Challenges and Opportunities in Using Data Science for Homelessness Service Provision

Charalampos Chelmis*
cchelmis@albany.edu
Department of Computer Science
University at Albany, SUNY
Albany, New York, USA

Wenting Qi
wqi@albany.edu
Department of Computer Science
University at Albany, SUNY
Albany, New York, USA

Wonhyung Lee
whlee@albany.edu
School of Social Welfare
University at Albany, SUNY
Albany, New York, USA

ABSTRACT

Homelessness service provision, a task of great societal relevance, requires solutions to several urgent problems facing our humanity. Data science, that has recently emerged as a potential catalyst in addressing long standing problems related to human services, offers immense potential. However, homelessness service provision presents unignorable challenges (e.g., assessment methods and data bias) that are are seldom found in other domains, requiring cross-discipline collaborations and cross-pollination of ideas. This work summarizes the challenges offered by homelessness service provision tasks, as well as the problems and the opportunities that exist for advancing both data science and human services. We begin by highlighting typical goals of homelessness service provision, and subsequently describe homelessness service data along with their properties, that make it challenging to use traditional data science methods. Along the way, we discuss some of the existing efforts and promising directions for data science, and conclude by discussing the importance of a deep collaboration between data science and domain experts for synergistic advancements in both disciplines.

CCS CONCEPTS

• **General and reference** → Surveys and overviews; • **Information systems** → *Decision support systems*; • **Applied computing** → *Enterprise applications*; • **Computing methodologies** → *Machine learning*.

KEYWORDS

human services; non-profit organizations; socially important data science

ACM Reference Format:
Charalampos Chelmis, Wenting Qi, and Wonhyung Lee. 2021. Challenges and Opportunities in Using Data Science for Homelessness Service Provision. In *Companion Proceedings of the Web Conference 2021 (WWW '21 Companion), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3442442.3453454

*Corresponding author.

WWW '21 Companion, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8313-4/21/04.
https://doi.org/10.1145/3442442.3453454

1 INTRODUCTION

Homelessness, which is described by the U.S. federal government as staying in a non-habitable residence for permanent living [1], presents a long-standing social, public health and policy problem across the world. In the US alone, 568, 000 Americans experienced homelessness on a single night in 2019 [24], whereas the number of homeless families in England increased from 50, 000 to 78, 000 in 2017 [36]. Multiple factors lead to people unable to afford housing including but not limited to increasing housing costs, limited affordable housing options and job opportunities, health and disabilities problems, family breakup, and unpredictable public health and social emergencies [6].

The application of data science to human services has been spurred by the scarcity of housing resources, and the need to prioritize waitlisted homeless, with the ultimate goal of stabilizing households and reducing future demands for assistance [20, 29]. Given the variety of disciplines that have recently become engaged in human services research, there is a great opportunity for cross-discipline collaborations and cross-pollination of ideas with the potential to advance both data science as well as social and behavioral science. At the same time, homelessness service provision presents several challenges that are strikingly different from those encountered in other domains, requiring novel problem formulations and data science methodologies. For example, human services is an overwhelmingly data-poor field, in which whatever little data is being collected is heavily restricted by privacy issues.

The purpose of this work is to summarize both the challenges and opportunities offered by the human services domain, with a particular focus on homelessness service provision. The remainder of this paper is organized as follows. Section 2 outlines important aspects where data science has the potential to yield major advances. Section 3 provides an overview of homelessness service data. Section 4 describes the challenges for data science arising from both the complex and dynamical nature of human services domain as well as the data collection itself. Section 5 discusses opportunities for data science methods, while Section 6 provides concluding remarks.

2 THEMES OF DATA SCIENCE RESEARCH FOR HOMELESSNESS SERVICE PROVISION

The ultimate goal of homelessness service provision is to decrease the total number of people being homelessness or mute the growth rate of homelessness [7]. To achieve this goal, existing data science solutions have thus far focused on:

- Matching service programs to need [55].

Table 1: Abbreviations and their corresponding description.

Abbreviation	Description
CES	Coordinated Entry System
CoC	Continuum of Care
DS	Day Shelter
ES	Emergency shelter
HMIS	Homeless Management Information System
HP	Homelessness Prevention
HUD	U.S. Department of Housing and Urban Development
PSH	Permanent supportive housing
RRH	Rapid rehousing
SNAP	Food stamp program
SSI	Supplemental security income
SSDI	Social security disability insurance
SO	Street Outreach
TH	Transitional housing
WIC	Supplemental nutrition program for women, infants and children

- Assessing the impact of service matching on the reduction of reentries [8].
- Prioritizing service allocation based on risk assessment or predicted outcomes [37].

Optimizing allocation based on predicted outcomes in particular, promises substantial gains in homelessness service delivery. However, the application of data-driven approaches in a domain with considerable individual and social costs requires careful consideration, as detailed in this work.

