Personality and Individual Differences 246 (2025) 113346

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Predictors of loneliness at school: An explainable artificial intelligence approach on a large-scale cross-cultural assessment☆

Simone Zasso [a,*], Lavinia De Marco [b], Stefania Sette [b], Massimo Stella [a,c], Enrico Perinelli [a]

[a] Department of Psychology and Cognitive Science, University of Trento, Italy
[b] Department of Developmental and Social Psychology, Sapienza University of Rome, Italy
[c] CogNosco Lab, Department of Psychology and Cognitive Science, University of Trento, Italy

ARTICLE INFO

Keywords:
Loneliness
Psychosocial skills
Machine learning
Adolescence
Explainable artificial intelligence
Personality traits
School

ABSTRACT

There is extensive evidence linking loneliness to adverse mental health outcomes in adolescents. Yet, research on school-based loneliness predictors is hindered by small sample sizes and a restricted number of variables tested in models. To fill this research gap, this study investigates predictive factors for loneliness at school using data from 60,498 students (30,909 aged 10 and 29,589 aged 15), derived from the SSES survey. Adopting a data-driven approach, we employ 3 machine learning algorithms (Elastic Net, Random Forests, and XGBoost) alongside eXplainable Artificial Intelligence (XAI) techniques to analyze the relationship between a broad range of psychosocial variables and loneliness. Results showed that bullying is the strongest negative predictor of loneliness for 10-year-olds, its influence persisting but getting dampened in 15-year-olds. In contrast, two personality traits, namely extraversion and emotional stability, emerge as key protective factors of loneliness for both 10 and 15-year-olds. Additional important predictors of loneliness include low-quality relationships with parents or friends and high screen use. By integrating classical and interpretable machine learning techniques, this study provides a nuanced understanding of the relative importance of a large number of predictors of loneliness at school, also taking into account nonlinear relationships within the data.

1. Introduction

Loneliness is defined as a negative and unpleasant feeling engendered by the discrepancy between one's actual social relationships and those desired or by the dissatisfaction with the personal social connections in terms of their quality and/or quantity (Asher & Paquette, 2003; Cacioppo et al., 2006). Since 2012, as highlighted by Twenge et al. (2021), there has been an increase in loneliness feelings at school among adolescents worldwide. Adolescents are particularly at risk: As demonstrated by Graham et al. (2024), loneliness follows a U-shaped distribution across the lifespan, i.e., loneliness is high during adolescence, decreases in adulthood, and increases again during old age. During adolescence, a reconfiguration and shift in relational focus occurs, moving from parents to peers. According to Cacioppo's evolutionary theory of loneliness (Goossens, 2018), this shift can lead to an increase in feelings of loneliness in this age group. However, this change does not imply that being immersed in an environment with peers is sufficient to prevent loneliness feelings. It is possible to experience loneliness even when surrounded by peers if the quality of relationships and personal expectations do not align with the expectations (Hawkley et al., 2003; Richardson et al., 2017), as it often happens in schools, where adolescents spend a significant fraction of their time with peers.

1.1. Predictors of loneliness

The sense of loneliness is an important factor to study, particularly given its relation with various negative mental health outcomes, such as depression, social anxiety, suicidality, and sleep disturbances (Goossens,

☆ This contribution was funded by European Union–Next Generation EU - PRIN 2022 PNRR (DD 1409, 14/09/2022) – PNRR – M4 – C2 – INV1.1 – PRIN – Project title "*Loneliness, non-cognitive skills, and academic achievement in Junior High School: Assessment and intervention based on Big Data, longitudinal, and intensive methods*" – Project Identification Code P2022ZPBYX – CUP University of Trento E53D23019470001, CUP Sapienza University of Rome B53D23030050001; Principal Investigator: Enrico Perinelli, Co-Principal Investigator: Stefania Sette.

* Corresponding author at: Department of Psychology and Cognitive Science, University of Trento, Corso Bettini 31, 38068 Rovereto, (TN), Italy.
E-mail address: simone.zasso@unitn.it (S. Zasso).

https://doi.org/10.1016/j.paid.2025.113346
Received 20 December 2024; Accepted 18 June 2025
Available online 27 June 2025
0191-8869/

2018; Park et al., 2020; Richardson et al., 2017). Moreover, research has shown that loneliness experienced during adolescence can have lasting effects into adulthood, leading to negative outcomes such as a higher risk of disability and lower income in midlife (von Soest et al., 2020). Given the negative consequences of loneliness for individuals, particularly adolescents, it becomes critically important to study the factors that may contribute to increasing this sense of loneliness to reduce the detrimental effects over time. However, while loneliness is among the most studied constructs in social and behavioural science (Goossens, 2018; Twenge et al., 2021), loneliness at school has received considerably less attention (Jefferson et al., 2023). One of the few studies examining the predictive factors of school loneliness is that of Olenik-Shemesh and Zeidner (2013), which investigated how personality traits might influence loneliness levels among 203 students. The findings revealed: (a) a positive and significant association between loneliness and neuroticism (e.g., low emotional stability), (b) a negative and significant association between loneliness and extraversion and agreeableness, and (c) non-significant associations between loneliness and openness and conscientiousness. These results partially align with a recent meta-analysis by Buecker et al. (2020), which explored the relationship between the Big Five personality traits and loneliness. Results from this meta-analysis indicated that taking into account measurement errors, the strongest associations were those with extraversion ($\rho = -0.403$ [−0.447, −0.357]) and neuroticism ($\rho = 0.394$ [0.352, 0.434]), followed by agreeableness ($\rho = -0.268$ [−0.304, −0.231]) and conscientiousness ($\rho = -0.219$ [−0.252, −0.186]). In contrast, openness showed the weakest correlation with loneliness ($\rho = -0.118$ [−0.147, −0.089]), which was no longer statistically significant ($\beta < 0.001$ [−0.038, 0.039]) once controlling for other personality traits (Buecker et al., 2020).

Beyond personality traits, the literature showed that another relevant predictor of loneliness is bullying, defined as a form of aggressive behavior characterized by repetition (i.e., the victim is targeted multiple times) and an imbalance of power (i.e., the victim is smaller or physically weaker than the bully; Olweus, 1999). Experiencing bullying victimization at age 10 has been identified as a positive and significant predictor of adolescent loneliness (Hosozawa et al., 2022). However, the association between loneliness and exposure to bullying varies by gender and age group (Cava et al., 2020; Putra & Dendup, 2022). Other factors that have proven to be predictors of loneliness, and which will be considered in this study, include the quality of relationships between students, teachers, and parents, as these relationships can significantly impact feelings of loneliness, especially during adolescence (Sette et al., 2023). Overall, during this period, peers and teachers often optimally meet the needs for inclusion, security, and emotional support rather than within the family context (Ahmed et al., 2010; Vargas-Madriz et al., 2024). Additionally, screen time (e.g., social network, videogame, smartphone) has also been found to be longitudinally positively related to loneliness (Lawrence et al., 2022).

