

Heartbreaker: modeling heart disease mortality rates

Kevin Wu, Luca Schroeder
{wukevin,lucschr}@stanford.edu

ABSTRACT

This paper investigates the factors driving local variation in heart disease mortality rates through the lens of machine learning classification models. We constructed a dataset spanning variables from food availability to demographics, from economic factors to healthcare statistics, and applied seven different classifiers to this data. Of these, we found that boosted trees yielded the best balance of precision and recall. By applying state-of-the-art feature attribution methods to this model, we then suggest promising directions for researchers and policymakers to explore in reducing heart disease risk. Code and data for our project is publicly available at <https://github.com/wukevin/heartbreaker>.

1 INTRODUCTION

Heart disease is the leading cause of death in the United States [1], but the incidence of heart disease varies dramatically across the country and even among neighboring counties [2]. For our application project, we used machine learning techniques to investigate these divergent local health outcomes. In particular, we pose the following question: Can cardiovascular disease mortality rates for a county be predicted from local statistics like median income, cost of healthcare, and access/affordability of fresh produce and fitness centers? If so, what are the most important factors in making such a prediction? Our results may be of significant interest to policymakers and researchers, both in suggesting the highest-impact factors to target to reduce heart disease and in identifying counties that may be at risk of elevated heart disease mortality rates in the future such that they can receive targeted preventative care funding today.

2 RELATED WORK

There exists a large body of research aiming to characterize heart disease’s association with various factors such as socioeconomic status [3, 4], dietary patterns [5, 6], dietary patterns in association with socioeconomic status [7], exercise patterns [8], and demographics [9], among many others. Although they provide key insights, many of these studies focus on understanding individual factors’ impact on heart disease, instead of zooming out and looking holistically at how many such factors can interplay to determine heart disease risk in regional populations. While a few studies do apply machine learning to population heart disease trends, such as the Framingham Heart Study [10], they rarely go

beyond logistic regression - a powerful but relatively simple model that may miss complex interactions between features. More complex machine learning methods such as neural nets have been applied to studying heart disease, but this work has focused on improving patient diagnosis [11–13] instead of population-scale risk prediction. We hope to fill this gap by taking local statistics summarizing factors reasonably well known to be associated with heart disease and building a state-of-the-art machine learning classifier that ingests this multivariate data and predicts heart disease mortality rate.

3 DATA

3.1 Data sources

We combined four distinct datasets for our analysis. The first is a dataset on county-level heart disease mortality rates among adults aged 35 years and older, provided by the Center for Disease Control and Prevention (CDC) [14]. These mortality rates are measured as age-adjusted 3-year average rates spanning the years 2013 through 2015, inclusive. While the CDC provides measurements stratified by race and/or gender for select counties, we ignored these in favor of overall per-county rates for simplicity and data completeness.

Our second dataset is the Food Environment Atlas produced by the United States Department of Agriculture [15]. This dataset contains over 275 county and state level variables on food environment factors such as access to farmer’s markets, prevalence of fast food restaurants, and number of grocery stores, socioeconomic variables such as income levels and poverty rates, general health indicators such as diabetes and obesity rates, and demographic data such as race and age. We excluded all variables in this dataset collected after 2015, the heart disease mortality data’s end year (to avoid including future knowledge in our model), retaining 191 features.

As comprehensive as the Food Environment Atlas is, it lacks a key class of variables related to heart disease risk: healthcare quality and access. To address this, we included two additional datasets containing county-level metrics: the Geographic Variation Public Use File (GVPUF) for all beneficiaries (regardless of age) as provided by the Centers for Medicare and Medicaid Services [16], and economic characteristics data provided by the United States Census Bureau [17]. We chose to include only a subset of features from these sources, as much of the data is presented at a granularity unnecessary for our purposes (e.g. breaking down healthcare

coverage rates by occupation). From the GVPUF table, we selected two features, "Average HCC Score" and "Standardized Risk-Adjusted Per Capita Costs." The HCC (Health Condition Category) score measures overall health risk with higher values indicating poorer health/higher risk, and the second feature expresses the per-capita cost of healthcare. From the US Census Bureau data, we chose 3 variables: percentage of people with healthcare coverage, average household income, and average per-capita income. (While the Food Environment Atlas contains summary socioeconomic data, it does not contain detailed values such as average income.) Both sources contain data collected on an annual basis; we selected data collected in 2014 from both.

