Modelling social vulnerabilities and eviction cases in the Fragile Families and Child Wellbeing Study with Gradient Boosted Decision Trees

Adriano Belisario

April, 2023

1 Introduction

This paper presents a data-driven approach to examining the results of the Fragile Families and Child Wellbeing Study (FFCWS), a survey on vulnerability in families from the United States. I evaluate the performance of four machine learning models using six key metrics: GPA, grit, material hardship, layoff, eviction, and job training. The methodology adopted aims to identify significant features and develop interpretable models. The study delves into eviction cases, providing a detailed analysis of this issue and discussing the challenges and opportunities associated with using the FFCWS data.

I employ three Gradient Boosted Decision Trees algorithms and linear/logistic regression to model all six metrics of interest. The results indicate that CatBoost outperforms all other models for nearly all target variables. I discuss the importance of economic factors and financial instability to understand eviction cases and address potential future research directions, including novel data collection approaches and methodological choices such as dimensionality reduction techniques.

2 Context

The Fragile Families and Child Wellbeing Study (FFCWS) is a birth cohort panel research that involved more than 4,000 children born in 20 large U.S. cities and their parents. Data was collected through surveys at five different moments, or survey waves, starting from the child's birth (wave 1) and continuing through their early adulthood (approximately age 15). Over the years, researchers have used FFCWS data to explore various topics, such as the effects of paternal incarceration or academic performance (Kumar, 2022; Wildeman, 2010).

The Fragile Families Challenge was designed to leverage this data and pose a question to scientists and researchers: how do socioeconomic and environmental factors experienced by families during childhood impact six key outcomes related to children's well-being? (Salganik et al., 2020). Early signals associated with issues like eviction or educational performance could help researchers and policymakers develop policies to promote well-being. The Challenge was a landmark initiative in social data science, attracting contributions from 160 teams, and required participants to use a version of FFCWS survey data to predict these life outcomes. However, the results were limited in terms of the accuracy of predictions. Despite the variety of academic expertise and methodologies, "the submissions were much better at predicting each other than at predicting the truth" (Salganik et al., 2019).

Unlike some of the previous works done for this challenge (Ahearn & Brand, 2019; Compton, 2019; Rigobon et al., 2019), my research focuses on interpretability rather than predictability. Salganik et al. (2020) discuss the tension between predicting and understanding life outcomes. Further debate on these concepts is beyond the scope of this article; nonetheless, to illustrate, according to the authors of the winning models for half of the Challenge's target variables (Rigobon et al., 2019), the most important features for predicting material hardship were the total length of the interview and the survey weight for the mother's city. While these variables may have some predictive power for this dataset, they do not enhance our understanding of the social causes of material hardship. Therefore, my focus on interpretability informed essential decisions regarding the selection of features and the choice of statistical model, which are presented in detail below.

3 Methods

This research applies a literature review, statistical methods and a case study. For the literature review, I concentrated on academic works that applied machine learning techniques to the Fragile Families Challenge and supplementary bibliography regarding eviction cases. The literature review also involved a thorough examination of the dataset documentation.

I also conducted exploratory data analysis on the metadata, dependent, and independent variables to better understand the underlying dataset structure. Because my primary goal was not to benchmark algorithms performance but use different methods appropriate to this dataset, I tested different models with different settings of data preprocessing. Finally, the case study used the best model and all data currently available to interpret relevant factors for understanding eviction cases. In the next section, I describe the data source, preprocessing steps, statistical models and evaluation methods.

3.1 Data source

I used structured metadata from the FFCWS Metadata Explorer[1] and the fifth version of the datasets available on the Fragile Families Challenge section on the FFCWS webpage. This webpage contains data available for training (background data) and files that were not available for researchers during the challenge, such as the target values used to score models (leaderboard and holdout data).

In order to ensure comparability with prior research on this Challenge, I refrained from using leaderboard or holdout data for training during the model development,

[1] `https://metadata.ffcws.princeton.edu/`

using them solely for evaluation purposes and for interpreting the most important variables with the top-performing model. The background data was partitioned into training and test sets using an 80/20 split and cross-validation.