Despite the above evidence, to our knowledge, only few studies have specifically focused on school-related loneliness, exploring predictors based on gender and different age groups. Moreover, predictors of loneliness were usually investigated with relatively small datasets (e.g., the median of the sample sizes reported in Buecker et al.'s [2020] meta-analysis was of 178 participants) and a small number of predictors (e.g., Olenik-Shemesh & Zeidner, 2013), and using linear models (Hosozawa et al., 2022), leaving room to study predictors of loneliness comprehensively.

1.2. The present study

To address the gaps identified in the literature, this study aims to expand the research on school-related loneliness during adolescence by identifying the main predictive factors, leveraging large-scale assessments and recent data-driven techniques. The dataset used in this study is the OECD's *Survey on Social and Emotional Skills* (SSES), encompassing over 60,000 participants from 10 countries evenly distributed between two age groups: Approximately 50 % are 10 years old, and 50 % are 15 years old. Together, these characteristics make the SSES a robust foundation for examining the nuances of loneliness at school during adolescence. To this end, we adopted a data-driven approach to analyze a large number of psychosocial predictors, using a combination of three supervised machine learning algorithms (Elastic Net, Random Forests, and XGBoost; Zollanvari, 2023) paired with Explainable Artificial Intelligence (XAI) methods (SHAP scores; Lundberg et al., 2020; Molnar, 2022). This combination of classical and interpretable machine learning techniques enables a deeper understanding of the importance of predictors while also accommodating nonlinear relationships within the data.

Machine learning approaches offer a powerful data-driven framework for conducting exploratory analyses that can contribute to theoretical advancements (Yarkoni & Westfall, 2017). This methodology is gaining traction in developmental studies due to its capacity to rigorously explore complex, nonlinear patterns and produce robust findings from previously unexamined data (Brieant et al., 2024; Van Lissa, 2023). In this study, we analyzed a wide range of individual and social factors to understand which may predict loneliness and which may act as protective factors within the school context. Additionally, the data were stratified by gender and age group to explore potential variations between these groups (Barreto et al., 2021; Cacioppo et al., 2006; Coplan et al., 2007).

2. Method

2.1. Participants

The original dataset for the OECD's *Survey on Social and Emotional Skills* (SSES) included $N = 61{,}035$ students. However, 25 students were excluded because they did not belong to either age cohort, resulting in a revised sample size of $N = 61{,}010$ students. This dataset was split by cohorts (10-year-old and 15-year-old) and by gender (females and males; see Table 1). The SSES is a survey involving 10-year-old and 15-year-old students from ten cities: Ottawa, Houston, Bogota, Manizales, Helsinki, Moscow, Istanbul, Daegu, Sintra, and Suzhou. The data were collected from 2018 to 2020.

2.2. Measures

In what follows, we describe the target variable and its respective measurement tool included in our models. A summary and composition of the composite scores are provided in Table A1 (Appendix). The other variables are described in detail in the Appendix.

2.2.1. Loneliness at school

The target (i.e., the dependent variable) of the models was *Loneliness at school.* It was measured using three items from the *Sense of Belonging Scale*. The content of these items is similar to that used in another widely used large-scale assessment (i.e., PISA 2022) and in several recently published studies (Jefferson et al., 2023; Twenge et al., 2021). Participants responded using a 4-point scale ranging from 1 (*Strongly Agree*) to 4 (*Strongly Disagree*). The composite score was calculated as the mean of

Table 1
Train and test partition.

Sample	Total	Train (75 %)	Test (25 %)
10 y/o Females	15,367	11,525	3,842
10 y/o Males	15,542	11,656	3,886
15 y/o Females	15,268	11,451	3,817
15 y/o Males	14,321	10,740	3,581
Total	60,498	45,372	15,126

Note: The *N* refers to the data after the imputation.

the items. Cronbach's alpha was 0.77.

2.3. Data analytic plan

2.3.1. Missing value imputation and dataset stratification

Before performing the machine learning analyses, the data from the entire sample ($N = 61{,}010$) were pre-processed. Composite scores for the variables were calculated by averaging their respective items, and cases with missing data were imputed using the Multiple Imputation by Chained Equations (MICE) method (Azur et al., 2011) in Python. After imputing missing data, the final sample size was $N = 60{,}498$. Next, we split the initial dataset by age and gender, thus creating 4 distinct datasets (see Table 1). Descriptive statistics and correlation matrices were calculated separately for each dataset.

2.4. Machine learning analysis

2.4.1. Machine learning algorithms

Our machine learning analyses were based on 3 supervised algorithms: Elastic Net, Random Forests, and XGBoost. We choose these 3 algorithms for several reasons: (a) the use of a linear algorithm (Elastic Net) would allow detecting whether results may change when accounting for non-linear relations (as it is in Random Forests and XGBoost); (b) the choice of Random Forests is due to its extensive use and refinement over the last 20 years; (c) XGBoost is chosen as it is a relatively recent algorithm which showed, in several scenarios, to outperform other algorithms in prediction performance on hold-out data, and thus it is nowadays recognized to be one of the most powerful algorithms (Zollanvari, 2023, p. 230).

2.5. Data analytic plan

For all four datasets, we carried out the same steps. First, each dataset was split into a training (75 %) and a test (25 %) set. Second, the features in the training set were standardized ($M = 0$, $SD = 1$), and the same mean and standard deviation values were then used to standardize the features in the test data (Géron, 2022). The third step regarded hyperparameter tuning (i.e., optimization) for each algorithm. For the Elastic Net model, the hyperparameters (i.e., alpha and l1_ratio) were tuned using 10-fold cross-validation. This cross-validation was performed with a grid search, specifying 24 values for alpha and 8 values for l1_ratio, resulting in a total of 192 possible combinations. For the Random Forest model, 6 hyperparameters (n_estimators, max_features, max_depth, min_samples_split, min_samples_leaf, and bootstrap) were optimized using Bayesian optimization paired with 5-fold cross-validation. Similarly, for the XGBoost model, 10 hyperparameters (learning_rate, n_estimators, max_depth, min_child_weight, subsample, colsample_bytree, colsample_bylevel, gamma, reg_lambda, and reg_alpha) were optimized using Bayesian optimization with 5-fold cross-validation. Fourth, the model was trained on the whole training partition using the best combinations of hyperparameters identified in the previous step. Fifth, the performance of the model was evaluated on the test data. Two performance metrics were used: R^2 and Mean Absolute Error (MAE). Sixth, we checked potential overfitting issues using a bootstrap approach: We calculated performance metrics using 1000 resampling iterations and compared the median and 95 % confidence intervals between the training and test sets.[1] Seventh, feature importance was assessed using model-specific methods. For Elastic Net, importance was derived from the absolute value of the model coefficients, indicating the strength and direction of the linear relationships. For Random Forests, importance was based on the impurity reduction method. In this approach, a feature is more important if it helps to improve the accuracy of the prediction when the trees in the random forest try to split the data. When a feature is able to create splits closer to the pure case, e.g., segregating completely data points with distinct values of the predicted variable, that feature gets higher importance (Scornet, 2023). For XGBoost, importance was quantified using the gain metric, which measures the improvement in the model's loss function from splits on each feature. These methods evaluate importance based on feature contributions during model training.