We engineered one additional feature, the per-capita cost of healthcare normalized by per-capita income. This feature attempts to capture the intuition that the raw dollar cost of healthcare may not be as informative as is the proportion of one's income represented by that cost. For example, a \$10,000 visit to a cardiac specialist is likely much more prohibitive in a region where the average per-capita income is only \$30,000 than in a region where the per-capita income is \$150,000.

The above datasets were inner-joined by their county identifiers, keeping only counties present in all datasets. Overall, our dataset contains 197 features for 2997 unique counties, all with published values for heart disease mortality rates.

3.2 Data pre-processing & cleaning

We define "extreme" values within a feature to be any value more than 3 times the inter-quartile range (IQR) away from the median. We replaced all such extreme values with the median plus or minus 3 times the IQR, based on whether the original value was larger or smaller than the median, respectively. Missing values were imputed by substituting in the mean value of that feature. When using models that perform better when features are on similar scales (e.g. logistic regression, support vector machine), we standardized our input features to be zero-mean with unit variance. Data processing was predominantly done using the pandas [18] and numpy [19] libraries.

4 METHODS

Our specific goal is to build a binary classifier that can discern counties of "high" risk and "low" risk, defined as being within the top 25th percentile or bottom 75th percentile of heart disease mortality rates, respectively. This 25th percentile cutoff corresponds to a heart disease mortality rate of 409.1 deaths per 100,000 population.

As a consequence of how we defined low and high heart disease mortality rates, we were faced with a class imbalance problem. To address this, we used the 'balanced' class weighting strategy utilized by the popular library sklearn [20, 21]

when training all of the following classifiers. Explicitly, the class weights are given by Equation 1, where w_c denotes the weight for class c , N denotes the total number of samples, $|C|$ denotes the total number of classes, and n_c denotes the number of samples in class c .

$$w_c = \frac{N}{|C| \times n_c} = \frac{\sum_{c' \in C} n_{c'}}{|C| \times n_c} \quad (1)$$

4.1 Models

Unless otherwise noted, we used model implementations provided by version 0.20.1 of the sklearn library [21].

4.1.1 Logistic regression. We used logistic regression to establish a performance baseline for classification models applied to our dataset. All input features were standardized.

4.1.2 Support vector classification (SVC). Support vector machines use the "kernel trick" to perform classification in high-dimension feature spaces. We tried three kernels: the linear kernel $\phi(x, z) = \langle x, z \rangle$, the radial basis function ('rbf') kernel $\phi(x, z) = \exp(-\gamma \|x - z\|^2)$, and the sigmoid kernel $\phi(x, z) = \tanh(\gamma \langle x, z \rangle + r)$. All input features were standardized.

4.1.3 Decision tree. As one of our project's goals was interpretability, it was only natural to include decision trees; since the split at each node of a decision tree is of the form $X_j > k$ (for feature j and value k), results are easy to interpret and visualize. To guard against overfit, we required that all leaf nodes contain a minimum percentage of all samples (see hyperparameter tuning). Gini loss $L(R_m) = \sum_k p_{mk}(1 - p_{mk})$ and cross-entropy loss $L(R_m) = -\sum_k p_{mk} \log p_{mk}$ were tried as measures of impurity.

4.1.4 Random forest. Random forests use bootstrapped samples to train a set of classification trees. Predictions are made by averaging the results obtained across these trees. Such ensemble approaches can result in a large reduction in variance with minimal increase in bias.

4.1.5 Boosted tree. Boosting combines weak learners to create a strong learner, and is commonly applied to decision trees. We used a popular implementation of this strategy, XGBoost [22] (version 0.80), which specifically uses gradient boosting. Intuitively, gradient boosting looks at residual errors made by a classifier and greedily fits subsequent models to minimize them, thus creating an additive model. More formally, the objective function minimized at iteration t is given by Equation 2, where $f_t \in \mathcal{F}$, \mathcal{F} denotes the space of all regression trees, l is a convex loss function, and Ω is a regularization function penalizing complexity of f_t .