3.2 Target variables

These outcomes represent answers reported by the child or their primary caregiver in the fifth wave, at approximately 15 years old. There are three continuous (GPA, grit and material hardship) and three binary outcomes (layoff, eviction and job training):

1. **Grade Point Average (GPA)**: measures a student's performance, calculated as the average of all their grades on a scale of 0-4. The FFCWS relies on self-reported GPA data provided by the child[2].

2. **Grit**: reflects a child's perseverance towards achieving long-term goals. The FFCWS uses a definition and scale adapted from the work of Duckworth et al. (2007).

3. **Material hardship**: represents the proportion of positive answers to 11 questions related to financial difficulties. The Challenge webpage [3] and Salganik et al. (2019) do not precisely reference the code of the variables used, but it was possible to identify some of them matching the text description of the question. Section 3.3 provides further details about how these features were handled.

4. **Layoff**: whether the child's primary caregiver has recently lost their job due to layoffs.

5. **Eviction**: whether the child's primary caregiver has been evicted from their home due to nonpayment of rent or mortgage, "regardless of whether a court ordered the eviction or a landlord carried it out informally" (Salganik et al., 2019, p.8).

6. **Job training**: whether the child's primary caregiver has taken classes to improve their job skills, such as computer training classes.

3.3 Feature selection

The FFCWS dataset has a high dimensionality with 13,026 independent variables and many missing values. According to Salganik et al. (2019), the background dataset has approximately 55 million distinct data entries, with more than 73% of missing values. Additionally, one-hot encoding further increases the number of dimensions due to the high cardinality of some categorical variables. To select relevant features for the model, I established the following rules:

[2] `https://www.fragilefamilieschallenge.org/gpa/`

[3] `https://www.fragilefamilieschallenge.org/material-hardship/`

- Variables listed as constant in the documentation or with low variance were deleted from the model.
- Features with strong correlations were eliminated to minimise multicollinearity. This rule not only improves computational efficiency during model training but also enhances the interpretability of Shapley values.
- Structured metadata was required for automated feature engineering, and only variables with numerical, categorical or binary data types were included. A few text string columns with constructed variables were discarded.
- Duplicated and malformed variables were deleted to ensure data integrity. An example of a malformed variable is "cf4fint"[4], which is referenced as a binary feature in the FFCWS Metadata Explorer but has a timestamp as values.
- To prevent target leakage, independent variables (e.g., "m5f23e"[5] or ""m5f23k"[6]) from the fifth wave potentially used to encode target variables (namely, material hardship) were deleted.
- Variables with less than 70% of non-missing values, including those with negative values indicating missing answers, were excluded.
- Variables related to survey weights and paradata (such as the total length of the interview) were excluded. Only features defined as asked in the survey or constructed by the Fragile Families research team were selected.

Applying the criteria above resulted in a significant reduction of independent variables from 13,026 to 1,204. Nevertheless, the number is still higher than the total of observations available for training (1,172). Although this standard, especially the last two rules, excluded numerous features, it ensured that only meaningful independent variables would be used to model the six metrics of interest.

3.4 Models

Tree ensemble models, specifically gradient-boosted decision trees (GBDT), are a highly reliable option for classification and regression tasks involving tabular data. Research has shown that they can outperform deep learning techniques in some cases (Shwartz-Ziv & Armon, 2022), and they have already proven their worth in online competitions, including the Fragile Families Challenge itself (Rigobon et al., 2019). On top of that, various implementations have built-in support for missing values. I have tested three popular GBDT models, namely XGBoost, LightGBM, and CatBoost, along with a linear and logistic regression model.

I applied one-hot encoding to transform all categorical variables into numerical values and used Bayesian hyperparameter optimisation with k-fold cross-validation (k=5) except for CatBoost. This approach allowed me to test different methods and

[4] "Constructed - Father interviewed at five-year follow-up"
[5] "Did not pay full amount of gas/oil/electricity bill in past 12 months"
[6] "Telephone service disconnected because wasn't enough money in past 12 month"

regularisation values to mitigate the high dimensionality of the dataset and prevent overfitting. Models were trained using a personal computer (Intel i7 CPU with no GPU), and I capped the number of iterations for hyperparameter search to 10.