[1] We point out that properly optimized models are expected to perform slightly worse on the test data than on the training data, indicating good generalization without overfitting.

2.5.1. Explainable artificial intelligence

We used Shapley Additive Explanations (SHAP), a model-agnostic approach (i.e., applicable to any machine learning model regardless of its structure), to interpret the machine learning models. Unlike feature importance analysis, SHAP scores provide detailed insights into the direction and magnitude of each variable's contribution to predictions, capturing both linear and nonlinear effects (Lundberg et al., 2020; Molnar, 2022). This methodology allows for mapping how low or high values of a variable influence model predictions relative to a baseline, offering a nuanced and richer picture of the directionality of how features influence machine learning results. Such an approach is particularly valuable in psychological contexts where understanding complex dynamics and interactions is crucial (Baker et al., 2023; Hassija et al., 2024; Pedregosa et al., 2011).

2.5.2. Software and packages

All analyses were conducted in Python (Version 3.12.2). Machine learning analyses were performed with the following Python packages: scikit-learn (Version 1.4.2; Pedregosa et al., 2011), shap (Version 0.45.1; Lundberg et al., 2020), and scikit-optimize (skopt, Version 0.10.1). For imputation data, we used pyampute (Version 5.7.0). Code are available at https://github.com/SimoneZasso/Loneliness_XAI

3. Results

3.1. Zero-order correlations

Descriptive statistics and zero-order correlations for all study variables, divided by cohort and gender dataset, are provided in Table 2 and Fig. 1, respectively. In the Appendix, descriptive statistics for the whole dataset are presented in Table A2, and zero-order correlations are shown in Fig. A1. Regarding zero-order correlations in Fig. 1, the target variable (Loneliness) exhibited medium correlations ($|r| \geq 0.30$; Cohen, 1992) across the entire sample. Specifically, for 10 y/o females, the highest correlations were with Emotion Regulation ($r = -0.42, p < .001$) and Bullying ($r = 0.41, p < .001$). For 10 y/o, the strongest correlations were found with Bullying ($r = 0.45$, $p < .001$) and Emotion Regulation ($r = -0.37, p < .001$). For 15 y/o females, the highest correlations were with Emotion Regulation ($r = -0.44$, $p < .001$) and Engaging with Others ($r = -0.43, p < .001$). For 15 y/o males, the highest correlations were with Emotion Regulation ($r = -0.42, p < .001$) and Engaging with Others ($r = -0.40, p < .001$).

3.2. Model training and performance evaluation

The results of model training and performance for the three algorithms across the four datasets are presented in Table 3. Regarding model optimization, the findings indicated no evidence of overfitting; of interest, XGBoost appeared to overfit less than Random Forest, further demonstrating its robustness. In terms of performance, all models explained between 35 % and 39 % of the variance in loneliness in the test data across age and gender groups. This narrow range suggests that neither age, gender, or algorithm choice significantly influenced predictive performance. However, XGBoost consistently demonstrated the highest R^2 and the lowest MAE, underscoring its ability to capture nonlinear relationships effectively and its overall superiority in predictive

Table 2
Descriptive statistics by gender and cohort.

Sample	Variable	N	Mean	SD	Median	Min	Max	Skewness	Kurtosis
10 y/o Females	Loneliness	15,367	1.72	0.68	1.67	1	4	0.92	0.62
	Parent_rel	15,367	3.45	0.63	3.67	1	4	−1.47	2.29
	Friend_rel	15,367	3.23	0.74	3.25	1	4	−0.85	0.08
	Teacher_rel	15,367	3.31	0.79	3.67	1	4	−1.04	0.17
	Bullying	15,367	1.41	0.60	1.25	1	4	1.96	3.90
	Task_Perf	15,367	598.66	87.97	586.65	215.86	885.33	0.66	0.66
	Emotion_reg	15,367	553.09	82.60	548.10	93.02	937.78	0.49	2.13
	Collaboration	15,367	605.76	82.73	593.76	295.61	912.78	0.70	0.56
	Open-mind	15,367	613.76	82.96	601.02	291.70	911.84	0.71	0.50
	Engaging_oth	15,367	561.26	73.62	551.82	238.09	890.45	0.66	1.21
	Screen_Time	15,367	1.98	0.67	2	1	4	0.76	0.37
	Class_friendly	15,367	3.77	1.21	4	1	5	−0.78	−0.39
	Class_part	15,367	3.74	1.2	4	1	5	−0.57	−0.67
10 y/o Males	Loneliness	15,542	1.75	0.69	1.67	1	4	0.94	0.70
	Parent_rel	15,542	3.44	0.65	3.67	1	4	−1.48	2.20
	Friend_rel	15,542	3.14	0.75	3.25	1	4	−0.69	−0.13
	Teacher_rel	15,542	3.23	0.80	3.33	1	4	−0.92	−0.06
	Bullying	15,542	1.58	0.71	1.25	1	4	1.46	1.67
	Task_Perf	15,542	588.71	84.49	577.35	238.35	886.02	0.72	0.84
	Emotion_reg	15,542	561.17	78.73	554.33	200.32	942.74	0.74	2.30
	Collaboration	15,542	594.07	81.78	581.43	254.75	904.56	0.83	1.01
	Open-mind	15,542	606.17	80.82	594.38	324.73	910.13	0.74	0.63
	Engaging_oth	15,542	567.11	73.78	557.53	204.70	888.33	0.67	1.23
	Screen_Time	15,542	2.13	0.70	2	1	4	0.54	−0.10
	Class_friendly	15,542	3.77	1.21	4	1	5	−0.79	−0.37
	Class_part	15,542	3.66	1.21	4	1	5	−0.51	−0.72
15 y/o Females	Loneliness	15,268	1.86	0.65	2	1	4	0.58	0.27
	Parent_rel	15,268	3.14	0.7	3.33	1	4	−0.89	0.49
	Friend_rel	15,268	3.17	0.67	3.25	1	4	−0.63	−0.07
	Teacher_rel	15,268	3.27	0.74	3.33	1	4	−0.95	0.13
	Bullying	15,268	1.23	0.39	1	1	4	2.77	10.16
	Task_Perf	15,268	569.06	74.49	561.73	242.36	884.64	0.76	1.44
	Emotion_reg	15,268	507.54	74.28	508.18	99.97	919.91	−0.02	1.99
	Collaboration	15,268	576.28	66.79	570.25	273.22	889.78	0.66	1.35
	Open-mind	15,268	596.06	74.63	585.62	283.76	906.72	0.82	1.06
	Engaging_oth	15,268	530.73	68.85	525.92	200.39	884.09	0.46	1.64
	Screen_Time	15,268	2.32	0.63	2.2	1	4	0.24	−0.3
	Class_friendly	15,268	3.81	1.11	4	1	5	−0.84	−0.07
	Class_part	15,268	3.16	1.13	3	1	5	−0.03	−0.72
15 y/o Males	Loneliness	14,321	1.77	0.62	1.67	1	4	0.71	0.54
	Parent_rel	14,321	3.3	0.64	3.33	1	4	−1.07	1.14
	Friend_rel	14,321	3.11	0.66	3	1	4	−0.55	0.07
	Teacher_rel	14,321	3.24	0.75	3.33	1	4	−0.93	0.21
	Bullying	14,321	1.35	0.5	1.25	1	4	2.2	6.09
	Task_Perf	14,321	569.21	72.71	561.8	240.48	884.12	0.85	1.84
	Emotion_reg	14,321	544.06	71.95	541.08	132.39	944.72	0.53	3.2
	Collaboration	14,321	571.87	67.54	565.35	219.5	888.14	0.81	2.14
	Open-mind	14,321	589.04	73.59	578.11	270.61	910.85	0.91	1.38
	Engaging_oth	14,321	548.18	70.96	542.07	150.19	884.09	0.55	2.04
	Screen_Time	14,321	2.46	0.66	2.4	1	4	0.07	−0.25
	Class_friendly	14,321	3.88	1.04	4	1	5	−0.92	0.26
	Class_part	14,321	3.17	1.12	3	1	5	−0.1	−0.7

