been conducted through literature-based approaches, such as Nafisah (2023), who found a negative correlation between father involvement and children's problematic behavior, yet predictive modeling was rarely employed. Meanwhile, studies by Vázquez et al (2021) and Lopez et al (2022) applied ML algorithms to predict mental health service needs and adolescents' involvement in family conflict, achieving accuracies ranging from 65% to 77.6%. %. Compared to these works, This Research introduces a methodological innovation by classifying father involvement using adolescents' psychosocial data through three ML algorithms. The findings demonstrate higher performance, with Logistic Regression achieving perfect accuracy (100%), Random Forest reaching 94.87–97.44%, and Decision Tree ranging from 92.31% to 94.87%. These results confirm that integrating ML with psychosocial data not only enriches academic literature but also provides practical implications for accurately mapping father involvement to support educational and counseling.

4. Conclusion

This study demonstrates that the application of Machine Learning (ML) algorithms in classifying father involvement (FI) among adolescents can produce highly accurate and consistent results. Logistic Regression proved to be the most effective model, achieving perfect accuracy (100%) across both labeling schemes, namely expert-based labels and score-based classifications. Random Forest also showed strong performance, reaching up to 97.44% accuracy with an f1-score close to 0.98, while Decision Tree remained stable with accuracy ranging from 92.31% to 94.87%. These findings confirm that integrating psychosocial data with modern computational approaches enables more systematic and data-driven mapping of FI. Theoretically, this research enriches the interdisciplinary literature between psychology and informatics, while practically, the results can be utilized by school counselors, teachers, and psychologists to more precisely identify levels of father involvement in order to support adolescent development.

Author Contributions

All authors have equal contributions to the paper. All the authors have read and approved the final manuscript.

Funding

No funding support was received.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/ or publication of this article.

Data Availability

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

Declaration on AI Use

The authors declare that no artificial intelligence (AI) or AI-assisted tools were used in the preparation of this manuscript.

© 2026 The Author(s)
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3110-1577

References

[1] Y. Sulistyawati, "The Effect of Perceived Father Involvement on Emotional Intelligence in Adolescents," *Dev. Clin. Psychol.*, 2024, doi: 10.15294/dcp.v5i2.17097.

[2] C. S. Walsh, W. Kliewer, and T. N. Sullivan, "Adolescents' Subjective Well-Being: The Unique Contribution of Fathers," 2024, doi: 10.1007/s10566-024-09801-z.

[3] A. F. Mulyani, "Perbedaan Self Esteem Ditinjau dari Peran Father Involvement Pada Remaja di SMA Swasta Muhammadiyah 2 Medan," 2023. [Online]. Available: https://repositori.uma.ac.id/handle/123456789/21678%0Ahttps://repositori.uma.ac.id/bitstream/123456789/21678/2/198600215 - Annisa Fitri Mulyani - Fulltext.pdf

[4] P. A. Pramesti *et al.*, "Classification of Clove Leaf Blister Blight Disease Severity Using Pre-trained Model VGG16, InceptionV3, and ResNet," *J. Ilmu Komput. dan Inf.*, vol. 17, no. 2, pp. 159–166, 2024, doi: 10.21609/jiki.v17i2.1237.

[5] R. Angger Saputra and Y. Sibaroni, "Multilabel Hate Speech Classification in Indonesian Political Discourse on X using Combined Deep Learning Models with Considering Sentence Length," *J. Ilmu Komput. dan Inf.*, vol. 18, no. 1, pp. 113–125, 2025, doi: 10.21609/jiki.v18i1.1440.

[6] P. Dehghan, H. Alashwal, and A. A. Moustafa, "Applications of machine learning to behavioral sciences: focus on categorical data," *Discov. Psychol.*, vol. 2, no. 1, p. 22, Dec. 2022, doi: 10.1007/s44202-022-00027-5.

[7] Kehinde Josephine Olowe, Ngozi Linda Edoh, Stephane Jean Christophe Zouo, and Jeremiah Olamijuwon, "Comprehensive review of logistic regression techniques in predicting health outcomes and trends," *World J. Adv. Pharm. Life Sci.*, vol. 7, no. 2, pp. 016–026, Nov. 2024, doi: 10.53346/wjapls.2024.7.2.0039.

