# IDENTIFYING A CUSTOMER CENTERED APPROACH FOR URBAN PLANNING: DEFINING A FRAMEWORK AND EVALUATING POTENTIAL IN A LIVABILITY CONTEXT

Golnaz Sarram

# IDENTIFYING A CUSTOMER CENTERED APPROACH FOR URBAN PLANNING: DEFINING A FRAMEWORK AND EVALUATING POTENTIAL IN A LIVABILITY CONTEXT

by

Golnaz Sarram

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Civil Engineering

The University of Memphis

May 2020

# DEDICATION

TO MY PARENT AND MY GRANDMOTHER

## ACKNOWLEDGMENTS

This dissertation marks the end of a long journey through an integrated level of problem-solving but also through an evolution of self. During this doctorate program, I made necessary changes in my career and lifestyle to align my inner passion and talents with where the future of science was about to lie.

This journey started with a mission that too many times changed. I would like to thank my advisor Dr. Stephanie Ivey, for accepting the challenge of supervising an interdisciplinary research project. I am grateful for her patience, constant encouragement, guidance and sharing sensible perspectives.

I would also like to thank my committee who provided the guidance and agreed to adjust the purpose of the research at each step, even when the work was lengthened due to the findings I uncovered.

During this long journey, there were many loving people that I could strongly rely on. Their patience, generosity and flexibility allowed me to prosper. I sincerely appreciate the sacrifices that my dear parents have made and the effort that my grandmother (Parvin) made to plant the seed of education in my mother and me.

To all my friends who intellectually challenged me to rise, I am always thankful for you and the warm supports that you provided during these days. I have them all in my heart and know they will continue to support me in a future of hard work. Because of you, today I am determined to spread your investment of love in me to whoever is in need.

# ABSTRACT

In transportation planning, public engagement is an essential requirement for informed decision-making. This is especially true for assessing abstract concepts such as livability, where it is challenging to define objective measures and to obtain input that can be used to gauge performance of communities. This dissertation focuses on advancing a data-driven decision-making approach for the transportation planning domain in the context of livability. First, a conceptual model for a customer-centric framework for transportation planning is designed integrating insight from multiple disciplines (chapter 1), then a data-mining approach to extracting features important for defining customer satisfaction in a livability context is described (chapter 2), and finally an appraisal of the potential of social media review mining for enhancing understanding of livability measures and increasing engagement in the planning process is undertaken (chapter 3). The results of this work also include a sentiment analysis and visualization package for interpreting an automated user-defined translation of qualitative measures of livability. The package evaluates users' satisfaction of neighborhoods through social media and enhances the traditional approaches to defining livability planning measures. This approach has the potential to capitalize on residents' interests in social media outlets and to increase public engagement in the planning process by encouraging users to participate in online neighborhood satisfaction reporting. The results inform future work for deploying a comprehensive approach to planning that draws the marketing structure of transportation network products with residential nodes as the center of the structure.

## TABLE OF CONTENTS

## List of Tables

## List of Figures

## Introduction

Datasets available through advances in technologies can immensely help in understanding how people use a city's infrastructure from the point of view of mobility, sustainability and environmental impact. Decisions in urban planning that can be improved from the analysis of demographics and personal location data include the mitigation of traffic congestion and planning for high-density development. Focused work in this realm has provoked new efforts in related areas such as livability. The concept of livability is often used for integrating the community quality of life, transportation facility access and neighborhood characteristics while supporting sustainability goals. A focus on improving livability via transportation systems leverages the economy of communities, businesses and consumers.

Addressing the related goals of livability can also help integrate planning processes between different agencies and levels of government. Transportation investment requires a decision-making process that includes planning, programming, implementation, and evaluation, accompanied by federal transportation funds. The process requires decision tools that evaluate up to date and immense information on land use, housing, people’s mobility, economic development, and many other factors. Meaningful information can be inferred from this data that allows support and testing of demographic and economic theories and can lead to sustainable development of smart cities.

An approach with the capability of increasing translation from metric to policy by determining a comprehensive set of influential factors and increasing input from stakeholders is needed. This research focuses on designing a framework for transportation planning using a livability context that integrates concepts from marketing and urban computing to create a richer understanding of stakeholder perceptions. C**hapter 1** develops a conceptual framework towards the application of a more objective means of quantifying livability. The aim was interpreting a linkage between society stated preferences and quantitative measures of livability using combined service industry and urban computing methodologies. **Chapter 2** describes the application of a non-linear model that approximates the relationship between indicator values and livability scores using supervised learning methods. This methodology is a classification and regression analysis associated with learning algorithms which builds a model that assigns each indicator into one category based on a given set of training examples using traditional survey data. **Chapter 3** proposes augmenting this approach using Natural Language Processing (NLP) on social media data to develop a sentiment analysis (SA) package to present how neighborhood satisfaction can estimate residential quality-of-life and identify users' preferences and priorities.

The goal of this research was bridging the gap between residential perspectives and assessment of livability by extracting information using innovative computing methods and data sources. This research developed a conceptual model for a customer-centric approach to transportation planning and evaluated its potential through a livability context. The tools developed through this research advance the state of practice for assessing livability of communities.