Note. Variables are not standardized. Parent_rel = Relationship with parents; Friend_rel = Relationship with friends; Teacher_rel = Relationship with teachers; Bullyng = Bullying experience; Task_Perf = Task Performance; Emotion_reg = Emotion regulation; Open-mind = Open-mindedness; Engaging_oth = Engaging with others; Class_friendly = Classmates - friendly to you; Class_part = Class participation.

accuracy.

3.3. Feature importance and eXplainable Artificial Intelligence (XAI)

The results of feature importance are shown in Fig. 2. For 10-year-old females and males, all algorithms ranked "bullying experience" as the most important predictor of school loneliness. The Elastic Net algorithm also highlights "engaging with others" as the second most important factor, while the Random Forest and XGBoost models emphasize "emotional regulation" as a key predictor. For 15-year-olds, the Random Forest and XGBoost algorithms prioritize "emotional regulation", followed by "engaging with others". The Elastic Net algorithm, however, ranks "engaging with others" as the top predictor for both groups, with "emotional regulation" second for females and bullying for males. XGBoost also highlights factors like "class friendly" and "relationships with friends", and for males, bullying remains significant.

These findings are further clarified by SHAP beeswarm plots, which integrate explainable artificial intelligence (XAI) to show how each feature influences model predictions. Fig. 3 presents the SHAP plots for the Random Forest and XGBoost algorithms, by gender and age. Features are ordered by importance, with wider areas indicating greater influence on predictions (Baker et al., 2023). The current one is a regression task, so here positive SHAP values correspond to the machine learning model producing higher predictions when tuning one feature and correcting for all others. Negative SHAP values are linked to lower predictions, instead. Red colours reflect higher feature values, while blue colours represent lower values.

Notably, high levels of bullying experience are the strongest predictor of school loneliness for 10-year-old females and males (Fig. 3). Low scores in "engaging with others" and "emotional regulation" also

