Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 185 (2021) 9–18

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference Theme: Big Data, IoT, and AI for a Smarter Future
Malvern, Pennsylvania, June 16-18, 2021

Smart Homelessness Service Provision with Machine Learning

Charalampos Chelmis[a]*, Wenting Qi[a], Wonhyung Lee[b], Stephanie Duncan[b]

[a]*Department of Computer Science, University at Albany, Albany, NY 12222, USA*
[b]*School of Social Welfare, University at Albany, Albany, NY 12222, USA*

Abstract

Homelessness presents a long-standing social problem for nearly every community across the world. A key goal of homelessness service provision is to reduce the number of individuals who experience repeated episodes of homelessness. The goal of this work is to determine the feasibility of an automated recommendation system designed to carefully match individuals to homelessness service facilities when they first experience homelessness. Specifically, machine learning methods are used to recommend the exact service facility that a homeless individual can benefit from among other numerous homeless-serving organizations in the Capital Region of New York, based on individual time-variant (e.g., monthly income, age) and time-invariant (e.g., race, sex) features. The data used in the study span a total of 38, 800 individuals seeking federally funded homelessness assistance from 2005 through 2019. The best performing method achieves an accuracy of 81.5%.

1. Introduction

In the wake of the global pandemic, and the ensuing rise of unemployment, homelessness, which is defined by the federal government as staying in a residence not habitable for permanent living conditions, remains to be a persistent problem in the United States, where an estimated 553,772 people experience homelessness on a given night [1].

Homelessness represents a complex public health and social problem that depends on various factors. Most often insurmountable challenges lead to homelessness, which include loss of a job, medical issues, family breakup or violence, substance abuse, mental health problems and/or involvement with the child welfare system at the individual

* Corresponding author. Tel.: +1-518-437-4948.
E-mail address: cchelmis@albany.edu

1877-0509 © 2021 The Authors. Published by Elsevier B.V.

10.1016/j.procs.2021.05.002

level. A lack of affordable housing is another significant factor at the societal level [2]. The United States has 7.4-million-unit shortage in affordable housing [3], which is also associated with increasing evictions. Between 2000 and 2016, an estimate of 61 million eviction cases were filed in the nation [4]. Due to the financial crisis during the pandemic, nearly 19-23 million renters are continually facing the risk of eviction by the end of 2020 [5].

Communities throughout the United States respond to homelessness with a variety of service programs, including emergency shelter (ES), day shelter (DS), transitional housing (TH), rapid rehousing (RRH), homelessness prevention (HP), and street outreach (SO), along with other resources for support. These services are expected to complement one another and form a desirable trajectory for a homeless person to exit homelessness. For example, a shelter provides a place to stay for emergencies whereas Permanent supportive housing (PSH) provides more long-term housing arrangement. It is noteworthy, however, that homelessness is often chronic and repetitive. A shelter is not usually just for a one-time visit; rather, the history of being in a shelter in the past becomes a significant predictor for who might become homeless later [6]. Even long-term solutions such as permanent housing does not mean that the residents will live there permanently. It is found that substantial numbers of PSH residents leave within months of entry, both voluntarily and involuntarily [7], which implies that the cycle of homelessness may continue even after people are placed in PSH.

Can this complex and tenacious issue be benefited from machine learning? In this study, we ask the question of whether accurate recommendations for matching individuals to homelessness service programs when they first experience homelessness can be algorithmically made. We present a proof of concept automated recommendation system which is capable of recommending the exact service facility that a homeless individual can benefit from among other numerous homeless-serving organizations in the Capital Region of New York, based on individual time-variant (e.g., monthly income, age) and time-invariant (e.g., race, gender) features. Our proposed solution could be used as a reference point by communities striving to improve the current housing systems in the future. In contrast, the administrative processes of assessing housing eligibility and assigning individuals to homelessness service programs that are available at any given time is to date manually performed [8].

2. Related Work

In recent years, significant attention has been given on the application of data science to social problems, which was triggered by the need to use resources more efficiently [9]. For example, machine learning can assist homeless-serving agencies to predict the number of beds needed for a shelter on any given night or the trajectories through which homeless people can utilize services and eventually exit homelessness [10]. Such insights can help to identify efficient areas (or blind spots) in the current resource allocation system and vulnerability assessment tools.

