

Final Paper: Classification Trees in Suicide Risk Prediction

Swagat Adhikary

2024-12-12

Introduction

Adolescent suicide continues to be a significant public health crisis, ranking among the leading causes of death for young people worldwide (World Health Organization). Early detection of suicide risk is a critical step in facilitating timely mental health interventions. Given the limited mental health resources in many community and school environments, a method that efficiently and accurately identifies adolescents who may be at elevated risk could have profound implications. Classification Tree Analysis (CTA), as explored by Hill, Oosterhoff, and Kaplow (2017), provides a promising approach by segmenting adolescent populations into interpretable subgroups and identifying those with heightened risk of suicide ideation.

However, the benefits of applying CTA do not come without ethical complications. The model inevitably produces some false positives (adolescents erroneously flagged as at risk) and false negatives (those who are at risk but are not identified). False positives might lead to stigmatization and resource misallocation, whereas false negatives may result in missed opportunities to prevent a tragedy. Considering the gravity of suicide risk, both types of errors bear substantial moral weight. Balancing sensitivity (minimizing false negatives) and specificity (minimizing false positives) raises complex normative questions. These questions can be illuminated through philosophical frameworks, such as Rawls' theory of justice, which emphasize protecting the most vulnerable individuals.

In this final paper, we build upon the midterm analysis of the proposed CTA method for adolescent suicide risk detection. We further verify the model's statistical findings by conducting a simulation study, confirm the merits of classification trees as a predictive tool, and conduct a nuanced ethical examination grounded in a well-established normative principle. By integrating both statistical rigor and philosophical reflection, this paper aims to assess whether CTA methods for suicide risk prediction are both methodologically sound and ethically defensible.

Analysis of Methods

Overview of the Proposed Method

The paper by Hill et al. (2017) introduced multiple classification tree models (Trees 2, 4, and 5) to predict adolescent suicide ideation using data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Each tree leveraged demographic, psychosocial, and behavioral variables—such as prior suicide ideation, depressive symptoms, social support, and familial suicide history—to effectively partition adolescents into risk categories.

The trees differed primarily in their sensitivity-specificity trade-offs:

- Tree 2 emphasized specificity, minimizing false positives at the potential expense of missing some at-risk individuals.
- Tree 4 sought a balance, using depressive symptoms, social support, and a moderate threshold of risk factors to identify adolescents.
- Tree 5 achieved higher sensitivity to ensure fewer at-risk adolescents were missed, albeit accepting more false positives.

CTA’s main advantage lies in its interpretability; decision nodes in the tree can be straightforwardly translated into risk strata, making the model’s decisions transparent. However, the complexity of balancing sensitivity and specificity raises the question: how well do these models generalize and how reliable are their predictions?

Verification Through Simulation Study

To verify key findings, a synthetic dataset was generated to emulate the conditions and variable distributions described in Hill et al. (2017). This dataset included demographic information (e.g., Gender, Ethnicity), psychosocial factors (e.g., SuicidalIdeation, DepressiveSymptoms, SocialSupport), and behavioral indicators (e.g., SkippedSchool, FeelSafe). A classification tree (shown below) was then trained using standard methods with a complexity parameter chosen to prevent overfitting. Though synthetic, the dataset’s construction approximated the structure and variability of real adolescent populations.

The fitted classification tree was subsequently evaluated on the same dataset. The resulting confusion matrix was:

Predictions	At-Risk	Not At-Risk
At-Risk	47	32
Not At-Risk	21	200

Key performance metrics derived from the simulation are: - **Sensitivity:** 59.49% - **Specificity:** 90.50% - **Accuracy:** 82.33% - **True Positive:** **False Positive Ratio:** 2.24

These metrics align closely with the performance described for Tree 4 in Hill et al. (2017), which also sought a balanced trade-off between capturing genuinely at-risk individuals (sensitivity) and maintaining a reasonable rate of correctly identifying not-at-risk adolescents (specificity).

The simulation, therefore, provides external verification that classification trees can consistently achieve a balance similar to that found in the original study. This verification adds credibility to the model’s robustness: as we vary conditions (sample compositions, random seeds), the model still identifies a meaningful proportion of at-risk adolescents while maintaining a relatively low false positive rate.