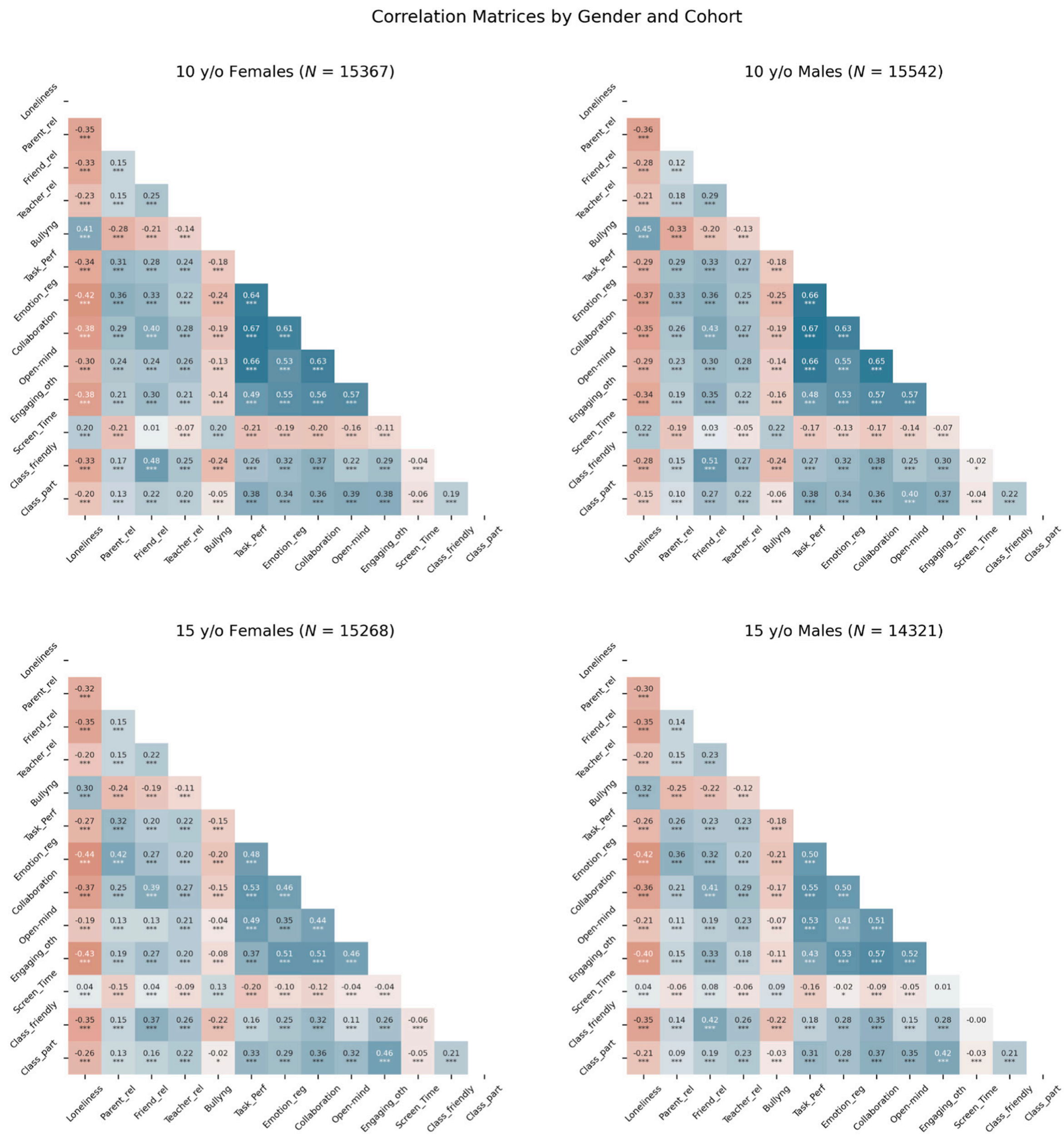

Fig. 1. Correlation matrices by gender and cohort.
Note. Variables are not standardized. Parent_rel = Relationship with parents; Friend_rel = Relationship with friends; Teacher_rel = Relationship with teachers; Bullyng = Bullying experience; Task_Perf = Task Performance; Emotion_reg = Emotion regulation; Open-mind = Open-mindedness; Engaging_oth = Engaging with others; Class_friendly = Classmates - friendly to you; Class_part = Class participation.

predict higher loneliness, while high scores in these traits reduce it. Poor relationships with parents and friends further predict loneliness. High levels of screen time are another important predictor of loneliness, but their influence follows a non-linear pattern. Specifically, the SHAP values reveal a long red tail on the right, indicating that high levels of screen time can have either a low or high positive effect on loneliness. In contrast, the compressed blue region around zero suggests that low levels of screen time have a minimal predictive impact on loneliness. For 15-year-olds, low scores in "engaging with others" and "emotional regulation" are the strongest predictors of loneliness, while high scores in these traits serve as protective factors. Although bullying's predictive power diminishes with age, it remains when its values are high (as demonstrated by the long red tail in the right part of the SHAP beeswarm plot). Finally, low levels of "relationship with parents" and "relationship with friends" are associated with increased loneliness, whereas positive relationships decrease loneliness. Interestingly, screen time continues to predict loneliness in males at age 15, though its effect is weaker than at age 10.

4. Discussion

Few studies have examined predictors of school loneliness, often relying on small samples and a restricted number of variables. Therefore, our aim was to address these gaps by using a large sample of students stratified by gender and age and a broad set of both individual and social variables. To this end, we used a data-driven approach with three machine learning algorithms and an XAI framework to identify key predictors of school loneliness, revealing age-specific differences.

For 10-year-old students, there were no substantial differences in terms of gender. Indeed, the experience of bullying emerged as the most

Table 3
Performance metrics.

Sample	Algorithm	Partition	R^2	MAE	Bootstrap (1000 resampling)	
					Median R^2 [95% CI]	Median MAE [95 % CI]
10 y/o Females	Elastic Net	Train	0.361	0.417	0.360 [0.332, 0.389]	0.418 [0.407, 0.428]
		Test	0.374	0.421	0.374 [0.330, 0.414]	0.420 [0.402, 0.439]
	Random Forests	Train	0.522	0.365	0.506 [0.482, 0.529]	0.371 [0.361, 0.380]
		Test	0.377	0.419	0.377 [0.334, 0.417]	0.419 [0.400, 0.437]
	XGBoost	Train	0.422	0.397	0.404 [0.372, 0.435]	0.406 [0.395, 0.419]
		Test	0.391	0.412	0.384 [0.336, 0.425]	0.419 [0.399, 0.438]
10 y/o Males	Elastic Net	Train	0.348	0.427	0.347 [0.318, 0.374]	0.428 [0.417, 0.439]
		Test	0.348	0.431	0.348 [0.301, 0.392]	0.431 [0.412, 0.449]
	Random Forests	Train	0.506	0.380	0.492 [0.469, 0.514]	0.384 [0.375, 0.394]
		Test	0.361	0.428	0.360 [0.314, 0.403]	0.428 [0.410, 0.447]
	XGBoost	Train	0.474	0.387	0.437 [0.410, 0.466]	0.401 [0.388, 0.414]
		Test	0.365	0.425	0.358 [0.309, 0.405]	0.430 [0.411, 0.448]
15 y/o Females	Elastic Net	Train	0.362	0.402	0.362 [0.336, 0.384]	0.402 [0.392, 0.412]
		Test	0.371	0.410	0.369 [0.325, 0.411]	0.411 [0.392, 0.428]
	Random Forests	Train	0.511	0.357	0.494 [0.472, 0.512]	0.363 [0.354, 0.372]
		Test	0.368	0.410	0.368 [0.327, 0.406]	0.411 [0.392, 0.429]
	XGBoost	Train	0.411	0.387	0.379 [0.331, 0.411]	0.400 [0.386, 0.418]
		Test	0.379	0.405	0.360 [0.304, 0.406]	0.415 [0.394, 0.438]
15 y/o Males	Elastic Net	Train	0.338	0.394	0.339 [0.308, 0.365]	0.393 [0.384, 0.404]
		Test	0.350	0.391	0.350 [0.297, 0.394]	0.390 [0.374, 0.408]
	Random Forests	Train	0.545	0.334	0.523 [0.500, 0.543]	0.341 [0.332, 0.350]
		Test	0.359	0.390	0.359 [0.313, 0.401]	0.390 [0.374, 0.406]
	XGBoost	Train	0.434	0.367	0.399 [0.356, 0.432]	0.379 [0.367, 0.394]
		Test	0.363	0.387	0.349 [0.292, 0.396]	0.392 [0.375, 0.411]

Note. MAE = Mean Absolute Error. 95%CI = 95 % confidence intervals.

important predictor of loneliness for both males and females, consistent with previous literature (Hosozawa et al., 2022). Then, high levels of two personality traits, namely "engaging with others" (extraversion) and "emotional regulation" (emotional stability), emerged to be the main protective factors since they predict lower levels of loneliness. These findings confirmed the important role played by extraversion and neuroticism (Buecker et al., 2020). Thus, our contribution showed that, albeit previous literature showed the importance of bullying and personality in determining loneliness levels, experiences of bullying (i.e., a contextual variable) may be a more important predictor than personality (i.e., individual dispositions), in particular during early adolescence.