Few studies on homelessness and social services have used machine learning, mainly focusing on individual outcomes of homelessness and within other social service systems. For example, [9] used unsupervised cluster algorithms to identify predictors of homelessness and long-term stays at the emergency shelter. [10, 11] used machine learning and secondary data from service provision to understand the prediction of reentry into homelessness. [12] examined the use of machine learning as a tool to aid caseworkers when assisting foster care youth who were at risk of being homeless to transition into adulthood. Thus, machine learning and data science applied to homeless service delivery systems can provide a novel response to this issue [11].

At the same time, the proposed two–stage classification method (c.f. Section 3.2) relates to the problem of hierarchical classification [13]. The most relevant methods include [14] (tree and Directed Acyclic Graph structured hierarchies), [15] (bottom-up multi-label classification), and [16] (global-model approach for hierarchical classification). Unlike such methods, our proposed approach relies on the hierarchical relationship of classes to narrow down the space of available options, and, in doing so, learn more specialized, highly accurate classifiers. To the best of our knowledge, no prior work has explored hierarchical classification for homeless service provision.

3. Data and Methodology

3.1. Data

In most local communities the government has appointed a regional planning body that coordinates services and funding for homeless families and individuals. According to federal mandates, a lead agency within each Continuum of Care (CoC) manages the Homeless Management Information System (HMIS), a local information technology system used to collect client level data, including personally identifiable information, socioeconomic characteristics, and service needs of people experiencing homelessness and seeking federally funded housing assistance [17].

The data used in this study has been collected through the HMIS for the Capital Region of New York State, administered by CARES of NY, Inc., a not-for-profit organization in Albany, New York. In total, the dataset comprises 92,586 records capturing details associated with housing services provided to a total of 38,800 individuals and families seeking federally funded homelessness assistance in communities around the Capital Region from 2005 through 2019. To respect the privacy and confidentiality of homeless individuals, both the name and Social Security Number of service users in the data have been double hashed. To link participants across programs we use a unique, anonymous identification number. This results in a dataset containing individual characteristics available upon entry into the system, as well as information on all entries and exits from different homeless services.

Six types of homeless service programs (hereafter referred to as project types) are considered (as mentioned in Introduction), each of which is further subdivided into subprograms, corresponding for instance to the specific characteristics of the population served (e.g., females or youth), and regional divisions (e.g., neighborhoods covered). Each subprogram is assigned a unique identifier (i.e., project ID) in HMIS, resulting in a total of 153 unique identifiers. Fig. 1(a) shows how the number of enrolments into subprograms varies by client, whereas Fig. 1(b) provides a temporal illustration of entry (and exit) dates into subprograms for individual clients. Evidently, most clients enter the system once (or only few times), with few extreme cases experiencing chronic homelessness and therefore requiring services for up to 50 times. Fig. 1(b) shows that although data collection goes as far back as 1950, most records are captured after 2005. For this reason, this study focuses on the population of individuals and families (99.6% of the total population in the dataset) who used homeless services from 2005 through 2019. Moreover, the target variable for our modelling is the project ID each individual or family has been assigned to. However, Fig. 2 shows that project IDs are not equally represented in the dataset. We therefore restrict our study to 16 project IDs with at least 500 clients (10.45% of all unique project IDs in the original dataset).

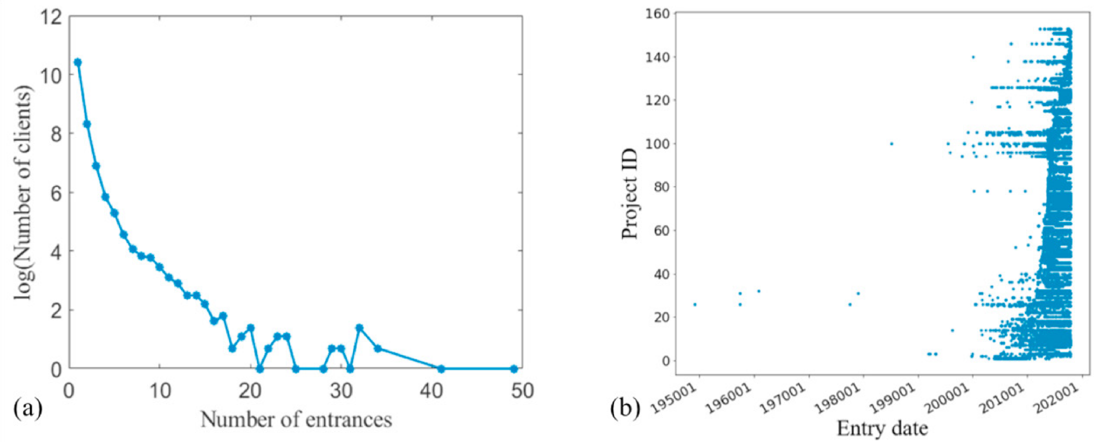

Fig. 1. (a) Number of enrollments per subprogram; (b) Temporal illustration of individual client entries per subprogram.