During exploratory data analysis on the dependent variables, I found a strong class imbalance in the three binary outcomes, with the minority class representing only a small sample ranging from 5 (eviction) to 25% (layoff) of the cases. In order to tackle issues like high dimensionality and class imbalance, various choices were made concerning the model parameters. A detailed account of the code, data analyses and the hyperparameters used can be found in the following anonymous repository: `https://anonymous.4open.science/r/ff_applied_ml-9B2C`.

Next, I provide a summary of each model. Other methods and feature selection procedures were considered or tested but not included in this final report. I mention some of them in my final considerations.

Model 1 - Lasso

The first model employed linear and logistic regression with Lasso (L1) regularisation. I used k-Nearest Neighbors (k=5) to impute data in rows with missing values and Synthetic Minority Over-sampling Technique (Chawla et al., 2002) to increase the representation of the minority class. The parameters optimised with Bayesian search were regularisation strength (*alpha*) and the number of iterations.

Model 2 - LightGBM

The second model used a LightGBM algorithm with Bayesian parameter optimisation for the learning rate and max depth of the tree. For binary outcomes, an additional parameter was optimised to scale class weights (*scale_pos_weight*).

Model 3 - CatBoost

The third approach ran CatBoost, a GBDT model with built-in algorithms to handle categorical features; hence, I did not use one-hot encoding to preprocess data. Similarly, instead of Bayesian search, I also adopted CatBoost built-in methods for determining the best hyperparameters values based on the number of iterations and characteristics of the dataset.

Model 4 - XGBoost

The XGBoost model was tuned using Bayesian search and cross-validation following the implementation of Rigobon et al. (2019). The hyperparameter search involved finding the optimal values for the learning rate, the number of decision trees in the ensemble (estimators), the maximum depth of each decision tree, the minimum number of samples required to create a new node in the tree (*min_child_weight*), and the minimum improvement required to create a new split in the tree (*gamma*).

3.5 Evaluation and metrics

I analysed the cross-validation, test, and leaderboard scores to ensure a robust evaluation. For continuous dependent variables, the main metric reported is Mean Squared Error, in line with the metrics adopted on the Challenge leaderboard. However, instead of relying only on the Brier score, which was adopted as the official metric for continuous variables in the public leaderboard, I also used the F1-score of the positive class to examine the model's performance. This decision was motivated by the fact that the Brier score has interpretability limitations for imbalanced data, as discussed by Wallace and Dahabreh (2012) and exemplified in the case study below.

Binary outcomes (eviction, layoff and job training) were evaluated in two ways: with and without threshold adjustment. I used Youden's J statistic to find an optimal threshold to distinguish positive and negative classes and plotted the ROC curve to analyse the performance of classifier models. For simplicity, the values reported in the findings section are based on the best results of each model for leaderboard data. However, all values obtained are available on the repository.

Once the best model was identified, I evaluated it using the holdout dataset and fitted it to all data currently available to list the five most important features for eviction. The importance here is defined in terms of impact on model output, according to each dependent variable's SHAP values (Shapley).

4 Findings

4.1 Model comparison

The model with the CatBoost algorithm performed better than the others for all target variables except Grit, to which the first model (Lasso) performed better. Because I use Mean Squared Error for continuous variables, lower values mean better performance for the three first lines. In contrast, higher values (F1-scores) represent better models for eviction, layoff and job training. Numbers in bold represent the best model for each of these life outcomes.

	Lasso	**LightGBM**	**XGBoost**	**CatBoost**
GPA	0,386	0,385	0,385	**0,382**
Material Hardship	0,028	0,025	0,025	**0,244**
Grit	**0,221**	0,225	0,289	0,249
Eviction	0,164	0,122	0,148	**0,177**
Job Training	0,339	0,377	0,298	**0,433**
Layoff	0,211	0,377	0,133	**0,382**

Table 1: Comparison of model performance.