3 DATA

Most communities in the U.S rely on the so called Coordinated Entry System (CES), according to which homeless people first sign up for housing support (e.g., emergency shelter), and are subsequently accessed for eligibility and vulnerability, and are subsequently prioritized for housing based on the assessments. This information is entered into the Homeless Management Information System (HMIS), a local information technology system, founded by the U.S. Department of Housing and Urban Development (HUD). HMIS serves as a network of local agencies that collaborate to provide a variety of service programs, including but not limited to emergency shelter (ES), day shelter (DS), homelessness prevention (HP) and rapid rehousing (RRH), transitional housing (TH), and permanent supportive housing (PSH) [32]. According to federal mandates, HMIS is operated by a lead organization for each Continuum of Care (CoC), and collects individual-level data including personally identifying information, socioeconomic backgrounds, healthy condition and educational history [37]. In addition, fields specifying the type of exit from homelessness (e.g., interim housing, hotel or motel) and whether an individual has reentered the system multiple times are also recorded in the system.

For illustration purposes, we rely on a dataset obtained by the CARES of NY Inc., a non-profit organization in the state of New York, managed by the HMIS. The dataset comprises 92, 586 records of housing services given to a total of 38, 800 individuals seeking federally funded homeless assistance between 2005 and 2019 in the Capital Region of New York. Each record in the dataset has both time-variant (e.g., monthly income, age) as well as time-invariant (e.g., race) features. Specifically, each record provides information about household relations (e.g., household head, spouse and child), health (e.g., dental status and mental state) and disability (e.g., physical or mental disability), income (e.g., source of income and amount earned), enrollment (e.g., length of stay, living condition, and housing information), service (i.e., services received), educational history (e.g., last grade completed), and working situation.

Table 2: Descriptive statistics of sample features in HMIS data.

Categorical Feature	Value	Percentage of Population
Sex	Female	40.49
	Male	59.61
	Transgender	0.31
Age	0-18	15.72
	19-30	31.33
	31-50	30.64
	51-65	19.40
	>66	2.88
Race	American Indian or Alaska Native	2.09
	Asian	0.86
	Black or African American	54.38
	Native Hawaiian or Other Pacific Islander	0.64
	White	37.27
	Race None	3.73
Reentry	Yes	19.82
	No	80.18

While substantial insight can be gained from HMIS data, homeless service agencies mainly collect data for reporting and record keeping purposes, rather than for systematized data–driven decision–making. Furthermore, any single agency may find it difficult, if not impossible, to track a homeless individual or family as they move from one shelter to another over the years. Organizations operating HMIS in a locale, have access to such comprehensive data, but there is no consistent technical expertise and/or resources to process and analyze such data.

4 CHALLENGES

In this section, we discuss several challenges in using data science for homelessness services provision.

4.1 Data Bias

Data bias is generally described as available data that is not representative of the population or phenomenon of the target study [3]. Bias in a dataset (e.g., imbalanced data distribution based on feature race shown in Table 2) that involves real humans [18] that

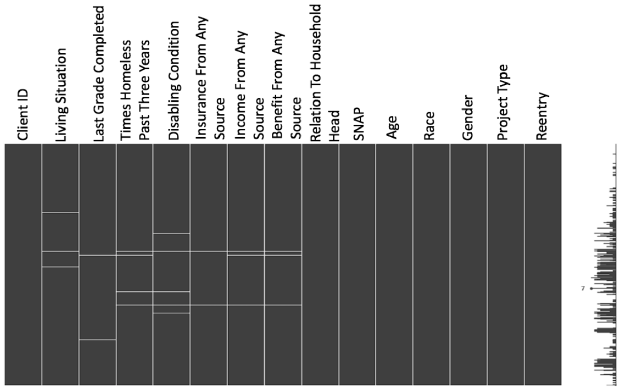


Figure 1: Visualization of missing information.

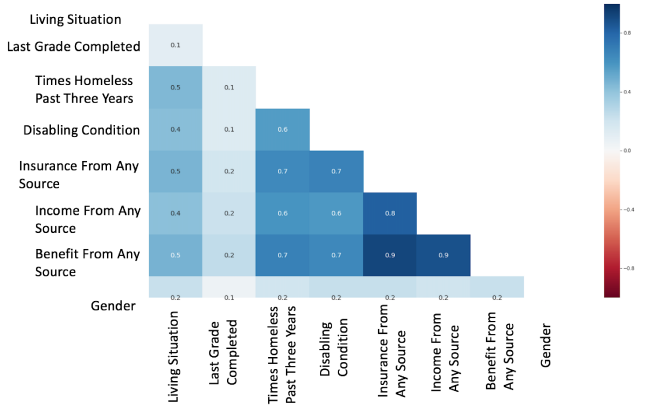


Figure 2: Correlation analysis shows that if a client refuses to provide information for one feature, it is highly likely that the completeness of other, highly correlated features will also be compromised.

causes greater harm because a biased dataset may induce human-like bias (e.g., sex, religion and race) in the learning model and further influence the decision-making process. For instance, when using algorithmic tools in the criminal justice system to predict individuals who are likely to get involved in crimes [39], a biased output that leads to a discriminatory result can be detrimental to wrongfully accused (and perhaps prosecuted) individuals.