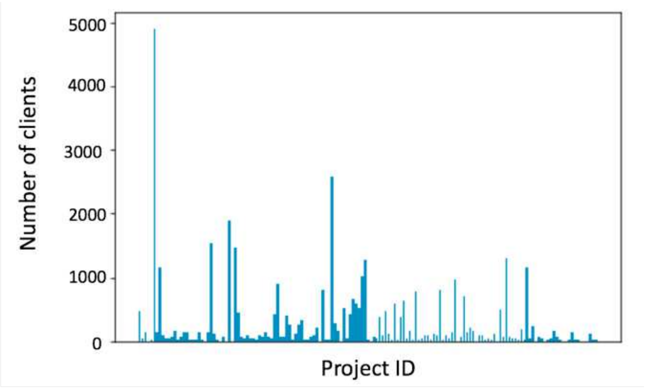

Fig. 2. Number of clients per project ID.

Each record in the HMIS dataset contains both time-invariant (e.g., race, gender, ethnicity) as well as time-variant (e.g., income) features. Specifically, each record comprises health and disability information (such as physical or mental disability and mental health status), income (such as income source and earned amount), enrollment (e.g., length of stay, living situation, and housing information), service (i.e., services received), educational background (such as last grade attended), and employment information (e.g., currently employed) for each client individually. In total, 215 socioeconomic demographics, education background, HUD-defined chronic homelessness, veteran status, and health insurance attributes are available. A complete description of the data elements in HMIS is available at [17].

Intuitively, personal information, such as name, birth date, and client ID cannot be used as accurate predictors of service requests, and thus we exclude such personal attributes from our study. We additionally disregard attributes recorded after the departure of a client from a program, since such data comprise privileged information, which is unavailable at the time of program assignment. Similarly, we consider data recorded during the period between a client entered a given program until the time they exit as relevant to the assignment. An illustration of such “valid period" is shown in Fig. 3. In the end, a total of 174 valid attributes with adequate amounts of available data for use in our model are used. Table 1 summarizes these attributes.

We begin by filtering out attributes whose fraction of missing values is larger than 60%. We then remove attributes with zero variance and low importance. To quantify importance, we utilize the commonly used feature importance function of the lgb.LGBMClassifier() library [18]. After applying this method, we get a score for each feature. The higher the score, the more important the feature is. We set a threshold of 500 to filter out low scoring features. The top informative features ranked by importance are shown in Fig. 4(a). The interested reader may refer to the complete description of HMIS data elements for more details [17]. We finally exclude redundant features [19] by excluding those features whose correlation coefficient with other features exceeds 90% (see Fig. 4(b)).

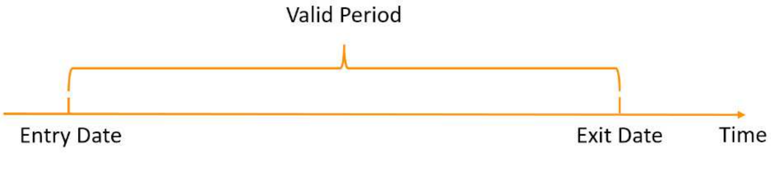

Fig. 3. Illustration of “valid" period. Data collected beyond the exit date is considered privileged information, and therefore unavailable at test time. As such, data collected beyond the “valid" period are excluded for training and testing purposes.

Table 1. Descriptive statistics by project type.

