Bridging the Gap: Social Work Insights for Ethical Algorithmic Decision-Making in Human Services

Rodriguez, M., DePanfilis, D., & Lanier, P.

Artificial Intelligence (AI), when combined with statistical techniques such as predictive analytics, has been increasingly applied in high stakes decision making systems seeking to predict and/or classify the risk of clients experiencing negative service outcomes. One such system is child welfare, where the disproportionate involvement of marginalized and vulnerable children and families raises ethical concerns about building fair and equitable models. one central issue in this debate is the over-representation of risk factors in algorithmic inputs and outputs, as well as the concomitant over-reliance on predicting risk. Would models perform better across groups if variables represented risk and protective factors associated with outcomes of interest? And, would models be more equitable across groups if they predicted alternative service outcomes? Using a risk and resilience framework applied in the field of social work, and the child welfare system as an illustrative example, the paper explores a strengths-based approach to predictive model building. We define risk and protective factors, then identify and illustrate how protective factors perform in a model trained to predict an alternative outcome of child welfare service involvement: the un-substantiation of an allegation of maltreatment.

Introduction

Recent trends have motivated human service organizations, and their funders, to leverage AI and sophisticated statistical methods. Most notable among these is predictive analytics (PA), used to predict the risk of experiencing adverse outcomes for human service system users,. The goals of these applications are, generally, to decrease system resource overload and increase favorable case resolutions [1, 2]. Despite PA's enthusiastic early reception (see, for example [3, 4, 5, 6]), recent research questions the ethics of its use within this sector [7, 8, 9]. In particular, the implementation of these methods in human services to augment high stakes decision making processes, as in the case of child welfare, has precipitated concerns regarding the ethical use of AI in sectors where decision making pathways are nebulous and where there is historic over-representation of marginalized groups [3, 8, 10]. Computer scientists, algorithmic developers, and engineers tasked with building AI projects in human service settings should be aware of ethical and theoretical debates in this context.

One central concern is the assumption that human behavior in the future will necessarily look like the past. Particularly in the case of risk assessment, the primary charge of PA in child welfare [11], this assumption relegates any current efforts of service users (i.e. change in socio-emotional circumstances, substance use rehabilitation, therapy, etc.) as irrelevant to their future outcomes. We argue PA requires context variables that discern what, if any, factors are mitigating the risk of experiencing an adverse outcome within the system's context. To do this, we propose reorienting model development in human services to a user-strengths perspective.

This paper briefly reviews the risk and resilience framework [6], advanced by social work scholars for developing social policy and delivering strengths-based service provision. The framework serves as a conceptual model for identifying context specific, risk-mitigating variables — termed protective factors — at the individual, family, and/or community levels. We illustrate the occurrence and effect of protective factors using a random forest model trained on the National Child Abuse and Neglect Data System (NCANDS) 2017 child file, a publicly available data set containing investigated reports of maltreatment allegations reported to State child protective agencies [12].

Risk and Resilience Framework

The 'strengths perspective' is an alternative to a pathology-oriented approach to interfacing with human service system users. Rather than focusing on the client's problems, deficits or risk profile, the strengths-based perspective focuses on clients' abilities, talents, and resources [13, 14]. Self-determination, empowerment, and social justice are inherent in the strengths perspective [15], and together comprise the ideal of human service provision. The risk and resilience framework builds on this perspective using empirically based knowledge of human behavior [16, 17].

Risk and resilience is a multi-theoretical framework for understanding how people achieve and sustain well-being despite adversity [18, 16]. The framework generally divides human characteristics and behaviors during times of crisis into two groups: risk factors and protective factors [17]. Risk factors are defined as characteristics or conditions that elevate the probability of an undesirable outcome [19] such as experiencing child abuse.

PA algorithms primarily focus on risk factors. Tables 1 and 2 below present an overview of risk factors in terms of their general conceptualization. Table 1 refers to individual level risk factors, while table 2 refers to family and community risk factors.

[Table 1 here]

[Table 2 here]

In order for PA outputs to be useful in building effective preventative human service interventions (their current goal), they must also include variables capturing characteristics which are known to help prevent negative outcomes, and thereby mitigate risk.