4.2 Study case: factors affecting families evicted

The model with better performance (CatBoost) was fitted using all training data available (background dataset) and presented an F1-score of 0.190 in the holdout dataset. Without threshold adjustment, the holdout and leaderboard Brier score loss are respectively 0.0589 and 0.0566. However, in this case, the model only predicts the majority class (F1-score equals zero). With threshold optimisation, the Brier scores for the holdout and leaderboard datasets are 0.270 and 0.351.

Figure 1 shows the independent variables with the highest impact on eviction cases. For each variable (vertical axis), the chart shows all data points coloured according to their values. Since all top-five variables are binary, blue dots represent "yes", and red dots indicate "no", while grey dots mean missing values. Positive

Shapley values (represented in log odds) increase the chances of being evicted, while negative values decrease them.

The first letter in the variable code indicates who answered the question in the survey ("m" for mother), and the second character, the number, indicates the survey wave. Therefore, the mother answered all variables referenced in the chart.

The most important variables are 'm5f23g' ("Borrowed money from friends/-family to help pay bills in past 12 months") and 'm5i3' ("Have completed training programs or years of schooling since last interview"). The third variable referenced is "m2h19f" ("In past year, did you not pay full gas/oil/electric bill?"), followed by "m4i23n", representing whether the phone service had been disconnected due to nonpayment in the past year before wave 4. The fifth variable is "m5i3c" ("You received any kind of employment counseling since last interview").

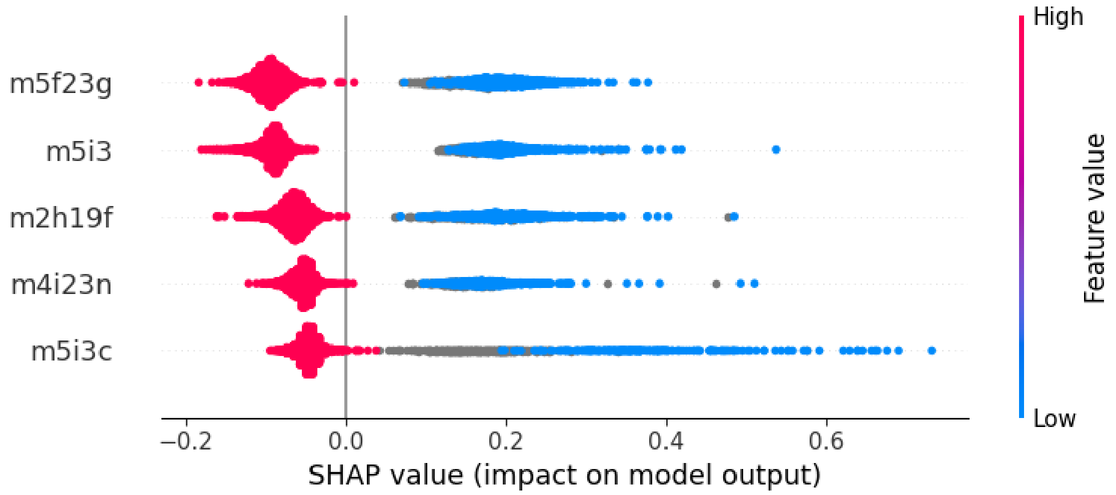

Figure 1: Top features ranked by importance.

5 Discussion

5.1 Model comparison

The CatBoost model outperformed all other approaches for almost all target variables. One possible explanation for this is that the CatBoost is a GBDT model designed to handle categorical variables efficiently, and most of the features in the Fragile Families dataset have this data type. CatBoost utilises various statistical methods for modelling categorical variables, such as target encoding, rather than relying solely on one-hot encoding. Another contributing factor could be the model's default methods for determining hyperparameters, which may have performed better than the Bayesian search.

It is important to note that the scores obtained were limited by the runtime constraints imposed by the iterations of the Bayesian search. Improved results might be achievable using more powerful machines and longer training times. Nevertheless, the F1-score values obtained for the positive classes in eviction cases surpassed the performance of previous works such as Compton (2019).