	ES (53.18%)	DS (29.04%)	HP (7.37%)	RRH (4.00%)	SO (3.05%)	TH (1.52%)
Living Situation (ES)	9852 (48.37%)	3132 (27.78%)	1224 (43.35%)	1200 (78.22%)	544(46.53%)	476(81.36%)
Living Situation (Friend/Family)	3786 (28.58%)	1867 (16.56%)	41 (1.45%)	127 (8.27%)	98(8.38%)	68(11.62%)
Living Situation (Rental)	149 (0.73%)	1419 (12.58%)	557 (19.73%)	35 (2.28%)	6(0.51%)	3(0.51%)
Education (> Grade 11)	17327 (85.56%)	9162 (81.28%)	2032 (71.98%)	1260 (82.13%)	1101(94.18%)	273(46.60%)
Homeless times (>2 times)	4440 (21.79%)	1052 (9.33%)	42 (1.48%)	188 (12.25%)	212(18.13%)	53(9.05%)
Disabling (Yes)	5008 (24.58%)	3819 (33.88%)	359 (12.71%)	153 (9.97%)	774(66.21%)	30(5.12%)
Insurance (Yes)	20924 (83.09%)	10498 (93.14%)	2491 (88.23%)	1400 (91.26%)	1106(94.61%)	528(90.25%)
Income (Yes)	4595 (22.56%)	4551 (40.37%)	1111 (39.35%)	367 (23.92%)	63(5.38%)	101(17.26%)
Household Head (Yes)	15907 (78.10%)	10968 (97.31%)	1199 (42.47%)	528 (34.41%)	1168(99.91%)	231(39.48%)

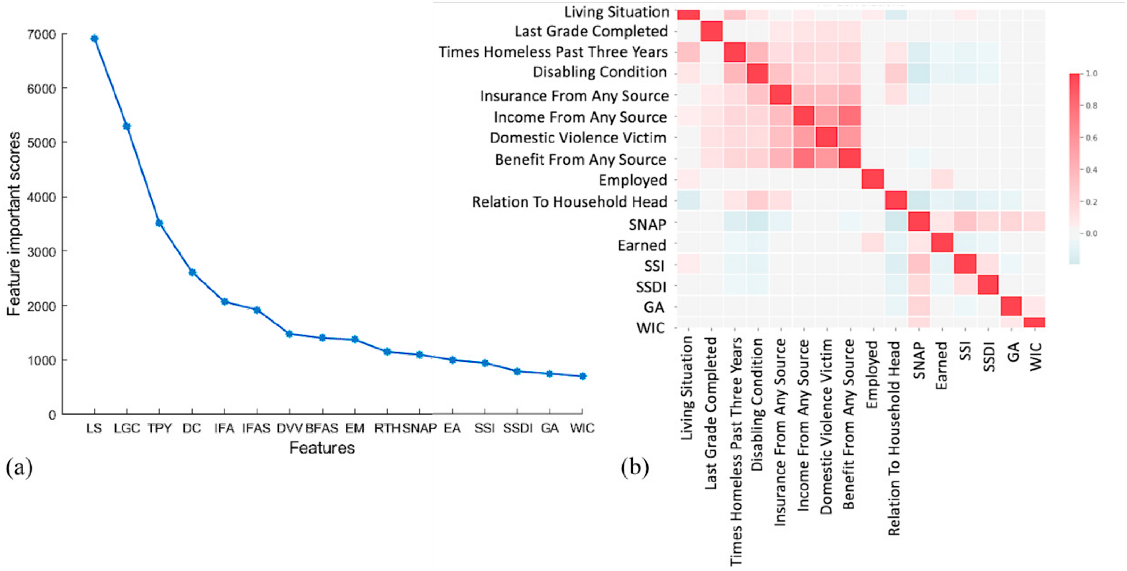

Fig. 4. (a) Features used for classification, ranked by importance. (b) Correlation coefficient of highly important features (better seen in color); the higher the correlation, the redder the corresponding cell is. SNAP, WIC, SSI and SSDI stand for food stamp program, supplemental nutrition program for women, infants and children, supplemental security income, and social security disability insurance.

3.2. Machine Learning for Service Provision

The task of algorithmically recommending a specific project ID for a client based on individual time-variant and time-invariant features can be formulated as follows. Consider dataset Dcomprised of N records $O_i = (x_i, y_i), 1 \leq i \leq N$, one for each client, where each observation O_i is represented by feature vector $x_i = [x_{i1}, \dots, x_{iM}]^T$ of M socioeconomic characteristics, education background, veteran status, disabling conditions, living situation, and benefits, as described in Section 3.1 above. Each observation is additionally associated with label $y_i \in \{1, \dots, 153\}$ that resents a project ID. The task is to learn a multi-class classification model J, capable of accurately predicting the label of a previously unseen observation O_{N+1}, represented by feature vector x_{N+1}, i.e., $y_{N+1} = J(x_{N+1})$.

As six project types exist in the dataset, six distinct multi-class classification models $\mathcal{F} = \{F_1, \dots, F_6\}$ can be trained to predict the project ID y_{N+1} for a new client, within a specific project type. The advantage of restricting classification to project ids within a project type, once the alignment of a client 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 based on their distinction by type), and (ii) class imbalance [20] can be avoided by flattening the probability distribution of project IDs within a given type. Fig. 5 illustrates our two-stage classification approach.