Termed 'protective factors', these are characteristics that promote resilience in times of crisis, or otherwise moderate the effect of risk factors [20, 19]. Protective factors operate by 1) reducing or buffering the impact of risk, 2) interrupting a chain of risk factors, or 3) preventing the onset of a risk factor [21]. Table 3 below summarizes the general conceptualization of protective factors.

[Table 3 here]

Research on protective factors within social work scholarship points to their dynamic nature. For example, one model of child maltreatment specifies risk and protective factors as either *transient*, fluctuating over time, or *enduring*, representing more stable conditions [22]. In child welfare, enduring protective factors may generally be found in initial assessment data, as they capture characteristics like parental education and prior involvement with child protective services. Transient factors include characteristics like participating in supportive services, difficult to include in static analyses not subject to regular iteration protocols.

In addition to the issue of iteration, administrative data, such as child welfare data, are generally cross-sectional and historical in nature: often times those entering the data (i.e. caseworkers) do not have the time or ability to update records with new information, and in fact are sometimes expressly prevented from doing so. For example, within child welfare, a prior incidence of domestic violence (DV) within a family will likely remain in a family's case record indefinitely. Should another call be made reporting suspicion of child neglect or abuse for the same family, the case may be automatically flagged as DV (a known risk factor), despite the possibility that family circumstances have changed during the intervening time period. To date, we have found no evidence that algorithm developers are aware of this, and similar, data gaps in contextualization.

Further complicating the issue of contemporality of data is the primary focus on risk. The United States' Children's Bureau (established in 1912 by the social work profession's founders [23]) is the federal department which regulates and funds local, state and tribal CPS jurisdictions. One of the Bureau's key missions is to "strengthen families and prevent child abuse and neglect" [24]. This prevention mandate is, arguably, one driving force behind the focus on risk in CPS PA applications. A second equally important driver is the nature of the CPS system itself: model building in this context relies exclusively on administrative data built specifically to assess risk. Together, these two forces result in a high tolerance for false positives within CPS analyses [25]. Importantly, given the general demographics of CPS involved communities, false positives predominantly impact marginalized or otherwise protected classes: namely, communities with low socio-economic status and communities of

color. Herein lies the greatest area of concern in the application of AI to human service decision-making.

Our aim in the current work is to offer a potential strategy to mitigate this tolerance for a high number of false positives by testing an alternative approach. Specifically, we train an algorithm to predict an alternative outcome of CPS involvement – the un-substantiation of an allegation of child maltreatment - then examine the variables most prominent in the prediction. This second step allows for the identification of protective factors within the existing data – what we define as case characteristics facilitating a less punitive exit from the CPS system following investigation.

Data

The National Child Abuse and Neglect Data System (NCANDS) collects administrative child welfare data from all 50 states, the District of Columbia, and Puerto Rico [12]. NCANDS data are comprised of records for all investigated reports of child maltreatment within a given fiscal year. These data are voluntarily submitted to NCANDS as a result of the Child Abuse Prevention and Treatment Act of 1988, and are publicly available through the National Data Archive of Child Abuse and Neglect hosted at Cornell University [12]. A full description of the data acquisition process, codebook, as well as the request form for the data themselves can be found on the NCANDS website [12].

We use the 2017 Fiscal Year Child File data set (v2), which contains demographic information on children and alleged perpetrators, types of alleged maltreatment reported, investigation results, risk factors, as well as services provided [12]. The data collection time frame is between October 1st, 2016 through September 30th, 2017. In total, the file contains N = 4,279,096 observations of 151 variables. We employ a systematic random sample of n = 12,239 of all variables. Removing columns containing more than 50% missing values, the final dataset used contains n=12,017 observations of 68 variables.

Methods

We train a Random Forest model using the randomForest package in R [26, 27]. The outcome of interest is the case disposition, or the finding that results from a CPS investigation of alleged child maltreatment. Within these data, this category is comprised of nine factor levels. We recode them to focus on three general investigation outcomes: substantiation (maltreatment is indicated by the investigation), un-substantiation (investigation finds no evidence of maltreatment), and alternative responses where the child(ren) is not reported as a victim(s).

The algorithm is trained to classify whether an investigation of alleged child maltreatment will result in un-substantiation. We choose this outcome for two reasons. First, given that over 60% of cases in the sample and full dataset are unsubstantiated, this outcome exemplifies an alternative pathway in the CPS system: exiting because no proof of maltreatment has been identified. Second, by learning the factors represented in the data that best predict case un-substantiation, we may learn more about how protective factors occur in existing human services datasets.