Model	Results on Training Set				Results on Test Set			
	Accuracy	Precision	Recall	F_1 score	Accuracy	Precision	Recall	F_1 score
Logistic Regression	0.8228	0.6091	0.8489	0.7078	0.8367	0.5955	0.8030	0.6839
SVC ('linear' kernel)	0.8235	0.6082	0.8642	0.7126	0.8300	0.5806	0.8182	0.6792
SVC ('rbf' kernel)	0.8339	0.6323	0.8313	0.7166	0.8167	0.5604	0.7727	0.6497
SVC ('sigmoid' kernel)	0.7953	0.5643	0.8533	0.6780	0.8000	0.5283	0.8485	0.6512
Decision Tree	0.7968	0.5709	0.7959	0.6637	0.7800	0.5000	0.7273	0.5926
Random Forest	0.8354	0.6381	0.8124	0.7137	0.8033	0.5368	0.7727	0.6335
Boosted Tree	0.8554	0.6861	0.7924	0.7354	0.8467	0.6250	0.7576	0.6849

Table 1: Performance for multiple classification models with optimized hyperparameters

$$\mathcal{L}_t = \sum_{i=1}^n l(y^{(i)}, \hat{y}_{(t-1)}^{(i)} + f_t(x^{(i)})) + \Omega(f_t) \quad (2)$$

5 RESULTS

5.1 Model evaluation & hyperparameter tuning

We divided our full data set into two partitions, a training set with 90% of our data (2697 observations) and a test set with the remaining 10% (300 observations).

We tuned hyperparameters for the classification models using k-fold cross validation within our training set with $k = 10$. Our primary metric for evaluating performance was the F_1 score, given by Equation 3. The F_1 score provides a balanced measure of precision and recall, and is often a better suited metric than accuracy in cases of class size imbalance.

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

Listed below are the hyperparameters we optimized for each model, all of which were evaluated via grid search enumerating all possible combinations. Final optimized hyperparameters are **boxed and in bold**.

- Logistic regression: **L1** vs. L2 regularization; regularization constant $C \in \{0.001, 0.01, \mathbf{0.1}, 1, 10, 100, 1000\}$
- SVC ('linear' kernel): Regularization constants $C \in \{0.001, \mathbf{0.01}, 0.1, 1, 10, 100, 1000\}$
- SVC ('rbf' kernel): Regularization constants $C \in \{0.001, 0.01, 0.1, \mathbf{1}, 10, 100, 1000\}$
- SVC ('sigmoid' kernel): Regularization constants $C \in \{0.001, 0.01, \mathbf{0.1}, 1, 10, 100, 1000\}$
- Decision tree: Gini vs. **cross-entropy loss**; maximum depths $\in \{25, 50, 100, \mathbf{\infty}\}$; minimum samples per leaf $\in \{0.1\%, \mathbf{1\%}, 10\%\}$ of total samples; max features for finding splits given by square root or \log_2 of total features, or $\mathbf{\infty}$

- Random forest: Gini vs. **cross-entropy loss**; number of trees in forest $\in \{10, \mathbf{100}, 1000\}$; maximum depth $\in \{25, 50, 100, \mathbf{\infty}\}$; min samples per leaf $\in \{0.1\%, \mathbf{1\%}, 10\%\}$, of total samples; max features for finding splits given by **square root** or \log_2 of total features, or ∞
- Boosted tree: max tree depth for base learners $\in \{3, 4, \mathbf{6}, 8\}$; num. boosted trees to fit $n \in \{\mathbf{150}, 200, 250, 300, 350, 400, 450\}$; L1 regularization term $\alpha \in \{0.001, 0.01, 0.1, 1, \mathbf{10}, 100\}$; learning rates $\in \{0.01, \mathbf{0.1}\}$; boosters **gbtree**, gblinear, or dart; min sum of instance weights in a child $\in \{\mathbf{1}, 2, 3, 4\}$

5.2 Model performance

Boosted Tree	Predicted Positive	Predicted Negative
True Positive	50	16
True Negative	30	204

Table 2: Confusion matrix on testing data

Table 1 shows model performance after aforementioned hyperparameter tuning. Training set results are averaged over the 10 cross validation folds. Test set performance is measured by training a single model based on all 2697 training examples, and evaluating that model on test data. The best value in each column is in **bold**.