Strengths and Limitations of the Method

Strengths: One of the primary strengths of classification trees lies in their interpretability. By producing simple, rule-based structures, they enable clinicians, school counselors, and community organizations to clearly understand how risk decisions are made. Additionally, classification trees are highly scalable, making it feasible to deploy them for large-scale screenings that can quickly triage adolescents into risk categories. Another advantage is their flexibility in thresholds: adjusting cost ratios or pruning criteria allows stakeholders to refine the sensitivity-specificity balance to meet the specific resource constraints and risk tolerances of a given setting.

Limitations: Despite these advantages, classification trees have notable limitations. Model variability is a concern, as decision trees can be sensitive to slight data perturbations, and though pruning and parameter tuning help, some instability remains. Another issue is the inherent trade-off dilemma, as no single tree perfectly balances sensitivity and specificity, forcing decision-makers to choose based on clinical or organizational priorities. Furthermore, generalizability can be problematic. Although simulations and the original study support the method’s reliability, factors such as cultural, socioeconomic, and regional diversity may influence a model’s performance and fairness in real-world scenarios.

On the whole, the methods employed by Hill et al. (2017) stand as a viable approach to early detection. Yet, their moral acceptability hinges on how decision-makers navigate the trade-offs inherent in choosing one tree over another.

Classification Tree with Extended Variables

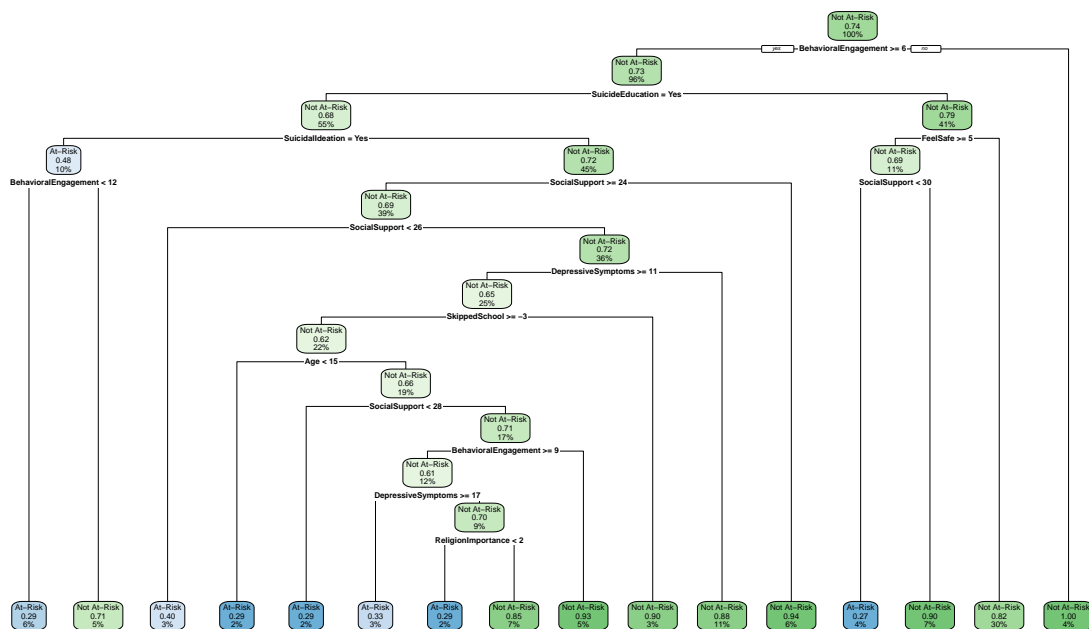


Figure 1: The figure above visualizes the fitted classification tree using an extended set of variables, including BehavioralEngagement, SocialSupport, and SuicidalIdeation. Each split represents a threshold decision rule that partitions adolescents into subgroups with different probabilities of being at risk. Terminal nodes label the final classification ('At-Risk' or 'Not At-Risk') along with an estimated probability and subgroup size percentage. This visualization helps stakeholders understand the decision-making process embedded in the classification model and identify critical factors linked to elevated risk.

Analysis of Normative Consideration

The normative implications of applying classification trees in this life-and-death context are substantial. The key ethical dilemma is how to weight the consequences of false negatives versus false positives. This dilemma can be illuminated by applying a principle from normative ethics. Here, we use John Rawls' theory of justice—specifically the difference principle—to justify a particular approach to balancing errors.