An estimated 568,000 Americans experienced homelessness on a single night in 2019 and 62.8% of them are sheltered [24]. With access to a small fraction of the general homeless population in the U.S., it is hard to generalize developed data science models across communities. For instance, our team has access to data collected from data of 38,800 homeless people that were assisted by a specific service provided in a city in upstate New York. Prior research (e.g., [27, 28]) was based on data from other locations. In addition, collected data may be imbalanced in terms of specific subpoluations, as indicated by features such as sex, age and race. Table 2, which presents statistics of different features in HMIS data, shows that 59.61% of clients are female versus 40.49% male, and 54.38% being Black or African American as compared with other races. Similarly, when examining the percentage of homeless individuals that receive homelessness services after their first exit from the system (i.e., *reentry*), discrepancies can be identified across datasets. For example, Table 2 shows that 19.82% people reenter into the homelessness services system in the Capital Region of New York as compared to 22.66% people requesting homelessness services sometime within 2 years of their exit in a different metropolitan community in another state [29].

Specifically, three major types of bias [38] have been identified in HMIS data as follows.

4.1.1 Response Bias. Response bias is commonly caused by inaccurate or untruthful answers from respondents, particularly when participants are asked to self-report on their background, and can be generally divided into social response bias and hostility bias [5]:

- **Social Response Bias:** Social response bias points to the phenomenon that homeless people who are affected by social desirability bias may over-report on good situations and under-report on bad situations [40]. For instance, health condition related information (e.g., chronic health condition and mental health problem) is collected from self-reports, clinic assessment and previous clinic records. Different collection ways share different confidence level, because health condition records provided by clinic assessment is more trustful than self-reports. This can be problematic when learning automated methods for homelessness service provision. For instance, in Bayesian classification (e.g., where the goal is to predict object-oriented software maintainability [54]) all feature values are treated equally in their contribution to the predicted class [30]. One possible method to address this challenge is to weight features based on their confidence scores [15].
- **Hostility Bias:** One key reason for clients refusing to provide certain responses is that they are asked for some information related to unpleasant memories or painful experiences (e.g., divorce, debt, death). This phenomena is known as hostility bias [14]. However, such information may be critical to the type of assistance that they truly need (e.g., drug-abuser may miss the detoxification service if he/she refuse to provide drug abuse history). In addition, information that homeless people omit may influence the completeness of other collected information. Figure 1 shows the information that homeless people omit in the HMIS dataset. In Figure 1, each row represents a client record and each column represents a feature extracted from the clients' response (e.g., age, race, times being homeless past three years). Black denotes available data, whereas a horizontal white line represents that the corresponding client refused to provide information. Moreover, each white line denotes multiple clients that refused to provide response (e.g., clients 395 and 440 refused to provide information with respect to their "Living Situation" and "Disabling Condition" respectively).
Figure 1 shows that a row with missing values in the 'Times Homeless Past Three Years' column, has a high chance of missing values in the 'Disabling Condition' columns. It follows, that missing information for one feature may in turn

influence the completeness of information of other features. This intuition is in fact confirmed by Figure 2, a heatmap showing the correlation analysis between the values of different features. Positive correlation is proportional to the level of darkness in blue as indicated by the bar on the right side. The highest correlation coefficient in the figure is 0.9. This means that information that a client refuses to provide is directly linked to the completeness of other inputs.

4.1.2 Omitted Feature Bias. Omitted feature bias typically reflects that the dataset misses the critical features that influence the outcome. Consider the study of using machine learning models in terms of predicting individuals who are likely to reenter the homelessness services system after exiting. Omitted features in HMIS data including calls (i.e., call hotline for assistant) and wait (i.e., time period from requesting to enrolling) are crucial information to follow through and assess the needs of individuals after exiting the system. [29] illustrates that those two features obtain high scores in the feature selection process, and the higher score a feature obtains, the greater influence it has on the outcome of reentry. Missing such key features can lead to inaccurate models and corresponding predictions. In addition, individual features related to received social support such as relief funds and household relations are also crucial for reentry prediction. For instance, [10] suggests that the more social support a client receives from family members and friends, the less likely he or she is to experience repeated homelessness episodes. Unfortunately, such information is not being currently recorded in HMIS.

4.1.3 Environmental Factor Bias. Environmental factors such as great depression, public health, and social emergencies are one of the reasons of being homeless. Compared with the time period from July to November in 2007, the number of families entering New York City homeless shelters jumped by 40% [45] during the recession period in 2008. Due to the financial crisis caused by the world-spread COVID-19, nearly 19 – 23 million renters are continually facing the risk of eviction and being homeless by the end of 2020 [6]. Those people being homeless by force majeure environmental factors need different assistance compared with those being homeless for a long period or experiencing repeated episodes of homelessness [45]. Models ignoring environmental factors risk inaccurate or unreasonable homelessness service allocation for those that become homeless in extraordinarily circumstances (i.e., economic depression, public health and social emergencies).

4.2 Data Sparsity

As mentioned in Section 1, one of the goals of using data science for homelessness service provision is to match service programs to clients needs. Figure 3 shows nine types of homelessness service programs, each of which is further divided into subprograms as shown in Figure 4, corresponding to the specific characteristics of the population served (e.g., females or youth) and service area covered (e.g., zipcodes). Each subprogram is assigned a unique identifier (i.e., project ID). In order to make fair allocation for both services programs and subprograms, data distribution of each services programs/subprograms should match the service availability and client demand. However, Figure 3 and Figure 4 illustrate uneven

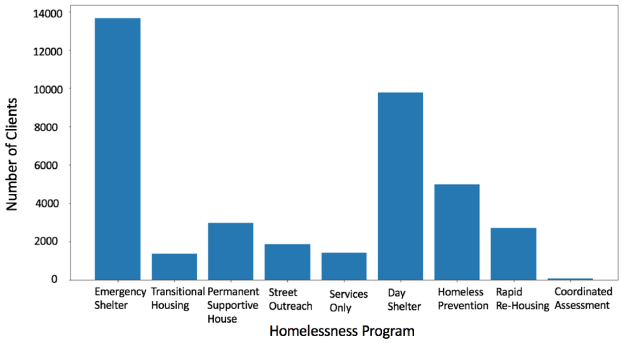


Figure 3: Distribution of service programs.