Comparing F_1 scores on training and test data, all of our models perform noticeably worse on the held-out testing data than on the training data; this seems to be particularly true with the SVC with 'rbf' kernel, decision tree model, and random forest classifiers. This suggests that our models overfit the training set to some degree, which is surprising as regularization strength (and sometimes regularization methodology) was among the hyperparameters tuned for every model during cross validation. Logistic regression appears to have suffered from this the least, as it has the smallest performance delta between testing and training. This is

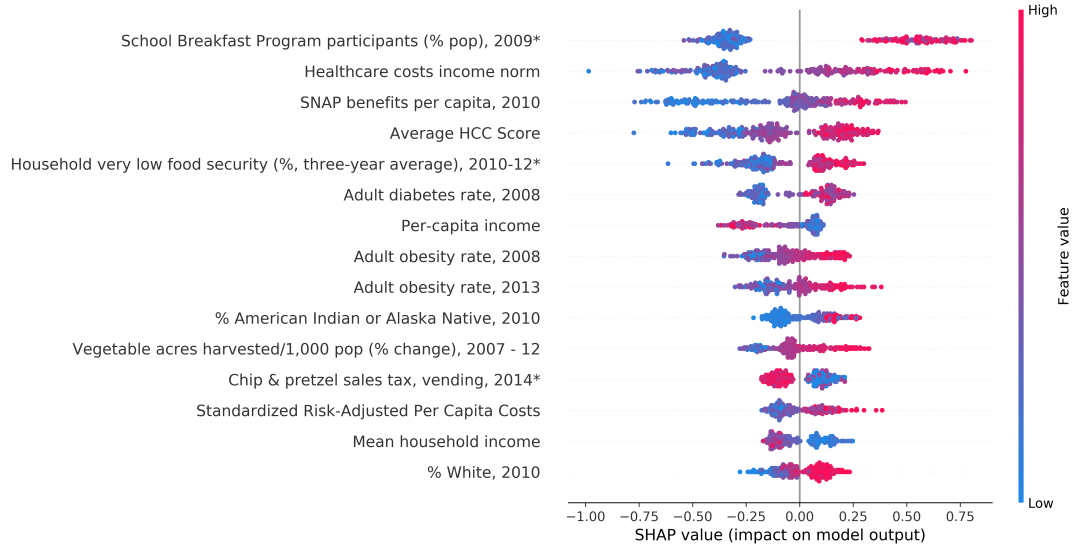


Figure 1: Fifteen most explanatory features as ranked by SHAP values

likely due to the logistic regression model having relatively high bias and thus lower variance.

Among the tree-based methodologies, we see that the decision tree has the worst performance; this is expected, as both the random forest and boosted tree use the decision tree as a foundation to build and improve upon. There is also a notably large gap in performance (especially on test data) between the random forest model and the boosted tree model, with the latter performing better. This may be due to the design of each algorithm; specifically, random forest aims to reduce variance by simply fitting many randomized tree models, while a boosted tree model attempts to systematically and iteratively remove errors, which could help it better approximate complex patterns in the data.

Overall, we see that the boosted tree model performs the best on both training and test data, with its confusion matrix on test data shown in Table 2. However, it is notable that logistic regression, a much simpler model less capable of expressing complex, nonlinear interactions between features, has only marginally worse performance on test data. This suggests that the underlying trends driving heart disease mortality rates may be predominantly linear, a possibility reinforced by the strong performance of the support vector classifier with linear kernel.