The holdout Brier score might appear worse (the smaller the Brier score loss, the better) than the top-ranked leaderboard models[7], but this metric may be biased for imbalanced classification tasks. For instance, a model that predicts only the majority class for evictions would have a better Brier score than a model that can effectively differentiate classes. To illustrate this point, consider the LightGBM and CatBoost models. Without threshold adjustment, they predicted only the majority class (non-evicted) for the leaderboard dataset, achieving a Brier score loss of 0.0566, better than dozens of models listed on the leaderboard. Nonetheless, these models have an F1-score of zero, indicating they are unlikely to be useful for interpretation or real-world applications.

5.2 Families affected by eviction

Results showed that the top-performing model for eviction cases had similar F1 scores for the leaderboard and holdout datasets, indicating that it is not overfitted and adequately captures general patterns in the data. Moreover, the most important features offered insights regarding the social context of evictions.

The study found that questions related to the families' economic condition and financial instability played an essential role in understanding eviction cases. Three of the five most important variables are related to difficulties paying regular bills. Figure 1 shows that mothers who have borrowed money to pay bills or have not paid them (blue dots on the right side of the chart for the variables m5f23g, m2h19f, and m4i23n) had increased changes to being evicted. This conclusion does not come as a surprise due to the very definition of this dependent variable, which means that the child's primary caregiver has been evicted from their home due to nonpayment.

My findings are consistent with previous academic research (Clarke et al., 2017; Rigobon et al., 2019) that has noted that eviction and poverty are two highly correlated problems. For the Fragile Families Challenge, it means that some of the variables used to encode material hardship are useful predictors for eviction. It is the case of "m5f23k"[8] that was the most important variable for predicting eviction, according to Rigobon et al. (2019). As noted before, these variables were removed to avoid target leakage; including them might improve the performance of eviction prediction models but would not necessarily increase our interpretability of real eviction cases.

Moreover, poverty and eviction are intertwined social issues that often occur together, with poverty leading to eviction and eviction exacerbating poverty. My findings support the fact that financial difficulties increase the chances of being evicted, but previous research has noted that the opposite is also true. For example, other research highlighted that "an eviction order increases homelessness, and reduces earnings, durable consumption, and access to credit" (Collinson et al., 2022, p.7).

On the other hand, the similarity between these variables ("m5f23g", "m2h19f",

[7] https://codalab.fragilefamilieschallenge.org/#results

[8] "Telephone service disconnected because wasn't enough money in past 12 months."

and "m4i23n") may also suggest that the threshold adopted to remove colinear features might have been still too permissive. As shown in the notebook available in the repository[9], an extended look into the top ten most important variables also reinforce this hypothesis: the sixth most important dependent variable was "m2h19h", which indicates whether the telephone service had been disconnected, an answer very similar to the fourth variable in the rank ("m4i23n"). The only difference between these variables is the survey wave, i.e. the year when they were answered. Therefore, feature engineering aggregating similar questions might achieve better statistical performance and interpretability; this methodology is further discussed below. Additionally, multicollinearity in the dataset might also hurt the interpretation of Shapley values because their calculation assumes that dependent features are uncorrelated (Basu & Maji, 2022).

Figure 1 demonstrates that mothers who received employment counselling were more likely to face eviction, supporting the link between financial instability and eviction. Employment counselling is often provided to individuals with difficulty securing stable job positions, indicating that job insecurity is a significant risk factor for eviction. This finding is consistent with prior research that has identified unemployment and underemployment as critical contributors to eviction (Campbell et al., 2013).

However, the finding that individuals who completed training programs since their last interview were more likely to be evicted does not have a straightforward interpretation. One hypothesis is that this feature might be correlated with "m5i3c", which asserts whether the respondent has completed training programs since the last interview. In that case, it would reinforce the link between unstable job positions and eviction. Unfortunately, the methods employed in this research cannot confirm such hypothesis, making it challenging to draw definitive conclusions about the relationship between education, training, and eviction.

It is worth noting that eviction cases might be the most challenging target variables to predict in the Fragile Families dataset due to the extreme class imbalance that provides only a few cases with positive classes (about 5% of the total). As well as previous attempts to model eviction using this dataset, my model has relatively low scores, indicating a low capacity for predicting eviction cases. These low rates suggest that factors and dynamics not captured by the survey questions play a significant role in defining this life outcome.