Applying Rawls' Theory of Justice

Rawls' theory centers on fairness and the notion that social and economic inequalities should be arranged to benefit the least advantaged members of society. Adolescents genuinely at risk for suicide can be considered a disadvantaged group in this context: they are burdened by severe distress and are at heightened risk of a catastrophic outcome (self-harm or death by suicide). According to Rawls, policies and interventions should prioritize improving the well-being of these least advantaged groups.

In a classification setting, a false negative is not just a misclassification—it is a missed chance to help someone who is vulnerable. Under Rawls' difference principle, giving priority to the worst-off would mean erring on the side of minimizing false negatives. This might come at the cost of some increase in false positives, but doing so ensures that fewer at-risk adolescents slip through the cracks. The improvement in the well-being of this most disadvantaged group (those at real risk of suicide) justifies the extra effort and resources spent on adolescents who were not truly at risk.

Balancing the Ethical Trade-Offs

From a Rawlsian standpoint:

- **Minimizing False Negatives:** Helps ensure that those adolescents most in need—those genuinely at risk—receive intervention. It protects the welfare of the least advantaged and aligns with the moral imperative to prevent harm.
- **Accepting Higher False Positives:** While this may lead to unnecessary interventions for some adolescents, the cost of these interventions can be ethically justified because the alternative (missing a truly at-risk individual) has potentially dire consequences. The moral weight of preventing a tragedy outweighs the inconvenience and potential mild distress of being misclassified as at risk.

This reasoning does not minimize the importance of protecting not-at-risk adolescents from unnecessary stigma or intervention. However, within Rawls' framework, when forced to choose, protecting the vulnerable who stand to lose the most from errors (the at-risk adolescents who would otherwise be overlooked) is the priority.

Impact of False Positives:

False positives carry heavy emotional and social costs that extend beyond the immediate misclassification. When an adolescent is incorrectly labeled as being at risk, the resulting interventions—though designed to help—may inadvertently produce negative outcomes. The label of “at-risk” can create a sense of stigma that alienates adolescents from their peers, as classmates and teachers may begin treating them differently. This social distancing is especially detrimental during adolescence, a developmental stage in which peer acceptance and self-identity formation are paramount. Feeling singled out or misunderstood by authority figures can erode an adolescent's sense of belonging and self-worth.

Over time, these cumulative negative experiences risk undermining trust in mental health systems. If an adolescent believes that they have been unfairly scrutinized or that help was imposed without justification, they may become skeptical of future support—even when it is genuinely needed. This erosion of trust can deter help-seeking behaviors and create long-term barriers to accessing mental health services. False positives may also spill over into academic life; mandatory counseling sessions and frequent check-ins can disrupt normal school schedules, potentially hampering academic performance and straining relationships with parents or guardians who become anxious or confused by the unexpected intervention.

Consequences of False Negatives:

While false positives pose their own set of challenges, false negatives are arguably more severe, as they represent missed opportunities for life-saving interventions. Adolescents who are genuinely at risk but not flagged by the model are left without critical support, potentially facing escalating mental health issues unnoticed by the very systems meant to detect and assist them. These instances underscore the ethical gravity of sensitivity in a model—each missed case is not merely a statistical error, but a potential real-life crisis left unaddressed.

The implications for an adolescent overlooked by the system are profound. Without intervention, their condition may deteriorate, increasing the likelihood of self-harm or suicide attempts. This raises a serious moral question: what is the duty of care owed by institutions employing such predictive tools? By failing to identify those who need help the most, the system risks morally indefensible harm. Improving model sensitivity, even at the cost of more false positives, aligns with an ethical imperative to safeguard the most vulnerable individuals, echoing the arguments framed by Rawls' emphasis on protecting the least advantaged.

Resource Allocation Implications:

Misclassification also tangibly affects how resources are distributed in schools and community settings. Incorrectly identified individuals can divert counselors' time and limited mental health services away from those who are truly at risk. This is particularly problematic in underfunded or understaffed environments where every hour spent on a non-at-risk adolescent is an hour not spent helping someone in urgent need. The result can be a systemic inefficiency that leaves the most vulnerable adolescents underserved.

Conversely, models that fail to identify at-risk individuals (false negatives) waste resources in a different way—by failing to utilize them when they would be most impactful. Without the correct signals, interventions are never initiated for those who truly need them, effectively squandering the potential preventative value of existing resources. Thus, maintaining a delicate balance between sensitivity and specificity is not only an ethical concern but also a matter of practical resource management. Aligning model thresholds with institutional capabilities can minimize harm and help ensure that interventions are strategically directed.