5.3 Visualization & interpretation

As mentioned previously, one of our goals for the project was to generate clear insights, not just in terms of which counties are predicted to have high risk of heart disease mortality, but *why* they are predicted as such. To this end, we applied SHAP (SHapley Additive exPlanations) [23, 24],

a recently developed framework for understanding model predictions and feature importance, to understanding our top-performing model, the boosted tree. A SHAP value for a particular feature for a particular observation is the expectation, taken over all possible feature orderings, of the change in the model prediction from adding that feature. Figure 1 visualizes the SHAP values for the 15 most explanatory features over every sample in the test dataset for our boosted tree model. The color of a dot (each representing a sample) indicates the feature’s value at that particular sample (red indicating a high value and blue a low value), and its horizontal position indicates its contribution toward a negative or positive prediction. The visualization for the ‘chip & pretzel sales tax’ feature, for instance, indicates that increased sales taxes on these foods on average push the model towards a low-risk prediction, while lower sales tax values tend to push the model towards a high-risk prediction; in both directions the impact seems modest.

Looking at the top 5 most influential features in Figure 1, we see a heavy emphasis on access to quality food and quality healthcare. The school breakfast program (SBP) is designed to “provide nutritious breakfasts to children at reasonable prices” [25] and increased participation could indicate that parents face challenges in providing healthy breakfasts for their children via other means. This observation, coupled with how elevated SNAP (Supplemental Nutrition Assistance Program) participation and decreased food security drive our model towards positive predictions, leads to the emphasis we place on access to quality food in our interpretation. The second most informative feature in this plot is healthcare costs normalized by income levels (the feature we engineered).

Heartbreaker: modeling heart disease mortality rates

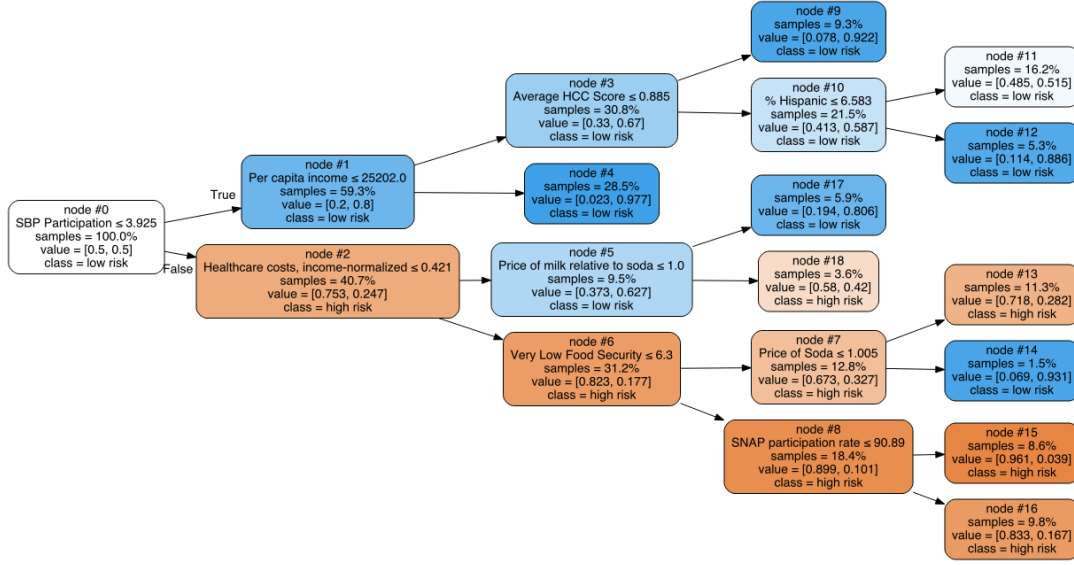


Figure 2: Visualization of decision tree model (truncated for brevity)

Naturally, higher values for this metric could dissuade individuals from going to the doctor, reducing the delivery and efficacy of critical preventive care. The fact that high HCC values push the model towards “high risk” predictions further reinforces our suspicion that proper preventive care may not be readily accessible in these areas.