The final model specifies 750 trees and 12 variables to be randomly sampled as candidates for each split, with an error rate of 24.4%. Table 4 below offers the resulting confusion matrix from prediction on the test set.

[Table 4 here]

Diagnostics for the number of trees and variables selected, as well as the entire R code for this analysis, can be found in the following public GitHub repository [https://github.com/MariaYR/NCANDS_RF]. Following model training, we use the randomForestExplainer package in R [28] to visualize variable importance, variable interactions, as well as predictions for demographic and other variables of interest.

Results

Figure 1 summarizes the variables of importance in the final model.

[Figure 1 here]

The type of alleged child maltreatment (chmal) and the source of the initial maltreatment allegation

(rptsrc), depending on the measure of accuracy used, emerge as the strongest predictors for un-substantiation in the final model. Type of alleged maltreatment is somewhat intuitive to interpret as a key predictor: as the type of maltreatment shifts in verifiable severity, one would expect the likelihood of maltreatment verification to be associated with each shift, resulting in a greater likelihood of un-substantiation. For example, psychological maltreatment may be less tangible, and may include more subjective assessment for verification, than physical maltreatment. Relatedly, the source of a maltreatment report can vary anywhere from a medical professional to an anonymous individual. As a result of the variability in child well-being knowledge, as well as differences in motivations for reporting suspected maltreatment, there is a significant association between report source and the likelihood of case substantiation [29]. These predictors illustrate why simply choosing the 'strongest' predictor in a machine learning model does not necessarily translate to actionable output, at least in the context of human services.

Examining additional top predictors offers a foundation for understanding protective factors that increase the likelihood of case un-substantiation. For example, a classification of un-substantiation is partially dependent on the child's living arrangements (chlvng) at the time of report. Living with two or more committed adults has been found to be protective against child maltreatment and neglect [30]. Importantly, whether or not a child is living in foster care at time of initial report (fostercr) is also a top predictor in the model, further undergirding the protective force of a child living with adults known to them on whether or not a case is likely to be classified as unsubstantiated.

CPS system level variables also impact un-substantiation classification: the number of days between the initial report and investigation (rpt_to_inv) and from investigation to service provision (inv_to_srv) emerge as top predictors in the final model. The number of days between an initial report and the beginning of an investigation is understood to be critical to the protective capacity of CPS: longer times between initial report and family contact may lead to greater risk for the child [31]. More significantly, a lag between report and initial contact also prevents the provision of supportive services, which may address structural issues that lead to the initial report. For example, reports alleging certain types of child neglect (i.e. a child missing school for many days or being malnourished) are significantly associated with family financial issues, which may be readily addressed with services such as referrals for housing assistance or food relief services. Cases with maltreatment allegations stemming from financial difficulties are also likelier to result in alternative responses to investigation, where the child is not reported as a victim (rptdisp = 3). The importance of these variables in the final model highlight how CPS jurisdictions themselves may either augment or mitigate risk, a factor not explicitly accounted for in applications of PA in human services.

Figures 2 and 3 offer two variations of variable importance visualization. Figure 4 provides the minimal depth for the top 15 predictive variables in the final model.

[Figure 2 here]

[Figure 3 here]

[Figure 4 here]

Initial report source, child living arrangements, and the number of days between report, investigation and service provision occur in at least 75,000 nodes. Initial report source is associated with both high gini and accuracy decreases, while all four variables are significantly responsible for node splitting within the model. These visualizations also highlight another protective factor: the child's age at the time of initial report (ChAge). The older a child is, the more likely allegations of maltreatment may be due to structural factors, such as financial difficulties.

We turn now to our second line of inquiry: how protective factors interact with demographic groups of interest. Figure 5 below offers the mean minimal conditional depth of the 30 most occurring interactions in the model.

[Figure 5]

For our purposes, the most interesting interactions occur between the foster care placement (fostercr) variable and the child's identified race or ethnicity at the time of report. Figures 6 through 8 depict the model prediction for case disposition for all values of

both variables.