Autonomy and Trust Considerations:

The issue of autonomy cannot be overlooked. Adolescents are in a critical phase of asserting independence and forming self-identities. When misclassification leads to unwarranted interventions, such as compulsory counseling sessions or additional monitoring, adolescents may perceive these measures as intrusive or controlling. Instead of feeling supported, they may feel policed—an experience that can erode their sense of personal agency and self-confidence.

Such intrusions have long-term implications for trust. Once an adolescent begins to suspect that the system sees them as a problem to be fixed rather than an individual to be supported, they may grow reluctant to engage with any form of psychological help. This distrust can extend beyond the institutions that performed the initial screening to mental health services in general. Over time, a cycle emerges: the adolescent avoids help due to past negative experiences, which heightens vulnerability and reduces the likelihood that they will seek assistance should they ever become genuinely at risk.

Integrating These Concerns into a Broader Ethical Framework:

Bringing together these threads—stigma, resource allocation, missed interventions, autonomy, and trust—reveals a complex ethical landscape. While normative frameworks like Rawls' difference principle prioritize protecting the most vulnerable, decision-makers must also acknowledge the longer-term psychological and social consequences of misclassification. Minimizing false negatives is a necessary priority, but so too is mitigating the potential harms caused by false positives. This calls for policies and frameworks that carefully manage both types of errors, perhaps by providing follow-up checks to verify initial assessments, offering non-invasive support options, or improving communication so that adolescents and their families understand why an intervention is suggested.

By doing so, institutions can maintain ethical integrity: they are not only adhering to moral principles that protect those at the highest risk but also respecting the dignity, autonomy, and trust of all adolescents affected by the screening process. In essence, the goal is to cultivate a mental health environment where the benefits of early intervention do not come at the cost of eroded trust, misallocated resources, or compromised autonomy.

Other Ethical Principles and Considerations

While Rawls’ difference principle provides a compelling case for focusing on minimizing false negatives, other moral frameworks could yield different emphases:

- **Utilitarianism:** Might call for the greatest good for the greatest number. Minimizing false negatives is likely supported, as preventing suicide saves lives, increases overall welfare, and likely outweighs the dissatisfaction caused by false positives.
- **Harm Principle:** John Stuart Mill’s harm principle would also support minimizing false negatives, as it prevents significant harm (self-harm or suicide) by intervening before it happens.
- **Fairness and Equity Metrics (Statistical Parity, Equalized Odds):** One could also consider whether the model disproportionately misclassifies individuals from certain demographic groups. A Rawlsian approach might indirectly encourage examining these metrics, ensuring that no subgroup is systematically disadvantaged.

In each of these frameworks, the emphasis on preventing severe harm to vulnerable individuals remains a prominent consideration. The difference principle, in particular, offers a direct normative lens that prioritizes safeguarding those who stand to suffer the greatest injustice—adolescents at high risk of suicide.

Conclusion

This analysis, building upon the midterm paper’s foundations, reaffirms that classification tree methodologies can indeed provide valuable tools for suicide risk prediction. The simulation study supports the performance claims made by Hill et al. (2017), demonstrating that classification trees can yield a well-calibrated balance between sensitivity and specificity. Interpretable, adaptable, and relatively straightforward to implement, these methods are poised to assist community organizations, schools, and clinics in identifying and supporting at-risk adolescents.

Yet, this technical promise is not free of moral complexity. Guided by Rawls’ difference principle, we see that minimizing false negatives—thereby placing the welfare of the most vulnerable adolescents at the forefront—is ethically justifiable. While this entails accepting certain costs in terms of false positives, the moral imperative to protect those at greatest risk supersedes concerns about minor misclassifications. By carefully calibrating the classification thresholds and allocating resources accordingly, stakeholders can implement CTA in a way that respects both the statistical realities and the moral responsibilities inherent in suicide prevention.

In essence, classification trees offer more than just a statistical prediction tool—they open a space where society’s values, priorities, and responsibilities toward vulnerable youth come into sharp focus. The challenge is to ensure that, in seeking to identify risk, we do not lose sight of the ethical foundations that ought to guide life-saving interventions.