To verify our interpretation of the SHAP values above, we turned to decision trees; despite not performing as well as boosted trees, decision trees have clear, easily understood structure as shown in Figure 2 (drawn using graphviz [26]), and should ideally reinforce the ideas we’ve formulated above. Indeed, we see the majority of those highly influential variables echoed here in the first few decision nodes. For instance, the path #0-#2-#6-#8 suggests that counties with high SBP participation, high income-normalized healthcare costs, and very low food security are more likely to be high-risk, as 89.9% of samples at the penultimate node #8 are high-risk, compared to 25% in the overall dataset. (The color of a node indicates the number of high-risk (orange) and low-risk (blue) counties at that node.) While correlation doesn’t imply causation, the above results corroborate that access to quality food and quality healthcare may play critical roles in managing heart disease mortality rates, and certainly should be explored further in ongoing efforts to improve public health.

We conclude our interpretation of these models with a more optimistic observation; both the SHAP values and decision tree nodes cite features like chip and pretzel sales taxes (Figure 1) or price of soda (Figure 2, node #7) as factors inversely related to heart disease risk. Without making statements regarding causality, this suggests that such seemingly

small measures may actually nudge a community towards decreased heart disease mortality, and certainly highlights such measures as important avenues for further research.

6 CONCLUSION & FUTURE WORK

For this project, we amassed features across four different datasets to build a classifier capable of predicting whether a county falls under high or low heart disease mortality rates. Through building this classifier, we have observed how different models’ strengths and weaknesses come into play when answering a real-world research question. By applying methods for elucidating these models, we have highlighted several important areas where continued research has a high potential for great impact in understanding and addressing heart disease mortality rates.

Future directions include several possible avenues of exploration. During the later stages of our project, we performed some cursory exploration into neural net classification using pytorch [27]; while our initial results haven’t been very promising, it would be interesting to continue tweaking these models. As for our dataset, we could add additional features such as the ratio of white to blue collar workers in an area and prevalence of risky health behaviors like smoking and alcoholism. Another avenue of exploration would be the inclusion of a time component in our model. Heart disease doesn’t develop overnight, so changes in local factors such as healthcare costs likely take years to be reflected in heart disease mortality rates. Being able to project how these effects manifest over time would be a very interesting, useful addition.

7 CONTRIBUTIONS

We have contributed equally to the project. The following lists the specific contributions each team member had to the project:

Kevin:

- Data collection, merging, and cleaning
- Boosted tree algorithm
- SHAP visualization of boosted tree
- Exploratory work with neural nets

Luca:

- Logistic regression
- Support vector classification
- Decision tree and corresponding visualization
- Random forests

We contributed equally to interpreting results, writing this report, and to generating figures/tables.