[Figure 6]

[Figure 7]

[Figure 8]

Across all three demographic groups, a CPS case involving a child not placed in foster care (probability_2), is more likely to result in a classification of 'un-substantiated'. Importantly, the converse relationship is found for substantiation (probability_1) for all groups. That is, a case involving a Black, Hispanic/Latinx, or White child not in foster care is more likely to result in a classification of 'un-substantiated' by the current model. This finding reinforces the evidence above: a child's living arrangement is a key protective factor when seeking to predict or classify risk of experiencing maltreatment.

Discussion

The analysis above offers a beginning exploration into how algorithms may benefit from predicting alternative outcomes when employed in human service contexts. We demonstrate how alternative pathway prediction can facilitate examining how protective factors, characteristics that mitigate the risk of experiencing an undesirable outcome, operate in the context of a classification model. The current model sought to classify child maltreatment investigations that resulted in 'un-substantiation', or cases where no maltreatment evidence was found.

We find that the source of the initial report and the type of maltreatment alleged are the strongest predictors in the model, and we contextualize these through a variety of visualizations. Results also demonstrate how protective factors, such as a child's living arrangements and services offered after an allegation of maltreatment, contribute to an 'un-substantiated' classification. That is, our results suggest that classifying cases likely to result in alternative system outcomes can facilitate identifying variables that decrease negative outcome risk. Further, we also demonstrate how such a model performs across demographic groups of interest, with implications for how protected classes might be better protected from undue surveillance by adopting a strengths-based lens to human service AI applications.

There are several noteworthy limitations to this exploratory work. First, due to computational resource limitations, we only employ a systematic random sample in the final model. Future work including the full data set, and spanning multiple years, would lend much support to this form of modeling. Second, we did not account for state differences in evidence thresholds for substantiating a child maltreatment case in the model. States vary in the amount of evidence required to confirm or deny maltreatment allegations, and future work would need to account for these variations for results to be meaningful in a policy or practice setting. Addressing both limitations would likely lead to an improvement in model accuracy as well.

Despite these limitations, we argue we have shown sufficient evidence to advocate for that inclusion of at least one derived variable in future applications of PA in human service contexts: probability of alternative outcome.

Acknowledgments

The analyses presented in this publication were based on data from the National Child Abuse and Neglect Data System (NCANDS) Child File, FFY 2017v2. These data were provided by the National Data Archive on Child Abuse and Neglect at Cornell University, and have been used with permission. The data were originally collected under the auspices of the Children's Bureau. Funding was provided by the Children's Bureau, Administration on Children, Youth and Families, Administration for Children and Families, U.S. Department of Health and Human Services. The collector of the original data, the funding agency, NDACAN, Cornell University, and the agents or employees of these institutions bear no responsibility for the analyses or interpretations presented here. The information and opinions expressed reflect solely the opinions of the authors.

The authors wish to thank Gleneara E. Bates-Pappas and Sebastian Hoyos-Torres for their research assistance.

References

[1] Vaithianathan, R., Maloney, T., Putnam-Hornstein, E., & Jiang, N. (2013). Children in the public benefit system at risk of maltreatment: Identification via predictive modeling. *American journal of preventive medicine*, *45*(3), 354-359.

[2] Pearsall, B. (2010). Predictive policing: The future of law enforcement. National Institute of Justice Journal, 266(1), 16-19.

[3] Fishman, T.D., Egger, W. D., & Kishani, P. (October 18, 2017). Ai-Augmented Human Services: using Cognitive Technologies to transform program delivery. Retrieved from https://www2.deloitte.com/insights/us/en/industry/public-sector/artificial-intelligence-technologies-human-services-programs.html.

[4] Teixeira, C. & Boyas, M. (October 2017). Predictive Analytics in Child Welfare: An assessment of current efforts, challenges, and opportunities. Retrieved from https://aspe.hhs.gov/system/files/pdf/257841/PACWAnAssessmentCurrentEffortsChallengesOpportunities.pdf.

[5] Microsoft (2017). Artificial Intelligence transforms even the most human services. Retrieved from https://news.microsoft.com/en-au/features/artificial-intelligence-transforms-even-human-services/.

[6] Jenson, J. M., & Fraser, M. W. (2016). A risk and resilience framework for child, youth, and family policy. In J. Jenson & M. Fraser (Eds), Social policy for children & families: A risk and resilience perspective, (pp. 5-21). Thousand Oaks, CA: Sage Publications, Inc.

[7] Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

[8] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. *ProPublica, May*, *23*.