For 15-year-old students as well, the results did not show substantial gender differences. Again extraversion and emotional stability emerged as the main predictors; when these traits have high scores, they act as protective factors in reducing loneliness. Conversely, low scores make them predictors of high levels of loneliness. For this group, bullying remains a strong predictor of school loneliness, though it had less impact compared to 10-year-old students. The data on personality traits were again consistent with the meta-analysis by Buecker et al. (2020). This study showed that, compared to 10-year-old students, personal dispositions are more important for predicting loneliness in 15-year-olds than contextual factors. This difference can be partially explained by the maturation of these traits during development (Slobodskaya & Kornienko, 2021).

In addition to bullying and personality, the results supported that the quality of relationships with parents and friends were also important predictors of loneliness (Jefferson et al., 2023). In more detail, for 10-year-olds, the bond with parents was more relevant, while for 15-year-olds, the quality of relationships with friends was more important. These findings align with Cacioppo's evolutionary theory of loneliness, which posits a shift in relational focus from parents to peers during adolescence (Goossens, 2018).

Finally, high screen time was associated with high levels of loneliness among 10-year-old and 15-year-old boys but not 15-year-old girls (Lawrence et al., 2022). Notably, the SHAP plot (Fig. 3) revealed a non-linear relationship between screen time and loneliness. This finding underscores the importance of employing techniques that account for non-linear patterns in data when investigating predictors of loneliness (and other psychological constructs, as well).

Our findings have thus expanded the existing literature by showing that various factors can influence loneliness levels and that these factors may change with the transition to adolescence.

4.1. Limitations and future research

Despite several strengths of this study, such as the use of machine learning methodologies and the Explainable Artificial Intelligence (XAI) approach, the large-scale assessment involving 10 countries,[2] stratification by gender and cohort, and the inclusion of multiple individual and social variables, our study has certain limitations. One of the main limitations is the lack of longitudinal data. This restricts us from inferring correlations between variables without being able to establish the directionality of effects. However, the findings of our study align partially with previous meta-analyses and longitudinal studies, which support the reliability of our results. To deepen the insights of this study, future research could explore the connections between loneliness and the variables considered here using longitudinal designs. A second limitation lies in our focus on only two specific age groups, namely, students aged 10 and 15 years old. Consequently, future studies should examine the robustness of our results across a broader range of age groups. Finally, our study relies solely on self-reported variables.

5. Conclusion

In conclusion, our study identified key predictive factors of school-related loneliness in > 60,000 individuals. To this end, we adopted a data-driven approach to analyze the psychosocial predictors of loneliness at school, using a combination of three supervised machine learning algorithms, paired with Explainable Artificial Intelligence (XAI) methods. This synergy enabled a deeper understanding of the importance of predictors while also accommodating nonlinear relationships within the data. Results indicate that bullying experience is the strongest

[2] We point out that in a subsequent supplementary analysis, we took into account the effect of countries (using 9 dummy-coded variables) and found that none of the country variables showed notable feature importance in predicting loneliness.

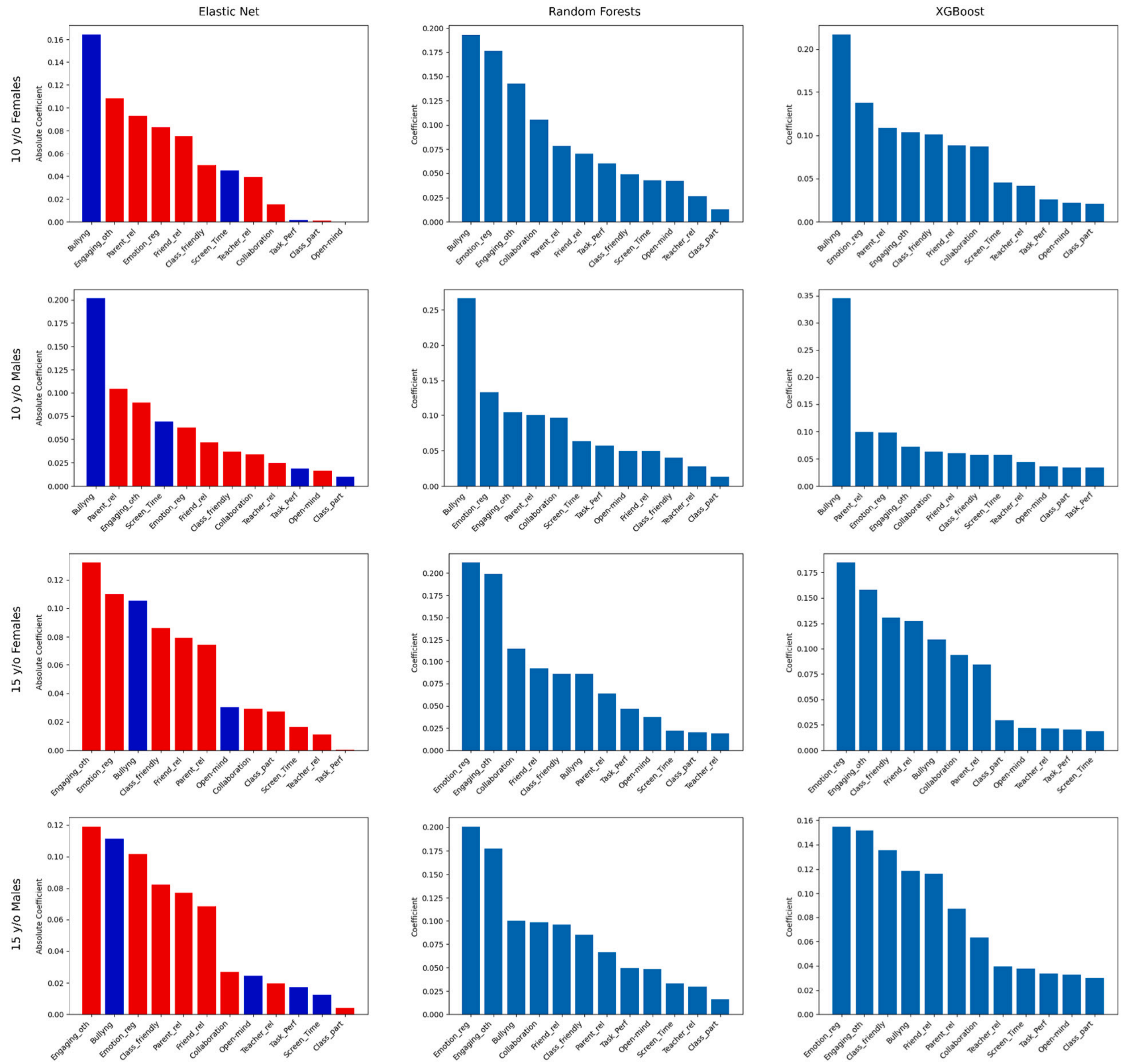

Fig. 2. Feature importance.
Note. In Elastic Net bar charts, blue and red represent positive and negative effects, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predictor of loneliness among 10-year-old students, remaining important for 15-year-olds as well, albeit to a lesser extent. Indeed, for 15-year-olds, personality traits (extraversion and emotional stability) emerged as the primary predictors: Low scores in these traits increase the risk of loneliness, whereas high scores play a protective role. This same pattern was observable, though with less intensity, among 10-year-old children. Additionally, we observed a predictive effect from a low-quality relationship with parents or friends. Finally, regarding findings related to screen time, our results show the importance of considering non-linear relationships in data. We hope this study offers valuable theoretical and practical insights to address loneliness within the school context. Future intervention programs should be tailored to address the diverse ways in which feelings of loneliness manifest over the course of adolescence, including activities involving class groups and/or training on individual skills.