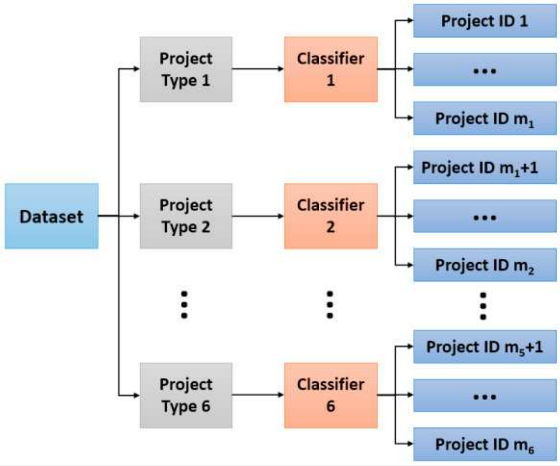

Fig. 5. Illustration of the proposed two-stage classification approach.

We consider three machine learning methods for the task of multi-class classification in this context:

- **K-Nearest Neighbors (KNN):** One of the most popular machine learning algorithms used for multi-class classification [21]. In our study, we use KNN to output a project ID by examining the assignments of k clients that are most similar to the one at hand. In the KNN algorithm, the number of neighbors K has the most influence on the performance of the KNN compared with other parameters such as distance metric and metric parameter [22]. To choose a reasonable value of k [22] and considering the size of the dataset, we use 5-fold cross-validation (i.e., the test to training set ratio is 1:4) and select the best performing parameter for our dataset. Specifically, we train a KNN classifier for different values of $k \in \{1,2,\dots,30\}$, and evaluate its performance on the test data.
- **Random Forest (RF):** Relatively fast compared to other classification models [23], a Random Forrest classifier constructs a collection of decision trees with random subsets of features during the classification process, and the label with the most votes is chosen as the output. To build the RF classifier, we mainly consider two most crucial parameters which are the number of tree n and the function to measure the quality of the split such as GINI and Entropy [30]. For parameter of maximum depth of the tree, we set it as none which means the nodes are expanded until all leaves are pure for keeping the completeness of the tree split. We tune the number of trees (parameter n) to be generated as 500, and the maximum depth unlimited, using the GINI criterion.
- **Multi-class AdaBoost (MA):** Boosting has been a popular technique for two-class classification [24], with multi-class variants having recently been proposed [25]. We experiment with Stagewise Additive Modeling using a Multi-class Exponential loss function (SAMME) [25]. The advantage of SAMME is that it does not simplify the multi-class classification problem as multiple two-class classification tasks. Instead, the label y is encoded into a C dimensional vector l, where C represents the number of classes. The loss function is defined as an exponential function of l, and the goal is to minimize the misclassification rate. For the optimal performance, we use decision tree as the base classifier, and thus, we need to tune for the maximum tree depth which is the most important parameter for decision tree [31].

4. Results and Discussion

4.1. Parameter tuning

Fig. 6 shows the results of our experimentation with the goal of fine tuning the corresponding parameters of each of the three models considered (c.f. Section 3.2). Specifically, Fig. 6(a) shows the accuracy of the KNN classifier as a function of the value of k. Although the plot gives the impression of high variability, the cross-validation accuracy deviation is $\pm 6 \times 10^{-3}$. Similarly, Fig. 6(b) shows the classification accuracy of RF as a function of the number of trees (i.e., parameter n), whereas Fig. 6(c) shows the classification accuracy of MA as a function of tree depth. The best accuracy is obtained for (i) k=14 for KNN, (ii) n=500 for RF, and (iii) maximum tree depth $= 13$ for MA.