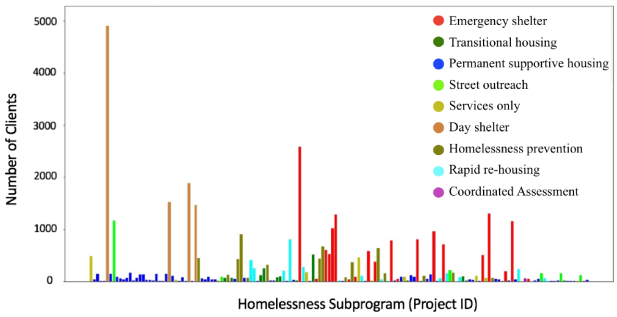
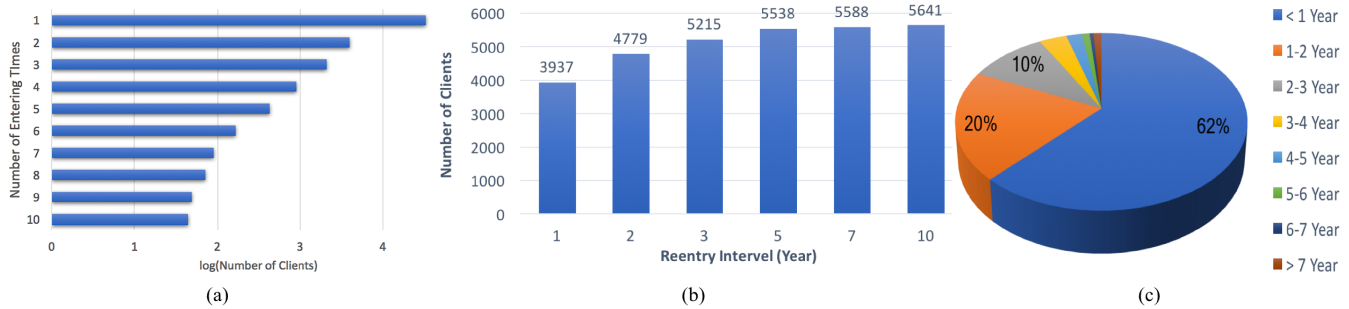


Figure 4: Distribution of service subprograms.

data distribution for both service programs and subprograms. If less data for specific programs (e.g., Coordinated assessment) is available, a machine learning model is prone to biases towards the more "frequent" programs (e.g., Emergency shelter, Day shelter).

Another goal of the homelessness service system is to reduce the number of individuals experiencing repeated episodes of homelessness. Identifying common attributes of reentered cases could add insights towards achieving this goal; this requires a learning model that is built based on sufficient data. However, Figure 5(a) shows that the number of multientry clients (i.e., those with more than one distinct period of residence in a homeless shelter) is limited. Specifically, 80.17% of the clients enter once compared to 19.83% clients who enter multiple times. Furthermore, with a different definition of reentry, the number of multientry individuals varies. For example, [29] defines reentry as "*requesting services within two years of exiting from the system*". Following this definition, the total number of multientry clients in our dataset is 4, 779, as shown in Figure 5(b). However, this number fluctuates depending on the definition of reentry, which we adjusted by considering varying lengths of time periods in which people reenter the homelessness services system. Nevertheless, it may be more advantageous to predict not only the probability of reentry, but also the time when a reentry may occur (i.e., how many years after one's exit one is expected to reenter the system). Figure 5 (c) shows that the distribution for the time interval of reentry is uneven, particularly when reentry is defined as requesting services within a time interval of more than four years.



(a) (b) (c)

Figure 5: Analysis of reentry data: (a) number of clients entering the system multiple times, (b) number of multientry clients given a time period threshold, and (c) percentage of reentry interval.

Besides, the level of aggregation can also cause sparsity. For example, when considering individuals record the total number is 38, 800. Each individual belongs to one household and has household one certain attribute from household head, spouse, and child. Compare individual to the number of records referring to households which aggregate multiple records (in our dataset, as many as 16) from multiple family members. In this case, the total number of data entries decreases to 24, 117.

4.3 Performance Measures

Evaluation metrics of data science methods used for homeless services provision can be classified into two main categories: better exits and fewer entries [9]. Better exits is quantified based on the ability of allocated services to place clients in permanent housing. Instead, the fewer entrie metric emphasizes on reducing the number of repeated episodes of homelessness, and is achieved by tracking relevant records of repeated enrollment, call and waiting time [29]. The Next Step Tool (NST) score, which is measured by multiple factors (e.g., history of housing, risks, socialization and daily functions), has been recently used by [11] to evaluate the quality of homelessness services received by youth. The above measurement methods aim to evaluate the quality of received services.