CRediT authorship contribution statement

Simone Zasso: Conceptualization, Methodology, Writing – original draft, Formal analysis. **Lavinia De Marco:** Writing – original draft, Writing – review & editing. **Stefania Sette:** Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing, Supervision. **Massimo Stella:** Methodology, Writing – review & editing. **Enrico Perinelli:** Conceptualization, Funding acquisition, Methodology, Project administration, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare no conflict of interest.

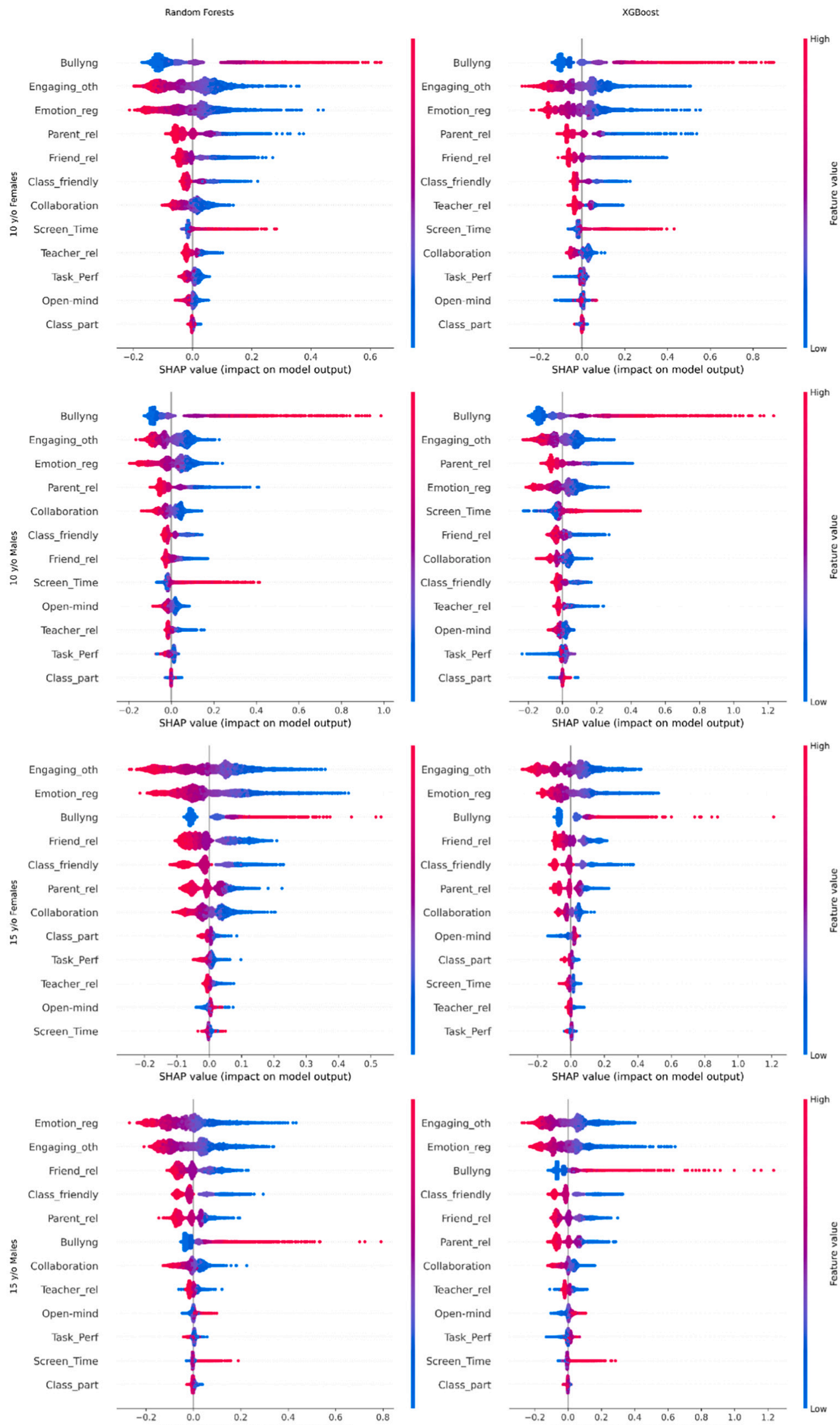

Fig. 3. Beeswarm plot for SHAP value.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.paid.2025.113346.

Data availability

Raw data are publicly accessible via the following link: https://www.oecd.org/en/data/datasets/SSES-Round-1-Database.html. The code necessary to replicate the preliminary analyses and machine learning procedures is available on GitHub: https://github.com/SimoneZasso/Loneliness_XAI

References

Ahmed, W., Minnaert, A., van der Werf, G., & Kuyper, H. (2010). Perceived social support and early adolescents' achievement: The mediational roles of motivational beliefs and emotions. *Journal of Youth and Adolescence, 39*, 36–46.

Asher, S. R., & Paquette, J. A. (2003). Loneliness and peer relations in childhood. *Current Directions in Psychological Science, 12*(3), 75–78. https://doi.org/10.1111/1467-8721.01233

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research, 20*(1), 40–49. https://doi.org/10.1002/mpr.329

Baker, O., Montefinese, M., Castro, N., & Stella, M. (2023). Multiplex lexical networks and artificial intelligence unravel cognitive patterns of picture naming in people with anomic aphasia. *Cognitive Systems Research, 79*, 43–54. https://doi.org/10.1016/j.cogsys.2023.01.007

Barreto, M., Victor, C., Hammond, C., Eccles, A., Richins, M. T., & Qualter, P. (2021). Loneliness around the world: Age, gender, and cultural differences in loneliness. *Personality and Individual Differences, 169*, Article 110066. https://doi.org/10.1016/j.paid.2020.110066

Brieant, A., Sisk, L. M., Keding, T. J., Cohodes, E. M., & Gee, D. G. (2024). Leveraging multivariate approaches to advance the science of early-life adversity. *Child Abuse & Neglect, 106754*. https://doi.org/10.1016/j.chiabu.2024.106754