REFERENCES

- [1] Centers for Disease Control and Prevention, “FastStats - Leading Causes of Death.” <https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>, 2017. [Online; accessed 2018-10-08].
- [2] Centers for Disease Control and Protection, “Interactive Atlas of Heart Disease and Stroke.” <https://nccd.cdc.gov/dhdspatlas/>. [Online; accessed 2018-10-01].
- [3] M. Woodward, S. A. Peters, G. D. Batty, H. Ueshima, J. Woo, G. G. Giles, F. Barzi, S. C. Ho, R. R. Huxley, H. Arima, X. Fang, A. Dobson, T. H. Lam, and P. Vathesatogkit, “Socioeconomic status in relation to cardiovascular disease and causespecific mortality: A comparison of Asian and Australasian populations in a pooled analysis,” *BMJ Open*, vol. 5, p. e006408, mar 2015.
- [4] M. M. Glymour, C. R. Clark, and K. K. Patton, “Socioeconomic Determinants of Cardiovascular Disease: Recent Findings and Future Directions,” *Current Epidemiology Reports*, vol. 1, pp. 89–97, jun 2014.
- [5] E. Mertens, O. Markey, J. M. Geleijnse, D. I. Givens, and J. A. Lovegrove, “Dietary patterns in relation to cardiovascular disease incidence and risk markers in a middle-aged british male population: Data from the caerphilly prospective study,” *Nutrients*, vol. 9, jan 2017.
- [6] M. A. Muga, P. O. Owili, C. Y. Hsu, H. H. Rau, and J. C.-J. Chao, “Association between Dietary Patterns and Cardiovascular Risk Factors among Middle-Aged and Elderly Adults in Taiwan: A Population-Based Study from 2003 to 2012,” *PloS one*, vol. 11, no. 7, p. e0157745, 2016.
- [7] T. Psaltopoulou, G. Hatzis, N. Papageorgiou, E. Androulakis, A. Briasoulis, and D. Tousoulis, “Socioeconomic status and risk factors for cardiovascular disease: Impact of dietary mediators,” jan 2017.
- [8] J. Stewart, G. Manmathan, and P. Wilkinson, “Primary prevention of cardiovascular disease: A review of contemporary guidance and literature,” *JRSM Cardiovascular Disease*, vol. 6, p. 204800401668721, 2017.
- [9] J. R. Betancourt, A. R. Green, J. E. Carrillo, and I. Owusu Ananeh-Firempong, “Defining cultural competence: A practical framework for addressing racial/ethnic disparities in health and health care,” *Public Health Reports*, vol. 118, no. 4, pp. 293–302, 2003. PMID: 12815076.
- [10] E. J. Benjamin, D. Levy, S. M. Vaziri, R. B. D’Agostino, A. J. Belanger, and P. A. Wolf, “Independent Risk Factors for Atrial Fibrillation in a Population-Based Cohort,” *JAMA*, vol. 271, p. 840, mar 1994.
- [11] R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” *Expert Systems with Applications*, vol. 36, pp. 7675–7680, may 2009.
- [12] S. Palaniappan and R. Awang, “Intelligent heart disease prediction system using data mining techniques,” in *2008 IEEE/ACS International Conference on Computer Systems and Applications*, pp. 108–115, IEEE, mar 2008.
- [13] M. Kukar, I. Kononenko, C. Grošelj, K. Kralj, and J. Feticich, “Analysing and improving the diagnosis of ischaemic heart disease with machine learning,” *Artificial Intelligence in Medicine*, vol. 16, pp. 25–50, may 1999.
- [14] Centers for Disease Control and Prevention, “Heart Disease Mortality Data Among US Adults (35+) by State/Territory and County - Data.gov.” <https://catalog.data.gov/dataset/heart-disease-mortality-data-among-us-adults-35-by-state-territory-and-county-5fb7> 2018. [Online; accessed 2018-10-11].
- [15] United States Department of Agriculture Economic Research Service, “USDA ERS - Data Access and Documentation Downloads.” <https://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads>, 2018. [Online; accessed 2018-10-11].
- [16] Centers for Medicare & Medicaid Services, “Public Use File.” https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html, 2018. [Online; accessed 2018-12-01].
- [17] United States Census Bureau, “American FactFinder.” <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>. [Online; accessed 2018-12-02].
- [18] W. McKinney, “Data structures for statistical computing in python,” in *Proceedings of the 9th Python in Science Conference* (S. van der Walt and J. Millman, eds.), pp. 51 – 56, 2010.
- [19] T. Oliphant, “NumPy: A guide to NumPy.” USA: Trelgol Publishing, 2006–. [Online; accessed <today>].
- [20] G. King and L. Zeng, “Logistic Regression in Rare Events Data,” *Political Analysis*, no. 9, pp. 137–163, 2001.
- [21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [22] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” mar 2016.
- [23] S. M. Lundberg, G. G. Erion, and S.-I. Lee, “Consistent individualized feature attribution for tree ensembles,” *arXiv preprint arXiv:1802.03888*, 2018.
- [24] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” pp. 4765–4774, 2017.
- [25] California Department of Education, “School Breakfast Program - School Nutrition.” <https://www.cde.ca.gov/ls/nu/sn/sbp.asp>, 2018. [Online; accessed 2018-12-10].
- [26] E. R. Gansner and S. C. North, “An open graph visualization system and its applications to software engineering,” *SOFTWARE - PRACTICE AND EXPERIENCE*, vol. 30, no. 11, pp. 1203–1233, 2000.
- [27] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in pytorch,” in *NIPS-W*, 2017.