Scholars and evidence-based research like the FFCWS have successfully identified predictors that increase the chance of someone being evicted. However, because eviction is a highly contextual decision made by an agent, and social dynamics are highly unpredictable, current methods have not achieved high accuracy scores for predicting these cases.

5.3 Future research

To better understand the complex issue of eviction, it is imperative to collect data on the house owner. Previous quantitative research, for example, has differentiated between private and social landlords (Clarke et al., 2017). While background infor-

[9]https://anonymous.4open.science/r/ff$_{a}pplied_{m}l - 9B2C$

mation of those facing eviction is crucial, it is essential to recognise that the house owner is the person that ultimately decides to evict someone. Various factors — such as whether the house is managed by an individual or a company or the economic landlord's status — may impact eviction rates. By incorporating this information, a more nuanced understanding of eviction can be attained.

Consequently, the FFCWS survey design focused on tenants adopted is a significant limitation to model eviction cases. The FFCWS also fails to capture geographical components relevant to understand eviction cases, such as neighbourhood gentrification (Desmond & Gershenson, 2017). Nevertheless, the dataset still holds valuable insights that can be used to comprehend eviction cases better. Further research with the FFCWS data might consider the following approaches.

Apart from its high dimensionality, the FFCWS dataset has collinearity issues, i.e. multiple questions measuring the same construct. Examples are variables associated with household income or constructed features that reuse information from survey variables. While regularisation parameters and modern machine learning models can mitigate this, structured metadata and dimensionality reduction techniques might also be applied to summarise variables meaningfully. For instance, variables can be grouped into topics like finance or cognitive skills. Creating these categories would require a careful review of the variables because the topics assigned to the features in the structured metadata are too generic, e.g. finances or education and school. Therefore, using them to group variables might result in uninformative predictions, such as that finances are relevant to define material hardship or education features are relevant to predict GPA.

Another possible methodology for feature selection is creating a set of variables with high predictive power using mixed methods. Qualitatively, a literature review combined with a search in the FFCWS data dictionary can identify independent features potentially correlated to a given target variable. Quantitatively, mutual information regression can help identify other independent variables with high predictive power that academic research might not have previously reported.

Regarding statistical models, three different options might also be considered. Firstly, weighted models can incorporate survey weights available in the FFCWS dataset, which might improve statistical inference capabilities. Secondly, multilevel models might be appropriate because the FFCWS dataset stacks variables asked for the same family in different years. Therefore, appropriate feature engineering techniques are necessary to allow any statistical model to fully capture the longitudinal dimension of the FFCWS. The year or survey wave number, referenced only in the variable name with the standard structure of the dataset, might be used as an extra variable to represent this time component. Then, multilevel regression or classification models could be applied, but this method would possibly require carefully assessing this dataset's complex skipping and missing value patterns.

Finally, automated machine learning is a recently developed field that might be useful. Open-source tools such as AutoGluon[10] or AutoSkLearn[11] can ensemble different machine learning models and apply Bayesian search to find optimal hyperparameters not only for the algorithm but also for the feature processing choices.

[10]`https://auto.gluon.ai`

[11]`https://automl.github.io/auto-sklearn/master/`

These approaches might help improve model performance but are computationally expensive and therefore were not suitable for this research with the limited resources and time.

6 Conclusion

In conclusion, this study compared different machine learning algorithms to predict life outcomes for families in the FFCWS dataset. The CatBoost algorithm outperformed other approaches for most target variables, demonstrating its effectiveness in handling categorical variables. The study then focused on eviction cases and identified variables related to financial instability as key predictors of eviction. While the findings offer valuable insights into the social context of eviction, the study also acknowledges the limitations of the survey and the challenges of predicting eviction due to its contextual nature. Future research can incorporate information on the house owner and the neighbourhood to attain a more nuanced understanding of eviction, as well as dimensionality reduction techniques to address collinearity in the dataset or multilevel models and automated machine learning to improve model performance.