The challenge is to evaluate performance based on counterfactual allocation using observational data. Assume for example that a client was actually assigned to an emergency shelter, and the predicted allocation for the same client is a day shelter. In this context, the predicted allocation is inconsistent with the observation (which is considered to be ground truth for evaluation purposes) and should be treated as false negative from a pure data science perspective. From a practical perspective, a high number of false negatives could mean that individuals that need special assistance do not receive it. However, when predicted services deviate from the ground truth, it does not necessarily mean that worse assistance is delivered to clients. Hence the challenge becomes how to measure the counterfactual service quality and how to simulate the outcome of the clients who are assisted by different services.

5 OPPORTUNITIES

In this section, we discuss several opportunities of using data science for homelessness services provision using HMIS data.

5.1 Long Term Predictions

Using data science related methods can make long term predictions on trajectories of homeless people, which are classified into three categories including consistently sheltered, inconsistently sheltered (i.e., reentry) and long-term exit [52]. Consistently sheltered refers to homeless people will keep staying in the homelessness services system for a long period. Inconsistently sheltered is described as the transition from one service to the next within a certain time interval as shown in Figure 5. Long term exit refers to homeless people will exit the homelessness system and step to another place for living (e.g., hospital facility, prison, permanent housing, hotel or motel). Providing such forecasting trajectories for current and incoming homeless people can help shelters provide targeted assistance and maximize the services allocation with limited, shared social resources. By identifying those consistently sheltered users in advance, shelter providers can avoid assigning them to short assistance programs (e.g., day shelter) which may cause further transition that has potentially additional shifting cost.

From the data science perspective, the trajectory prediction problem can be treated as a time-series problem where the future trajectories are predicted based on the present condition and previous records. Regression based algorithms (i.e., methods used to learn the relationship between an outcome variable and multiple features), such as support vector regression, decision tree regression and multiple linear regression, which have been used in application domains such as stock price [53], consumers interest [25], and electricity demand [12] may be useful in this context. State-space models [2] and probabilistic models, such as hidden Markov models [42], may be viable alternatives for long term predictions. Finally, deep neural networks (DNN), and particularly long short-term memory (LSTM) networks [46], have recently been successful in capturing temporal dynamics for timeseries prediction tasks in domains including machine translation [51], health care [47], and geoscience [50]. Given the strong biases that may be present in HMIS data as discussed in prior sections, naive application of DNNs should be avoided. More general, DNNs have been thus far hard to interpret. Therefore although developing models that can achieve high accuracy on average may be desirable, care must be taken to additionally address fairness and interpretability considerations.

Figure 6: Correlation coefficient of highly important features; the higher the correlation, the redder the corresponding cell.

5.2 Mining Relationships in the Data

Data science can help identify hidden or previously unknown relationships within the data and external relationships with exogenous factors. Using data science analysis (e.g., correlation coefficient of features) can provide relationship between features of each data compared with comprehension by textual description. Figure 6 shows identified correlation between features. Note that the feature pair of "domestic violence victim" and "benefit from any source" has a significantly higher correlation score compared with the feature pair of 'employed' and "earned", despite common sense that the latter pair should exhibit a stronger relationship. Identifying correlations among features automatically can assist in predicting one feature from another, which can be used for instance to impute missing values.

Also, identifying patterns can help the government and service providers to allocate resources to prevent homelessness. For example, the school and society can pay more attention to teenagers' living, mental, and physical conditions in advance by knowing who have higher risks of becoming homeless based on the analyses of demographic and societal data. Without the intervention of data science methods to identify the pattern, it is hard to manually find the potential target and offer timely assistance in advance. Similarity estimation is one of the methods for data patter identify, which can be achieved by Minkowski distance [49] and Jaccard index [22]. Minkowski distance [49] can be used to measure the high dimensional distance between samples, and the shorter the distance, the more similarity between the samples. Jaccard index [22] can be used to measure the overlap union between samples, and the more union between samples, the more similarity of them.

Data science can also help to mining the relationship between data and certain service programs by extracting the common characteristics of homeless people who received the same service program.

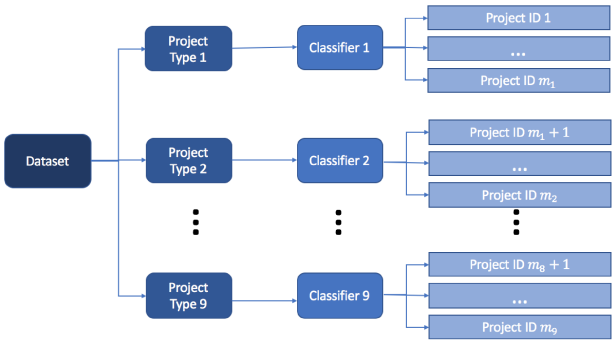


Figure 7: Illustration of hierarchical learning structure; m denotes project ID.

By extracting the common characteristics of the served people, service programs can be improved in an efficient way. For example, assuming that a common characteristic for people who were allocated to an emergency shelter is poor health conditions. With an acknowledgment of that point, more doctors or medical equipment should be devoted to the emergency shelter. It is hard to manually conclude the common characteristics from a large number of homeless people with different background entered a certain services program. However, multiple regression [23] is one potential method for such purpose. The advantage of multiple regression is the ability to determine the relative influence of one or more predictor variables on the outcome variable and the ability to identify outliers especially for the real-word dataset.