[8] M. A. Azkiya and Y. N. Ifriza, "Analysis of Coding Stress Impact on Students Programming Skills with Random Forest and C4.5 Algorithms," *J. Ilmu Komput. dan Inf.*, vol. 18, no. 2, pp. 167–180, 2025, doi: 10.21609/jiki.v18i2.1487.

[9] F. AABOUB, H. CHAMLAL, and T. OUADERHMAN, "Analysis of the prediction performance of decision tree-based algorithms," in *2023 International Conference on Decision Aid Sciences and Applications (DASA)*, IEEE, Sep. 2023, pp. 7–11. doi: 10.1109/DASA59624.2023.10286809.

[10] T. Yamamoto *et al.*, "Clinical psychoinformatics," in *Methodological Approaches for Sleep and Vigilance Research*, Elsevier, 2022, pp. 255–279. doi: 10.1016/B978-0-323-85235-7.00013-2.

[11] M. J. Burger and N. T. P. Pang, "Well-being in Asia," *Front. Psychol.*, vol. 16, p. 1661988, Aug. 2025, doi: 10.3389/fpsyg.2025.1661988.

[12] W. A. Rothenberg *et al.*, "Predicting Adolescent Mental Health Outcomes Across Cultures: A Machine Learning Approach," *J. Youth Adolesc.*, vol. 52, no. 8, pp. 1595–1619, Aug. 2023, doi: 10.1007/s10964-023-01767-w.

[13] A. J. Mena, N. De Oliveira Cardoso, C. E. Xavier, and I. I. De Lima Argimon, "Técnicas de Machine Learning utilizadas en estudios psicológicos con adolescentes: Una revisión sistemática," *Edupsykhé. Rev. Psicol. y Educ.*, vol. 19, no. 2, pp. 23–38, Nov. 2022, doi: 10.57087/edupsykhe.v19i2.4440.

[14] Aisyah Durrotun Nafisah, Yuli Kurniawati Sugiyo Pranoto, and Siti Nuzulia, "The Impact of Father Involvement in the Early Childhood Problematic Behavior," *JPUD - J. Pendidik. Usia Dini*, vol. 17, no. 1, pp. 14–30, Apr. 2023, doi: 10.21009/JPUD.171.02.

[15] A. L. Vázquez, T. Chou, C. M. Navarro Flores, T. S. Barrett, M. T. Villodas, and M. M. Domenech Rodríguez, "High value correlates of caregiver reported counseling service need and utilization for adolescents at-risk for childhood maltreatment and neglect," *PLoS One*, vol. 16, no. 10, p. e0258082, Oct. 2021, doi: 10.1371/journal.pone.0258082.

[16] S. Lopez-Larrosa *et al.*, "Using Machine Learning Techniques to Predict Adolescents' Involvement in Family Conflict," *Soc. Sci. Comput. Rev.*, vol. 41, no. 5, pp. 1581–1607, Oct. 2023, doi: 10.1177/08944393221084064.

[17] Q. Ma, "Recent applications and perspectives of logistic regression modelling in healthcare," *Theor. Nat. Sci.*, vol. 36, no. 1, pp. 185–190, Jul. 2024, doi: 10.54254/2753-8818/36/20240614.

[18] K. Tajima, N. Ichijo, Y. Nakahara, and T. Matsushima, "An Algorithmic Framework for Constructing Multiple Decision Trees by Evaluating Their Combination Performance Throughout the Construction Process," *arXiv.Org*, vol. abs/2402.0, 2024, doi: 10.48550/arxiv.2402.06452.

[19] A. Moumen, A. Shafqat, T. Alraqad, E. S. Alshawarbeh, H. Saber, and R. Shafqat, "Divorce prediction using machine learning algorithms in Ha'il region, KSA," *Sci. Rep.*, vol. 14, no. 1, p. 502, Jan. 2024, doi: 10.1038/s41598-023-50839-1.

[20] J. H. Cabot and E. G. Ross, "Evaluating prediction model performance," *Surgery*, vol. 174, no. 3, pp. 723–726, Sep. 2023, doi: 10.1016/j.surg.2023.05.023.

© 2026 The Author(s)
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License
ISSN: 3110-1577