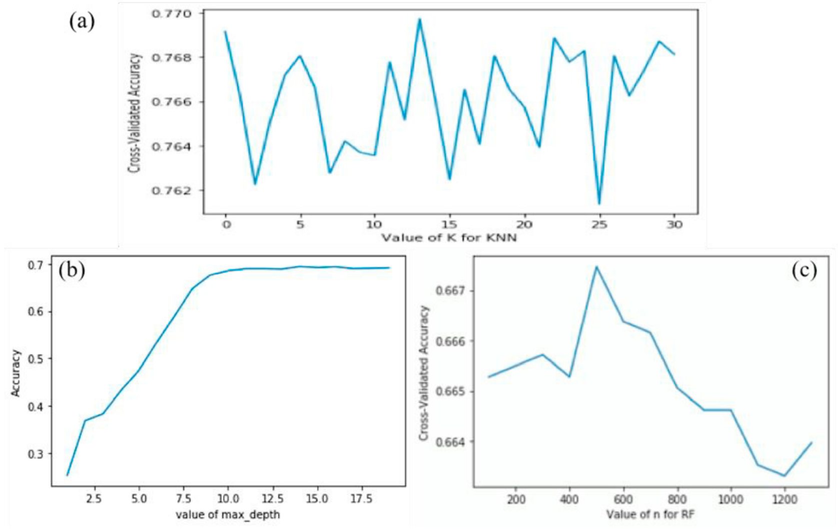

Fig. 6. (a) Cross-validation accuracy of KNN classifier as a function of k (b) Cross-validation accuracy of RF classifier as a function of n (c) Accuracy of MA as a function of tree depth.

4.2. Analysis of Predictive Accuracy

We use average Accuracy and F1-score over five folds for evaluation, as predicted labels fall into one of four categories, namely, true positive (TP), true negative (TN), false positive (FP) or false negative (FN). Two averaging methods are used to calculate the average F1-score as follows: we compute the metric for each class and then take the average (i.e., macro-average), and (ii) aggregate all classes to compute the average metric (i.e., micro-average). We also compute Hamming-loss to quantify the fraction of wrongly predicted labels and the total number of labels.

Finally, note that Section 3.1 details two methods to automatically recommend a project ID for a client, namely, (i) treat the project ID recommendation as a multitask classification task, or (ii) determine a project type first, and restrict recommendations to project IDs for that project type only. In the second scenario, we first split the data into 6 subsets according to project type. We then apply the same classifier on each subset (for each classifier respectively) with parameters being set to the same values as above. Let the fraction of clients in the i-th project type over the total number of clients be denoted as m_i, and the classification accuracy for the i -th project type be denoted as a_i. The overall accuracy of such approach is to be computed as $A_s = \sum_{i=1}^{6} m_i a_i$. The same process is used to compute macro-score and micro-score and Hamming-loss.

Table 2. Overall performance comparison for direct and contextual (i.e., after project type has been determined first) project ID recommendation. Best performing method is highlighted in bold.

Metric	KNN	RF	MA
Accuracy (Higher is better)	0.6217 / 0.7214	0.6219 / 0.7721	**0.6944 / 0.8150**
F1-macro score (Higher is better)	0.5927 / 0.6891	0.6117 / 0.7648	**0.6870 / 0.8011**
F1-micro score (Higher is better)	0.6217 / 0.7214	0.6219 / 0.7721	**0.6944 / 0.8150**
Hamming loss (Lower is better)	0.3782 / 0.2675	0.3780 / 0.2268	**0.3029 / 0.1840**

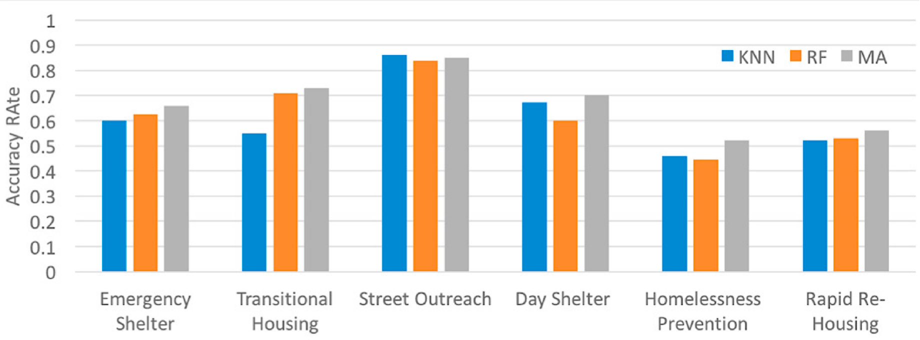

Fig. 7. Direct recommendation accuracy per project ID.