5.3 Incorporating Hierarchical Relationships

Homelessness service programs (i.e., project type) shown in Figure 3 and their corresponding subprograms (i.e., project ID) shown in Figure 4, form a natural hierarchy which can be beneficial in improving classification accuracy for both levels. One possible hierarchical learning model structure is shown in Figure 7. As nine project types exist in the dataset, nine distinct multi-class classification models can be trained to predict the project ID for a new homeless individual, within a specific project type, in accordance to the nine project types in the HMIS data. The advantage of restricting classification to project IDs within a project type, once the alignment of a homeless people to a project type has been determined, is twofold: (i) specialized classifiers can be learned given some context (i.e., by exploiting implicit relationships between project IDs and data of an individual based on their distinction by type), and (ii) class imbalance can be avoided by flattening the probability distribution of project IDs within a given type.

5.4 Causal Discovery and Causal Attribution

In the homeless service domain, where observations depend not only on available resources, but also on unobserved dynamics due to human behavior and interactions (e.g., relationship with family members), it is important to quantify potential causal relationships. For instance, [56] examined the relationship between homeless families' stay patterns in homeless shelters and the likelihood of

readmission [48]. Without randomized experiments and causal reasoning, it is almost impossible to draw safe conclusion on the common characteristics of particular programs with certain group of people. Data science methods (e.g., Bart (Bayesian Additive Regression Tree) Model [13] [29]) provide opportunity for causal mining using counterfactual data. Both Granger causality [21], which defines causality in terms of predictability, and Pearl causality [41], which defines causality in terms of changes resulting from intervention, have yielded tremendous breakthroughs in other domains over the past decade, and have great potential for tasks related to homeless service provision. Such tasks range from variable selection for estimation and prediction to identifying causal interactions, and causal attribution [34].

5.5 Explainability

In data science, explainability refers to peoples' ability to explain the output of a machine learning model [43]. For example, in homelessness services systems, the reasons for a decision matter [44]. For example, we can provide a convincible allocation explanation which may be inquired by the clients [44]. One possible way to achieve this is based on learning different weights across features (i.e., a feature with higher weight means that this feature has more influence on the decision). At the same time, for regulatory purposes, in homelessness service allocation system where questions of accountability and transparency are particularly important, for each allocation decision we need to properly deliver evidence-based reason from the scientific perspective.

5.6 Fairness

Fairness becomes an issue when automated methods are involved in the decision making process [4]. Machine learning algorithms are intrinsically fair, however, biases in the data itself used for training can introduce systematic biases that if left unchecked can lead to the perpetuation of inequities [29]. Recent advances in data science can help to both formalize and quantify biases with respect to individual– and group–fariness [4]. Specifically, individual fairness refers to a classifier's ability to consistently produce similar outcomes for similar individuals [16, 31]. Group fairness is described as the ability of a classifier to predict a particular outcome for individuals across groups with almost equal probability [19]. In HMIS data, measuring group fairness is critical because of potential biases towards sex, age and race (Section 4.1). One cluster-based visualization method Principal Component Analysis have recently been used for group fairness evaluation [33]. Alternative methods to be explored include but are not limited to spotting the sensitive features by counterfactual approaches [35] and then deleting those features from the feature space to reduce the effect of bias.

We illustrate this point using a non-parametric approach to Bayesian regression (BART) [13]. BART is based on the "sum of trees model, where each tree is restrained by a regularization prior, and samples are drawn from a posterior distribution by the Bayesian back-fitting MCMC algorithm [17]. In our study, we train BART using a susbet of available data, and use the trained model to generate posterior draws for each individual, which allows inference on both population- and individual-level. We use the R package bartMachine [26] to tain a BART model over N records $O_i = (x_i, y_i), 1 \leq N$ in our HMIS dataset (Section 3). Each record O_i captures information about a homeless client, and is represented by feature vector $x_i = [x_1, ..., x_M, x_d, x_t]^T$, where x_d refers to the assigned project ID , and x_t refers to the assigned service program. Each record is additionally associated with a binary label $y_i \in \{-1, 1\}$, which represents reentry, i.e.,$y_i = 1$ means that the individual reentered the homelessness service system, otherwise -1. We split the dataset into test and training sets with a ratio of $1 : 4$. BART predicts that $1,020$ (21.52%) individuals will reenter the homeless support system in the test set, as opposed to $1,093$ (23.06%) individuals who actually reentered. In line with [29], this result demonstrates that the BART model is well-calibrated and fair because outcomes closely match actual observational data with the assumption that the actual homelessness service allocation is fair.

6 CONCLUSION

Homelessness service provision is a social, public health, and housing policy challenge of great societal interest that impacts the lives of actual human beings. The list of challenges and promising research directions provided in this work may not not be exhaustive, but it illustrates the emerging possibilities of future data science research in this important area. We strongly believe that successful application of data science methods in this domain will be driven by a specific question arising in the homelessness service provision, and that the best recipe for success is for a data science researcher (or team of researchers) to collaborate very closely with a domain expert or practitioner during all phases of research. That is because domain experts and practitioners are in a better position to understand which variables "make sense" to use to answer a given question, and the weaknesses inherent in the data collection process. Likewise, data science researchers are better placed to decide which methods are better suited to a given question, and what are the realistic expectations practitioners should have when applying such methods. Furthermore, inherently transparent methods should be preferred so as to improve interpretability and reduce the perpetuation of inequalities. Finally, evaluation methods must be revisited, particularly when performance is measured based on counterfactual allocation using observational data. In conclusion, frequent communication and continuous collaboration between data science researchers and domain experts is required to ensures that data science methods can indeed have a positive outcome for homelessness service provision.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. ECCS-1737443.