Buecker, S., Maes, M., Denissen, J. J., & Luhmann, M. (2020). Loneliness and the Big Five personality traits: A meta-analysis. *European Journal of Personality, 34*(1), 8–28. https://doi.org/10.1002/per.2229

Cacioppo, J. T., Hawkley, L. C., Ernst, J. M., Burleson, M., Berntson, G. G., Nouriani, B., & Spiegel, D. (2006). Loneliness within a nomological net: An evolutionary perspective. *Journal of Research in Personality, 40*(6), 1054–1085. https://doi.org/10.1016/j.jrp.2005.11.007

Cava, M.-J., Tomas, I., Buelga, S., & Carrascosa, L. (2020). Loneliness, depressive mood and cyberbullying victimization in adolescent victims of cyberg dating violence. *International Journal of Environmental Research and Public Health, 17*(12). https://doi.org/10.3390/ijerph17124269

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Coplan, R. J., Closson, L. M., & Arbeau, K. A. (2007). Gender differences in the behavioral associates of loneliness and social dissatisfaction in kindergarten. *Journal of Child Psychology and Psychiatry, 48*(10), 988–995. https://doi.org/10.1111/j.1469-7610.2007.01804.x

Géron, A. (2022). *Hands-on machine learning with Scikit-learn, Keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly.

Goossens, L. (2018). Loneliness in adolescence: Insights from Cacioppo's evolutionary model. *Child Development Perspectives, 12*(4), 230–234.

Graham, E. K., Beck, E. D., Jackson, K., Yoneda, T., McGhee, C., Pieramici, L., … Ong, A. D. (2024). Do we become more lonely with age? A coordinated data analysis of nine longitudinal studies. *Psychological Science, 35*(6), 579–596. https://doi.org/10.1177/09567976241242037

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., … Hussain, A. (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation, 16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Hawkley, L. C., Burleson, M. H., Berntson, G. G., & Cacioppo, J. T. (2003). Loneliness in everyday life: Cardiovascular activity, psychosocial context, and health behaviors. *Journal of Personality and Social Psychology, 85*(1), 105–120. https://doi.org/10.1037/0022-3514.85.1.105

Hosozawa, M., Cable, N., Yamasaki, S., Ando, S., Endo, K., Usami, S., Nakanishi, M., Niimura, J., Nakajima, N., Baba, K., Oikawa, N., Stanyon, D., Suzuki, K., Miyashita, M., Iso, H., Hiraiwa-Hasegawa, M., Kasai, K., & Nishida, A. (2022). Predictors of chronic loneliness during adolescence: A population-based cohort study. *Child and Adolescent Psychiatry and Mental Health, 16*(1), 107. https://doi.org/10.1186/s13034-022-00545-z

Jefferson, R., Barreto, M., Jones, F., Conway, J., Chohan, A., Madsen, K. R., … Qualter, P. (2023). Adolescent loneliness across the world and its relation to school climate, national culture and academic performance. *British Journal of Educational Psychology, 93*(4), 997–1016. https://doi.org/10.1111/bjep.12616

Lawrence, D., Hunter, S. C., Cunneen, R., Houghton, S. J., Zadow, C., Rosenberg, M., … Shilton, T. (2022). Reciprocal relationships between trajectories of loneliness and screen media use during adolescence. *Journal of Child and Family Studies, 31*(5), 1306–1317. https://doi.org/10.1007/s10826-021-02066-3

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., … Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Molnar, C. (2022). *Interpretable machine learning* (2nd ed.). Available at https://christophm.github.io/interpretable-ml-book/.

Olenik-Shemesh, D., & Zeidner, M. (2013). Personality predictors of school loneliness in adolescent students. *Psychology Research, 3*(10), 579.

Olweus, D. (1999). Sweden. In P. K. Smith, Y. Morita, J. Juner-Tas, D. Olweus, R. Catalano, & P. Slee (Eds.), *The nature of school bullying: A cross-national perspective* (pp. 2–27). Routledge.

Park, C., Majeed, A., Gill, H., Tamura, J., Ho, R. C., Mansur, R. B., … McIntyre, R. S. (2020). The effect of loneliness on distinct health outcomes: A comprehensive review and meta-analysis. *Psychiatry Research, 294*, Article 113514. https://doi.org/10.1016/j.psychres.2020.113514

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research, 12*, 2825–2830.

Putra, G. N. E., & Dendup, T. (2022). Health and behavioural outcomes of bullying victimisation among Indonesian adolescent students: Findings from the 2015 Global School-based Student Health Survey. *Psychology, Health and Medicine, 27*(3), 513–527. https://doi.org/10.1080/13548506.2020.1826546

Richardson, T., Elliott, P., & Roberts, R. (2017). Relationship between loneliness and mental health in students. *Journal of Public Mental Health, 16*(2), 48–54. https://doi.org/10.1108/JPMH-03-2016-0013

Scornet, E. (2023). Trees, forests, and impurity-based variable importance in regression. *Annales de l'Institute Henri Poincare, Probabilites et Statistiques, 59*(1), 21–52. https://doi.org/10.1214/21-AIHP1240

Sette, S., Zava, F., Laghi, F., Baumgartner, E., & Coplan, R. J. (2023). Linking social relationships at school, loneliness, and academic self-perceptions among primary school children. *Journal of Applied Developmental Psychology, 88*, Article 101568. https://doi.org/10.1016/j.appdev.2023.101568

Slobodskaya, H. R., & Kornienko, O. S. (2021). Age and gender differences in personality traits from early childhood through adolescence. *Journal of Personality, 89*(5), 933–950. https://doi.org/10.1111/jopy.12624

Twenge, J. M., Haidt, J., Blake, A. B., McAllister, C., Lemon, H., & Le Roy, A. (2021). Worldwide increases in adolescent loneliness. *Journal of Adolescence, 93*(1), 257–269. https://doi.org/10.1016/j.adolescence.2021.06.006

Van Lissa, C. J. (2023). Developmental data science: How machine learning can advance theory formation in Developmental Psychology. *Infant and Child Development, 32*(6), Article e2370. https://doi.org/10.1002/icd.2370

Vargas-Madriz, L. F., Konishi, C., & Wong, T. K. (2024). A meta-analysis of the association between teacher support and school engagement. *Social Development, 33*, Article e12745. https://doi.org/10.1111/sode.12745

von Soest, T., Luhmann, M., & Gerstorf, D. (2020). The development of loneliness through adolescence and young adulthood: Its nature, correlates, and midlife outcomes. *Developmental Psychology, 56*(10), 1919–1934. https://doi.org/10.1037/dev0001102

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100–1122.

Zollanvari, A. (2023). *Machine learning with Python: Theory and implementation*. Springer.