[9] Lanier, P., Rodriguez, M., Verbiest, S., Bryant, K., Guan, T., & Zolotor, A. (2019). Preventing Infant Maltreatment with Predictive Analytics: Applying Ethical Principles to Evidence-Based Child Welfare Policy. *Journal of Family Violence*, 1-13.

[10] Fluke, J., Jones Harden, B., Jenkins, M., & Ruehrdanz, A. (2010). *Research synthesis on child welfare disproportionality and disparities*. Washington, DC: Alliance for Racial Equity in the Child Welfare System.

[11] Russell, J. (2015). Predictive analytics and child protection: Constraints and opportunities. *Child abuse & neglect*, *46*, 182-189.

[12] U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau (2017). National Child Abuse and Neglect Data System (NCANDS) Child File, FFY 2017 [Dataset]. Available from the National Data Archive on Child Abuse and Neglect Web site, http://www.ndacan.cornell.edu

[13] Kim, J. (2013). Strengths perspective. *Encyclopedia of Social Work*. Washington, DC: National Association of Social Workers.

[14] Saleebey, D. (2006). *The strengths perspective in social work practice* (4th ed.). Boston: Allyn and Bacon.

[15] Blundo, R. (2013). Strengths-based framework. *Encyclopedia of Social Work*. Washington, DC: National Association of Social Workers.

[16] Greene, R. R. (2013). Resilience. *Encyclopedia of Social Work*. Washington, DC: National Association of Social Workers.

[17] Rutter, M. (1987). Psychosocial resilience and protective mechanisms. American Journal of Orthopsychiatry, 57(3), 316-331.

[18] Fraser, M. W. (2004). *Risk and resilience in childhood* (2nd edition). Washington, DC: NASW Press.

[19] Masten, A. S., & Wright, M.O. (1998). Cumulative risk and protection models of child maltreatment. In B.B.R. Rossman & M.S. Rosenberg (Eds). *Multiple victimization of children* (pp. 7-30). New York: The Haworth Press, Inc.

[20] Capacity Building Center for States (2018). *Protective Capacities and Protective Factors: Common Ground for Protecting Children and Strengthening Families*. Retrieved from https://library.childwelfare.gov/cwig/ws/library/docs/capacity/Blob/107035.pdf?r=1&rpp=25&upp=0&w=NATIVE%28%27SIMPLE_SRCH+ph+is+%27%27protective+factors+framework%27%27%27%29&m=1&order=native%28%27year%2FDescend%27%29. Washington, DC.

[21] Fraser, M. W., & Terzian, M. A. (2005). Risk and resilience in child development: Principles and strategies of practice. *Child Welfare for the Twenty-First Century: A Handbook of Practices, Policies and Programs*, 55-71.

[22] Cicchetti, D., & Lynch, M. (1993). Toward an ecological/transactional model of community violence and child maltreatment: Consequences for children's development. *Psychiatry*, *56*(1), 96-118

[23] Rodriguez, M. Y., Ostrow, L., & Kemp, S. P. (2017). Scaling up social problems: Strategies for solving social work's grand challenges. *Research on Social Work Practice*, *27*(2), 139-149.

[24]CB Fact Sheet. (n.d.). Retrieved July 17, 2019, from Children's Bureau | ACF website: https://www.acf.hhs.gov/cb/fact-sheet-cb

[25] Kahn, N. E., Gupta-Kagan, J., & Eschelbach Hansen, M. (2017). The standard of proof in the substantiation of child

abuse and neglect. *Journal of Empirical Legal Studies*, *14*(2), 333-369.

[26] Liaw, A. and Wiener , M. (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

[27] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

[28] Paluszynska, A. & Biecek, P. (2017). randomForestExplainer: Explaining and Visualizing Random Forests in Terms of Variable Importance. R package version 0.9. *https://CRAN.R-project.org/package=randomForestExplainer*

[29] Eckenrode, J., Powers, J., Doris, J., Munsch, J., & Bolger, N. (1988). Substantiation of child abuse and neglect reports. *Journal of consulting and clinical psychology*, *56*(1), 9.

[30] Brown, S. L. (2004). Family structure and child well-being: The significance of parental cohabitation. *Journal of Marriage and Family*, *66*(2), 351-367.

[31] DePanfilis, D. (2018). Child Protective Services: A Guide for Caseworkers. Child Abuse and Neglect User Manual Series. Capacity Building Center for States. Washington, D.C.