REFERENCES

[1] Kate Amore, Michael Baker, and Philippa Howden-Chapman. 2011. The ETHOS definition and classification of homelessness: an analysis. *European Journal of Homelessness* 5, 2 (2011).
[2] Masanao Aoki. 2013. *State space modeling of time series.* Springer Science & Business Media.
[3] Ricardo Baeza-Yates. 2016. Data and algorithmic bias in the web. In *Proceedings of the 8th ACM Conference on Web Science.* 1–1.
[4] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2017. Fairness in machine learning. *NIPS Tutorial* 1 (2017).
[5] Mark F Basquill, Christine Maguth Nezu, Arthur M Nezu, and Tamara L Klein. 2004. Aggression-related hostility bias and social problem-solving deficits in

adult males with mental retardation. *American Journal on Mental Retardation* 109, 3 (2004), 255–263.
[6] EA Benfer, DB Robinson, S Butler, et al. 2020. The COVID-19 eviction crisis: an estimated 30-40 million people in America are at risk. The Aspen Institute. Published August 7, 2020.
[7] S Berg. 2015. Ten-year plans to end homelessness. *National Low-Income Housing Coalition* (2015).
[8] Molly Brown, Danielle Vaclavik, Dennis P Watson, and Eric Wilka. 2017. Predictors of homeless services re-entry within a sample of adults receiving Homelessness Prevention and Rapid Re-Housing Program (HPRP) assistance. *Psychological services* 14, 2 (2017), 129.
[9] Martha R Burt et al. 2002. Evaluation of continuums of care for homeless people. (2002).
[10] Thomas Byrne, Dan Treglia, Dennis P Culhane, John Kuhn, and Vincent Kane. 2016. Predictors of homelessness among families and single adults after exit from homelessness prevention and Rapid Re-Housing Programs: Evidence from the Department of Veterans Affairs Supportive Services for Veteran Families program. *Housing Policy Debate* 26, 1 (2016), 252–275.
[11] Hau Chan, Eric Rice, Phebe Vayanos, Milind Tambe, and Matthew Morton. 2017. Evidence From the Past: AI Decision Aids to Improve Housing Systems for Homeless Youth.. In *AAAI Fall Symposia*. 149–157.
[12] Yongbao Chen, Peng Xu, Yiyi Chu, Weilin Li, Yuntao Wu, Lizhou Ni, Yi Bao, and Kun Wang. 2017. Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. *Applied Energy* 195 (2017), 659–670.
[13] Hugh A Chipman, Edward I George, Robert E McCulloch, et al. 2010. BART: Bayesian additive regression trees. *The Annals of Applied Statistics* 4, 1 (2010), 266–298.
[14] Dennis R Combs, David L Penn, Melanie Wicher, and Evan Waldheter. 2007. The Ambiguous Intentions Hostility Questionnaire (AIHQ): a new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive neuropsychiatry* 12, 2 (2007), 128–143.
[15] Lorcan Coyle and Pádraig Cunningham. 2004. Improving recommendation ranking by learning personal feature weights. In *European Conference on Case-Based Reasoning*. Springer, 560–572.
[16] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*. 214–226.
[17] Daniel Eaton and Kevin Murphy. 2012. Bayesian structure learning using dynamic programming and MCMC. *arXiv preprint arXiv:1206.5247* (2012).
[18] Daniel James Fuchs. 2018. The dangers of human-like bias in machine-learning algorithms. *Missouri S&T's Peer to Peer* 2, 1 (2018), 1.
[19] Pratik Gajane and Mykola Pechenizkiy. 2017. On formalizing fairness in prediction with machine learning. *arXiv preprint arXiv:1710.03184* (2017).
[20] Yuan Gao, Sanmay Das, and Patrick Fowler. 2017. Homelessness service provision: a data science perspective. In *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*.
[21] Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* (1969), 424–438.
[22] Lieve Hamers et al. 1989. Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula. *Information Processing and Management* 25, 3 (1989), 315–18.
[23] Timothy Hanson. 2010. Multiple regression.
[24] M Henry, R Watt, A Mahathey, J Ouellette, and A Sitler. [n.d.]. The 2019 Annual Homeless Assessment Report (AHAR) to Congress. Part 1: point-in-time estimates of homelessness [Internet]. United States: 2020.[access at: July 31, 2020].
[25] Hsien-You Hsieh and Shih-Hung Wu. 2015. Ranking online customer reviews with the SVR model. In *2015 IEEE international conference on information reuse and integration*. IEEE, 550–555.
[26] Adam Kapelner and Justin Bleich. 2013. bartMachine: Machine learning with Bayesian additive regression trees. *arXiv preprint arXiv:1312.2171* (2013).
[27] Bonnie D Kerker, Jay Bainbridge, Joseph Kennedy, Yussef Bennani, Tracy Agerton, Dova Marder, Lisa Forgione, Andrew Faciano, and Lorna E Thorpe. 2011. A population-based assessment of the health of homeless families in New York City, 2001–2003. *American Journal of Public Health* 101, 3 (2011), 546–553.
[28] Paul Koegel, Greer Sullivan, Audrey Burnam, Sally C Morton, and Suzanne Wenzel. 1999. Utilization of mental health and substance abuse services among homeless adults in Los Angeles. *Medical Care* 37, 3 (1999), 306–317.