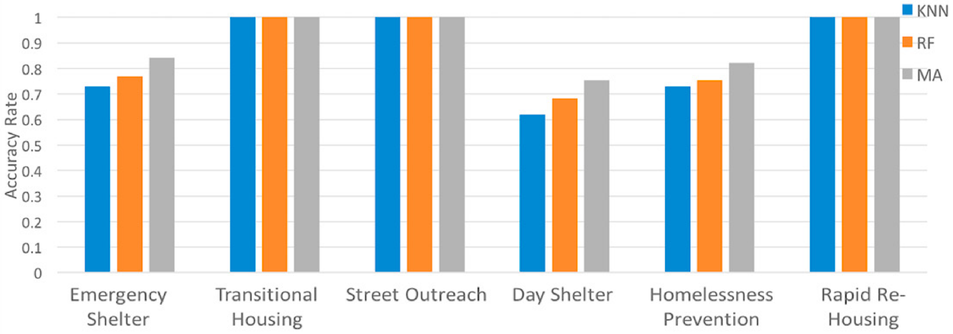

Fig. 8. Contextual recommendation (i.e., after project type has been determined first) accuracy per project ID.

Table 2 summarizes the performance of the three methods for direct project ID recommendation, as well as recommendations when the project type has been determined first. The best performing method is highlighted in bold. Evidently, the context provided by first determining an appropriate project type for a homeless individual, leads to a significant performance improvement (e.g., 17.36% improvement in accuracy; similarly, for the other metrics). Fig. 7 and Fig. 8 show the accuracy of the three methods for direct and contextual project ID recommendations based on the project type. By comparing Fig. 7 with Fig. 8, for all three methods, it is clear that when project type is determined first, the accuracy per project ID improves, particularly for transitional housing, street outreach and rapid re-housing.

5. Ethical Considerations

The ethics of social work have been influenced by the bioethics of human rights and the common principle of protecting the vulnerable, and to advocate for justice, human dignity as well as social equity [26]. Accordingly, the use of machine learning in social services invites more rigorous ethical considerations because it can cause algorithmic bias and discrimination that may continue the pre-existing cycle of injustice [27]. This study is not an exception to concerns of fairness and accountability [28], as it uses real-life data that implies a disproportional representation of certain demographic groups. Additionally, no attempt was made to avoid unintentionally introduced systematic biases and misuses [29], as addressing such issues was not the main goal of this study. Nevertheless, this study promotes human interpretability and encourages recommendations assessment by shedding light into the features used to derive recommendations (Fig. 4). In future work, we plan to first study the correlations among sub-variables in depth, for example, the association between each demographic variable and its service use patterns, identify disproportionate representation or service use, and incorporate such information into machine learning models.

6. Conclusion and future work

This study investigated the feasibility of an automated recommendation system designed to algorithmically match individuals to homelessness service facilities when they first experience homelessness. Experimental evaluation using a real-world dataset indicate the potential of such system. At the same time, it is unclear from the data whether assignments based on the extant homeless system actually address housing needs or if "better" assignments can be made to help prevent future homelessness. Equally important is the challenge of human interpretability of recommendations in order to ensure equity and fairness when deciding how to best allocate scarce, shared societal resources using machine learning approaches. We intent to explore these promising research directions in future work.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. ECCS—1737443 and IIS—1850097.

References

[1] United States Department of Housing and Urban Development. (2019) *HUD's Definition of Homelessness: Resources and Guidance*. Retrieved December 18, 2020. https://www.hud.gov/

[2] Freeman, L. (2002) "America's affordable housing crisis: A contract unfulfilled." *American Journal of Public Health*, **92(5)**, 709-712.

[3] Thurber, A., Krings, A., Martinez, L. S., and Ohmer, M. (2019) "Resisting gentrification: The theoretical and practice contributions of social work." *Journal of Social Work*, 146801731986150. doi:10.1177/1468017319861500

[4] Eviction Lab. (n.d.). *The Eviction Lab*. Retrieved December 11, 2020, https://evictionlab.org/

[5] Aspen Institute. (2020) "The COVID-19 Eviction Crisis: An Estimated 30-40 Million People in America Are at Risk." (2020, September 16). Retrieved December 11, 2020, https://www.aspeninstitute.org/blog-posts/the-covid-19-eviction-crisis-an-estimated-30-40-million-people-in-america-are-at-risk/

[6] Shinn, M., and Cohen, R. (2019) "Homelessness prevention: A review of the literature." Center for Evidence-Based Solutions to Homelessness. http://www.evidenceonhomelessness.com/wp.

[7] Wong, Y., Hadley, T., Culhane, D., Poulin, S., Davis, M., Cirksey, B., and Brown, J. (2006) "Predicting Staying in or Leaving Permanent Supportive Housing That Serves Homeless People with Serious Mental Illness." (Rep.). Philadelphia, PA: University of Pennsylvania Center for Mental Health Policy and Services Research (CMHPSR).