[32] Centers for Disease Control and Prevention (2018). Child Abuse and Neglect: Risk and Protective Factors. Washington, D.C.

Maria. Y. Rodriguez, MSW, PhD. Silberman School of Social Work (Hunter College, City University of New York). 2130 Third Avenue New York, NY 10035 (mr3284@hunter.cuny.edu). Dr. Rodriguez received her Ph.D. from the University of Washington (2016). She is an assistant professor at the Silberman School of Social Work and a faculty associate at the Berkman Klein Center for Internet and society (2019-2020). Dr. Rodriguez currently serves on the editorial boards for the Journal of Technology in Human Services and the Journal of Community Practice. She has presented at numerous conferences, including the Society for Social Work and Research, the Population Association of America, and the Urban Affairs Association. Her work examines the ethical implications of algorithmic decision-making in human services, as well as the use of social media data for social work intervention development.

Diane DePanfilis, MSW, PhD. Silberman School of Social Work (Hunter College, City University of New York). 2130 Third Avenue New York, NY 10035 (diane.depanfilis@hunter.cuny.edu). Professor Diane DePanfilis teaches and conducts research related to child welfare policies, programs, and practices. She has published extensively on the epidemiology of child maltreatment recurrences and has led the design, testing, and implementation of federally funded community based interventions focused on preventing child maltreatment and on supporting systems to use evidence and data to inform decision-making related to child welfare policies and programs. Dr. DePanfilis is a former Vice President of the Society for Social Work and Research and a former President of the American Professional Society on the Abuse of Children.

Paul Lanier, MSW, PhD University of North Carolina School of Social Work. 325 Pittsboro Street Chapel Hill, NC 27516. Dr. Lanier is an Assistant Professor in the School of Social Work and is a faculty affiliate with the Cecil G. Sheps Center for Health Services Research. He received his Ph.D. from the Brown School at Washington University in St. Louis where he was a pre-doctoral fellow with the National Institutes of Health Ruth L. Kirschstein Institutional National Research Service Award. His research focuses on the prevention of child maltreatment and children's mental health services.

Tables

Table 1 Conceptualization of Individual Level Risk factors (adapted from [32])

Individual Level Risk Factors	Description
Age	Below age of 18 at behavior
Special Needs	Mental health, developmental differences
Prior History	Prior experience of adverse behavior/outcome
Substance abuse	Unmediated history of substance addiction
Lack of knowledge	No formal or informal understanding of adverse event
Transient social connections	Lack of consistent peer or role models, unstable family connections

Table 2 Conceptualization of Family & Community Risk Factors (adapted from [32])

Risk Factor	Description
1. Social Isolation	Little engagement with civic/neighborhood institutions, no formal/informal social supports
2. Community violence	Individual/family resides in area with high visible police presence, number of arrests
3. Family disorganization, dissolution and/or violence	Little formal or informal attachments between family members, domestic violence (DV), estrangement between caregivers/family members
4. Concentrated Neighborhood Disadvantage	High rates of poverty, residential tenure, unemployment, etc.

Table 3 Conceptualization of Protective Factors (adapted from [32])

Protective Factor	Description
1. Nurturing and Attachment	Evidence of positive emotional bonding between a child and caregiver
2. Parental Resilience	Evidence of caregiver emotional flexibility and coping skills to support navigating daily and unforeseen life stressors
3. Social Connections	Evidence of strong social ties, emotionally supportive social networks
4. Concrete support in times of need	Ability to provide for basic needs, and/or knowledge and ability to access required services
5. Knowledge of Parenting and Child Development	Evidence of formal and informal knowledge regarding child development and caregiving skills that foster positive youth development
6. Social and Emotional Competence	Evidence caregiver models effective communication, self-regulation, and positive social interactions

Table 4 Confusion Matrix of Prediction on Test Set

	Un-substantiation	Substantiation	Alt. Response
Un-substantiation	2098	465	214
Substantiation	114	267	13
Alt. Response	79	12	344

Figures

Figure 1 A plot depicting variable importance in the final random forest model. Depending on the metric of importance used, the type of maltreatment alleged (chmal1) or the report source (rptsrc) emerge as most predictive variables. Importantly, other variables indicate unsubstantiation classification is also dependent on a child's living arrangements (chlvng), whether family supportive services were offered (famsup), whether a caretaker has emotional health issues (fcemotnl), as well as the number of days between initial report and investigation (rpt_to_inv) and between investigation and service provision (inv_to_srv).