[29] Amanda Kube, Sanmay Das, and Patrick J Fowler. 2019. Allocating interventions based on predicted outcomes: A case study on homelessness services. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 622–629.
[30] K Ming Leung. 2007. Naive bayesian classifier. *Polytechnic University Department of Computer Science/Finance and Risk Engineering* 2007 (2007), 123–156.
[31] Binh Thanh Luong, Salvatore Ruggieri, and Franco Turini. 2011. k-NN as an implementation of situation testing for discrimination discovery and prevention. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. 502–510.
[32] Maggie McCarty, Libby Perl, and Katie Jones. 2014. Overview of federal housing assistance programs and policy. Congressional Research Service, Library of Congress.
[33] Marius Miron, Songül Tolan, Emilia Gómez, and Carlos Castillo. 2020. Addressing multiple metrics of group fairness in data-driven decision making. *arXiv preprint arXiv:2003.04794* (2020).
[34] Raha Moraffah, Mansooreh Karami, Ruocheng Guo, Adrienne Raglin, and Huan Liu. 2020. Causal interpretability for machine learning-problems, methods and evaluation. *ACM SIGKDD Explorations Newsletter* 22, 1 (2020), 18–33.
[35] Ramaravind K Mothilal, Amit Sharma, and Chenhao Tan. 2020. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 607–617.
[36] House of Commons Committee of Public Accounts. 2017. Homeless households: eleventh report of session 2017 - 2019. (2017).
[37] United States Department of Housing and Urban Development. 2019. HMIS Guides and Tools. *Retrieved December 14, 2020, from https://www.hudexchange.info/programs/hmis/hmis-guides/* (2019).
[38] Kristen Olson. 2006. Survey participation, nonresponse bias, measurement error bias, and total bias. *International Journal of Public Opinion Quarterly* 70, 5 (2006), 737–758.
[39] Marion Oswald, Jamie Grace, Sheena Urwin, and Geoffrey C Barnes. 2018. Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental'proportionality. *Information & Communications Technology Law* 27, 2 (2018), 223–250.
[40] Delroy L Paulhus. 1991. Measurement and control of response bias. (1991).
[41] Judea Pearl and Thomas S Verma. 1995. A theory of inferred causation. In *Studies in Logic and the Foundations of Mathematics*. Vol. 134. Elsevier, 789–811.
[42] Lawrence Rabiner and Biinghwang Juang. 1986. An introduction to hidden Markov models. *ieee assp magazine* 3, 1 (1986), 4–16.
[43] Ribana Roscher, Bastian Bohn, Marco F Duarte, and Jochen Garcke. 2020. Explainable machine learning for scientific insights and discoveries. *IEEE Access* 8 (2020), 42200–42216.
[44] R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke. 2020. Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access* 8 (2020), 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199
[45] Barbara Sard. 2009. Number of homeless families climbing due to recession. *Center on Budget and Policy Priorities, Special Series: Economic Recovery Watch* (2009).
[46] Jürgen Schmidhuber and Sepp Hochreiter. 1997. Long short-term memory. *Neural Comput* 9, 8 (1997), 1735–1780.
[47] Ying Sha and May D Wang. 2017. Interpretable predictions of clinical outcomes with an attention-based recurrent neural network. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 233–240.
[48] Xinpeng Shen, Sisi Ma, Prashanthi Vemuri, and Gyorgy Simon. 2020. challenges and opportunities with causal Discovery Algorithms: Application to Alzheimer's pathophysiology. *Scientific reports* 10, 1 (2020), 1–12.
[49] Archana Singh, Avantika Yadav, and Ajay Rana. 2013. K-means with Three different Distance Metrics. *International Journal of Computer Applications* 67, 10 (2013).
[50] Ziheng Sun, Liping Di, and Hui Fang. 2019. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *International journal of remote sensing* 40, 2 (2019), 593–614.
[51] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215* (2014).
[52] Heather D Tevendale, W Scott Comulada, and Marguerita A Lightfoot. 2011. Finding shelter: Two-year housing trajectories among homeless youth. *Journal of Adolescent Health* 49, 6 (2011), 615–620.
[53] Theodore B Trafalis and Huseyin Ince. 2000. Support vector machine for regression and applications to financial forecasting. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, Vol. 6. IEEE, 348–353.
[54] Chikako Van Koten and AR Gray. 2006. An application of Bayesian network for predicting object-oriented software maintainability. *Information and Software Technology* 48, 1 (2006), 59–67.
[55] Till Von Wachter, Marianne Bertrand, Harold Pollack, Janey Rountree, and Brian Blackwell. 2019. Predicting and Preventing Homelessness in Los Angeles.
[56] Yin-Ling Irene Wong, Dennis P Culhane, and Randall Kuhn. 1997. Predictors of exit and reentry among family shelter users in New York City. *Social Service Review* 71, 3 (1997), 441–462.