References

Ahearn, C. E., & Brand, J. E. (2019). Predicting Layoff among Fragile Families. *Socius : sociological research for a dynamic world*, *5*, 10.1177/2378023118809757. Retrieved 2023-04-18, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8455106/ doi: 10.1177/2378023118809757

Basu, I., & Maji, S. (2022). Multicollinearity Correction and Combined Feature Effect in Shapley Values. In (Vol. 13151, pp. 79–90). Retrieved 2023-04-18, from http://arxiv.org/abs/2011.01661 (arXiv:2011.01661 [cs, stat]) doi: 10.1007/978-3-030-97546-3_7

Campbell, I., Parkinson, S., & Wood, G. (2013). The housing security consequences of underemployment.
(Publisher: Australian Housing and Urban Research Institute)

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321–357.

Clarke, A., Hamilton, C., Jones, M., & Muir, K. (2017). Poverty, evictions and forced moves. *York: Joseph Rowntree Foundation*.

Collinson, R., Humphries, J. E., Mader, N. S., Reed, D. K., Tannenbaum, D. I., & van Dijk, W. (2022, August). *Eviction and Poverty in American Cities* [Working Paper]. National Bureau of Economic Research. Retrieved 2023-04-18, from https://www.nber.org/papers/w30382 doi: 10.3386/w30382

Compton, R. (2019). A data-driven approach to the fragile families challenge: Prediction through principal-components analysis and random forests. *Socius*, *5*, 2378023118818720. (Publisher: SAGE Publications Sage CA: Los Angeles, CA)

Desmond, M., & Gershenson, C. (2017, February). Who gets evicted? Assessing individual, neighborhood, and network factors. *Social Science Research*, *62*, 362–377. Retrieved 2023-04-12, from https://www.sciencedirect.com/science/article/pii/S0049089X16300977 doi: 10.1016/j.ssresearch.2016.08.017

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, *92*(6), 1087. (Publisher: American Psychological Association)

Kumar, T. (2022). *Gendered differences in the factors influencing adolescent academic outcomes : An analysis of the Fragile Families and Child Wellbeing Study (FFCWS) using decision trees, a supervised machine learning method.* Retrieved 2023-04-18, from http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-190683

Rigobon, D. E., Jahani, E., Suhara, Y., AlGhoneim, K., Alghunaim, A., Pentland, A. & Almaatouq, A. (2019, January). Winning Models for Grade Point Average, Grit, and Layoff in the Fragile Families Challenge. *Socius*, *5*, 2378023118820418. Retrieved 2023-03-22, from https://doi.org/10.1177/2378023118820418 (Publisher: SAGE Publications) doi: 10.1177/2378023118820418

Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., ... McLanahan, S. (2020, April). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, *117*(15), 8398–8403. Retrieved 2023-03-23, from `https://www.pnas.org/doi/10.1073/pnas.1915006117` (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.1915006117

Salganik, M. J., Lundberg, I., Kindel, A. T., & McLanahan, S. (2019, January). Introduction to the Special Collection on the Fragile Families Challenge. *Socius*, *5*, 2378023119871580. Retrieved 2023-01-27, from `https://doi.org/10.1177/2378023119871580` (Publisher: SAGE Publications) doi: 10.1177/2378023119871580

Shwartz-Ziv, R., & Armon, A. (2022, May). Tabular data: Deep learning is not all you need. *Information Fusion*, *81*, 84–90. Retrieved 2023-04-18, from `https://www.sciencedirect.com/science/article/pii/S1566253521002360` doi: 10.1016/j.inffus.2021.11.011

Wallace, B. C., & Dahabreh, I. J. (2012, December). Class Probability Estimates are Unreliable for Imbalanced Data (and How to Fix Them). In *2012 IEEE 12th International Conference on Data Mining* (pp. 695–704). (ISSN: 2374-8486) doi: 10.1109/ICDM.2012.115

Wildeman, C. (2010). Paternal incarceration and children's physically aggressive behaviors: Evidence from the Fragile Families and Child Wellbeing Study. *Social Forces*, *89*(1), 285–309. (Publisher: The University of North Carolina Press)