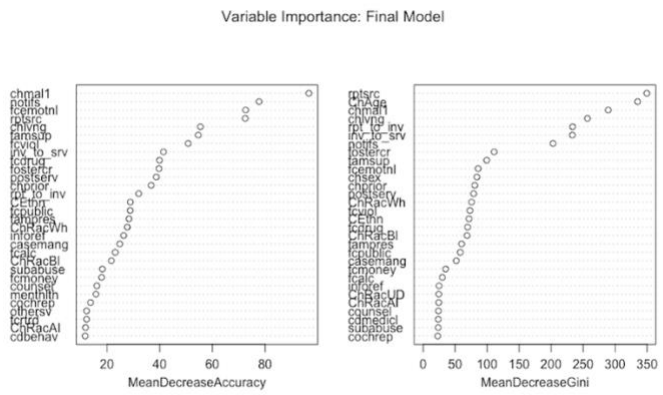

Figure 2 A plot depicting variable importance in the final model. The x-axis refers to the accuracy decrease associated with removing the variable from the model, while the y-axis refers to the Gini decrease. The size of the circle corresponding to each variable indicates the number of nodes the variable appears in.

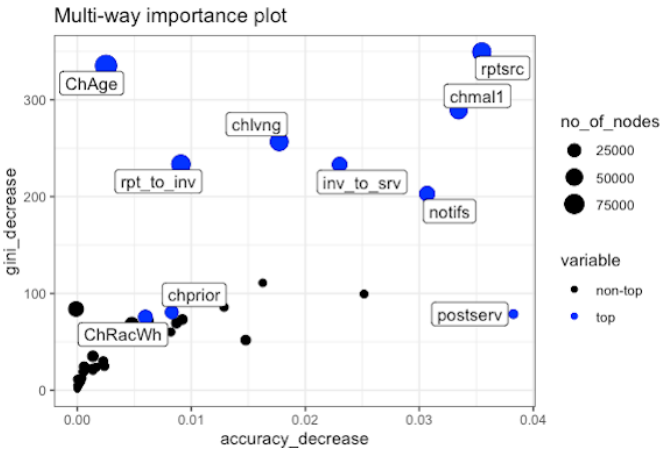

Figure 3 A variation of the multi-way importance plot in figure 2, with p-values for the prediction. P-values indicate variables are used for splitting more than expected if the selection were random.

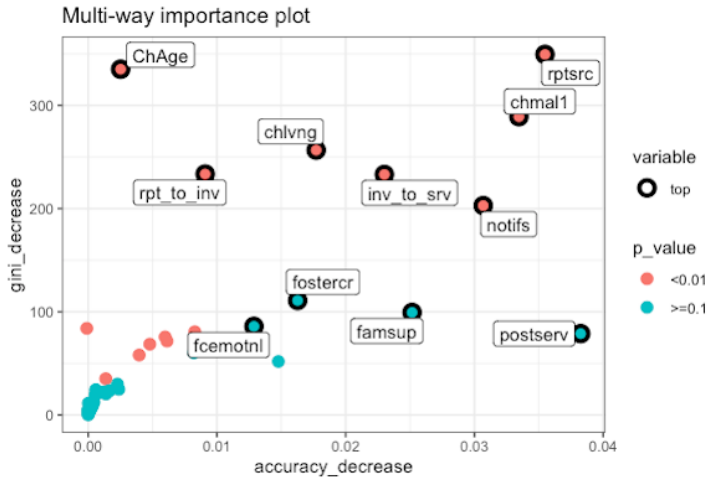

Figure 4 A plot depicting the distribution of minimal depth and minimal depth mean for the top 15 predictive variables in the model. The mean of the variable's distribution is marked with a vertical bar.