[8] PHA Guidebook to Ending Homelessness. (2013) Programs and Polices: Administering CoC Rental Assistance. https://www.usich.gov/resources/uploads/asset library/PHA Guidebook Final.pdf

[9] Hong, B., Malik, A., Lundquist, J., Bellach, I., and Kontokosta, C.E. (2018) "Applications of Machine Learning Methods to Predict Readmission and Length-of-Stay for Homeless Families: The Case of Win Shelters in New York City." *Journal of Technology in Human Services*, **36:1**, 89-104. doi: 10.1080/15228835.2017.1418703

[10] Kube, A., Das, S., and Fowler, P. J. (2019) Allocating Interventions Based on Predicted Outcomes: A Case Study on Homelessness Services. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 622-629. doi:10.1609/aaai.v33i01.3301622

[11] Gao, Y., Das, S., and Fowler, P. (2017). "Homelessness service provision: a data science perspective." In Workshops at the Thirty-First AAAI Conference on Artificial Intelligence.

[12] Brindley, Meredith & Heyes, James and Booker, Darrell. (2017) "Can Machine Learning Create an Advocate for Foster Youth?" *Journal of Technology in Human Services*. doi: 10.1080/15228835.2017.1416513.

[13] Silla, C. N., and Freitas, A. A. (2011) "A survey of hierarchical classification across different application domains." *Data Mining and Knowledge Discovery*, **22(1-2)**, 31-72.

[14] Bi, W., and Kwok, J. T. (2011) "Multi-label classification on tree-and dag-structured hierarchies." *In Proceedings of the 28th International Conference on Machine Learning*, ICML-11, 17-24.

[15] Cesa-Bianchi, N., Gentile, C., and Zaniboni, L. (2006) “Hierarchical classification: combining bayes with svm.” *Proceedings of the 23rd international conference on Machine learning*, 177-184.

[16] Silla Jr, C. N., and Freitas, A. A. (2009) “A global-model naive bayes approach to the hierarchical prediction of protein functions.” *2009 Ninth IEEE International Conference on Data Mining*, 992-997, IEEE.

[17] 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/

[18] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017) “Lightgbm: A highly efficient gradient boosting decision tree.” *Advances in neural information processing systems*, 3146-3154.

[19] Guyon, I., and Elisseeff, A. (2003) “An introduction to variable and feature selection.” *Journal of machine learning research*, 3(Mar), 1157-1182.

[20] Japkowicz, N., and Stephen, S. (2002) “The class imbalance problem: A systematic study.” *Intelligent data analysis*, **6(5)**, 429-449.

[21] Deng, Z., Zhu, X., Cheng, D., Zong, M., and Zhang, S. (2016) “Efficient kNN classification algorithm for big data.” *Neurocomputing*, 195, 143-148.

[22] Zhang, S., Li, X., Zong, M., Zhu, X., and Cheng, D. (2017) “Learning k for knn classification.” *ACM Transactions on Intelligent Systems and Technology* (TIST), **8(3)**, 1-19.

[23] Quinlan, J. R. (1986) “Induction of decision trees.” Machine learning, **1(1)**, 81-106.

[24] Freund, Y., Schapire, R., and Abe, N. (1999) “A short introduction to boosting.” *Journal-Japanese Society for Artificial Intelligence*, **14(771-780)**, 1612.

[25] Hastie, T., Rosset, S., Zhu, J., and Zou, H. (2009) “Multi-class adaboost.” *Statistics and its Interface*, **2(3)**, 349-360.

[26] Leslie, David, Holmes, Lisa, Hitrova, Christina, and Ott, Eleanor. (2020) “Ethics Review of Machine Learning in Children’s Social Care: Ethics Review of Machine Learning in Children’s Social Care.”

[27] Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., and Cave, S. (2019) “Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research.” London: Nuffield Foundation.

[28] O'neil, C. (2016) “Weapons of math destruction: How big data increases inequality and threatens democracy.” Broadway Books.

[29] Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., and Huq, A. (2017) “Algorithmic decision making and the cost of fairness.” *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, 797-806.

[30] Menze, B.H., Kelm, B.M., Masuch, R. *et al.* (2009) “A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data.” *BMC Bioinformatics* **10,** 213. https://doi.org/10.1186/1471-2105-10-213.

[31] Quinlan, J.R. (1986) “Induction of decision trees.” *Mach Learning*, **1,** 81–106.