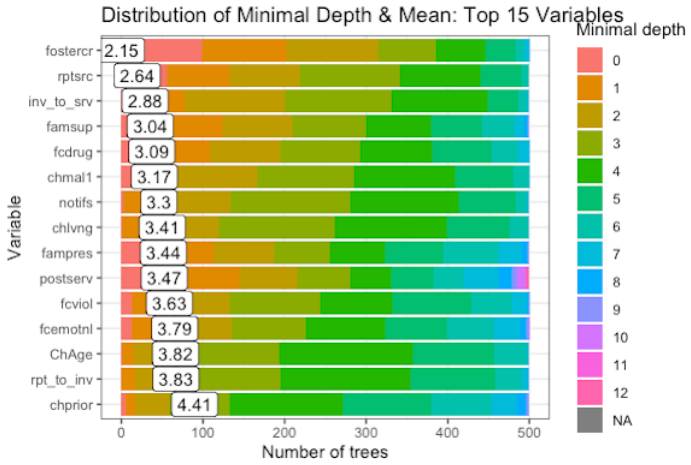

Figure 5 A plot depicting the mean conditional minimal depth for the 30 most frequent interactions in the final model. This is a generalization of minimal depth derived by the randomForestExplainer package [28], which measures the depth of the second variable in a tree where the first variable is the root.

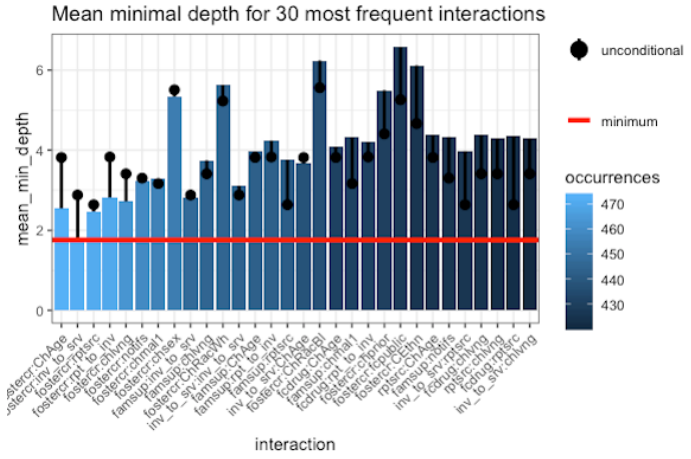

Figure 6 Model prediction for foster care placement (1 = yes) and child's race identified as black (1 = yes) interaction.

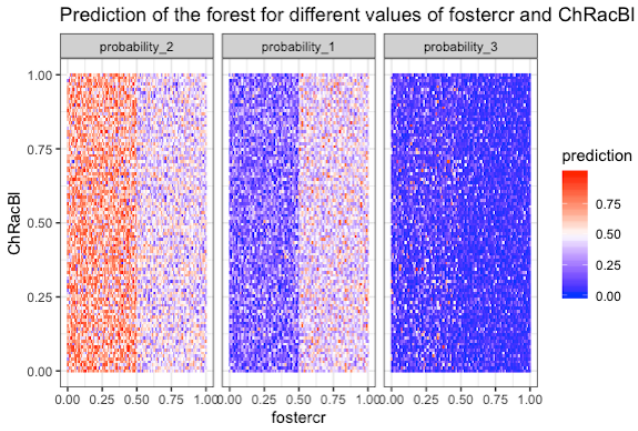

Figure 7 Model prediction for foster care placement (1 = yes) and child's ethnicity identified as Hispanic/Latinx (1 = yes) interaction.

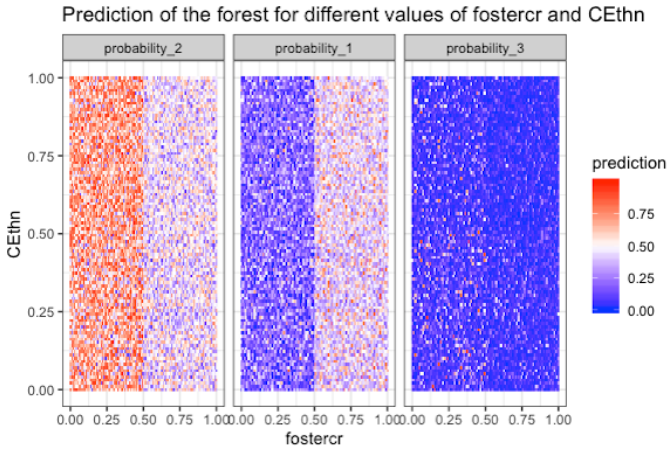

Figure 8 Model prediction for foster care placement (1 = yes) and child's race identified as white (1 = yes) interaction.