# Chapter 1:

## Investigating Customer Satisfaction Patterns in a Community Livability Context: An Efficiency-Oriented Decision-Making Approach

A comprehensive understanding of neighborhood facilities distribution and functions along with residential quality of life satisfaction is a key asset for relating livability management to transportation networks. Due to the simultaneous involvement of varied factors with an individual's perception of livability, this concept is difficult to measure. Therefore, a more objective means of quantifying livability is needed. The service industry has demonstrated the intersection of machine learning classifiers and survey domain knowledge for evaluating users' quality of experiences; however, this process of inquiry-based learning has never been considered for solving the communication difficulties between community stakeholders and transportation agencies. Another area of overlap is that of urban computing, which integrates computing technology in the traditional context of urban areas, connecting ubiquitous sensing technologies, computational power, and data about the urban environment to promote quality of life for people living in a particular community. To this aim, the focus of this study is on interpreting a linkage between society stated preferences and quantitative measures of livability by extracting information from survey-based methods and translating it to a quantitative framework using combined service industry and urban computing methodologies. This work focuses on four transportation planning-related research questions in this blended framework: understanding existing livability patterns, predicting heterogeneous perceptions of quality of life, prioritizing public preferences and developing a multidimensional livability index (MLI).

### 1.1 Introduction

Supporting livability goals through decision-making processes necessitates an effective and interactive discussion among planners and stakeholders leading to identification of a preferred set of indicators that moves beyond traditional contributors. Incorporating state-of-the-art livability performance measures such as accessibility and public health requires defining measureable livability metrics reflecting stakeholders' needs, wants, and behaviors. On the other hand, transportation infrastructure and services are gaining growing attention by decision-makers for the vital role they play in supporting the quality of life of the people served. As of yet, there is no efficiency evaluation involved in the project prioritization process; however, measuring the quality of the transportation network is essential for creating an effective system. This process brings another need for measuring the changes in the condition of the transportation system to determine where to invest or improve it.

Urban computing is an interdisciplinary field integrating computing technology in the traditional context of urban areas. It connects ubiquitous sensing technologies, computational power, and data about the urban environment to promote the quality of life for people influenced by the populated areas. It can help to predict the future of the cities, and to shape the future of urbanism to align with human expectations. The availability of massive amounts of data provides the opportunity to employ urban computing methods to interpret and identify the subjective concepts and attitudes of individuals. Ultimately, relationships between subjective and objective measures of service performance can help to establish a baseline for data-driven performance metrics in the urban-planning domain. Data-driven metrics can address the challenge in connecting planning simulations and configuration designs through augmenting conventional approaches with the new efficiency-based proxy (Zheng et al. 2014).

The service industry has demonstrated the intersection of machine learning classifiers and survey domain knowledge for evaluating users' quality of experiences; however, this process of inquiry-based learning has never been considered for solving the communication difficulties between community stakeholders and transportation agencies (Diaz-Aviles et al. 2015). Thus, this work focuses on four transportation planning-related research questions in the framework of urban computing: understanding existing livability patterns, predicting heterogeneous perceptions of quality of life, prioritizing public preferences and developing a multidimensional livability index (MLI).

## 1.2 Background

The U.S. Department of Transportation (DOT), the U.S. Environmental Protection Agency (EPA), and the U.S. Department of Housing and Urban Development (HUD) entered into an interagency "Partnership for Sustainable Communities" in 2009 outlining an approach to improving quality of life in communities through increasing transportation mode choices while reducing transportation costs and supporting the environment through the incorporation of six principals of livability. These six principles include (FHWA 2011):

1. Provide more transportation choices. Develop safe, reliable, and economical transportation choices to decrease household transportation costs, reduce our nation's dependence on foreign oil, improve air quality, reduce greenhouse gas emissions, and promote public health.
2. Promote equitable, affordable housing. Expand location- and energy-efficient housing choices for people of all ages, incomes, races, and ethnicities to increase mobility and lower the combined cost of housing and transportation.
3. Enhance economic competitiveness. Improve economic competitiveness through reliable and timely access to employment centers, educational opportunities, services, and other basic needs by workers, as well as expanded business access to markets.
4. Support existing communities. Target Federal funding toward existing communities—through strategies like transit oriented, mixed-use development, and land recycling—to increase community revitalization and the efficiency of public works investments and safeguard rural landscapes.
5. Coordinate and leverage Federal policies and investment. Align Federal policies and funding to remove barriers to collaboration, leverage funding, and increase the accountability and effectiveness of all levels of government to plan for future growth, including making smart energy choices such as locally generated renewable energy.

6. Value communities and neighborhoods. Enhance the unique characteristics of all communities by investing in healthy, safe, and walkable neighborhoods—rural, urban, or suburban.

Under this partnership proposal, the FHWA has introduced a set of indicators and performance measures to be able to track progress toward achieving community goals. The provided basic set of livability indicators mostly requires neighborhood demographic percentages (e.g. jobs and housing within one-half mile of transit, household income) that can be applied as livability data into the appraisal system clarifying the range of transportation strategies that should be followed in the ensuing phases of the planning process.

Therefore, identifying the right combination of transportation scenarios and development of a livability scorecard for comparing indicators across different scenarios by mapping potential corridor improvements is recognized as a significant need by the FHWA. The FHWA also emphasizes that performance measures should be directly linked to local and regional vision and goals (ICFI, FHWA 2011). However, considering public involvement in decision making processes efficiently has been a long-time challenge for practitioners. Various numbers of heterogeneous effects drive fluctuations in public perception across different geographical regions. Understanding patterns of public perceptions related to livability can provide a formulated understanding for decision making purposes.

In addition, one of the newly considered performance outcomes by a growing number of transportation agencies related to livability is providing access to opportunity through transportation investments as it gets to the heart of what makes communities livable, and regions economically prosperous and equitable. This led to a partnership of the U.S. Environmental Protection Agency, U.S. Department of Transportation, and Smart Growth America has adopted accessibility as a priority in performance management. While measure of access to the right destinations for meeting community needs seems relatively simple, it can be very complex to measure how well-connected people are to opportunities and resources in their neighborhood, and consequently directing the investment for improving access (GICD 2017). Thus, there is demand from planning agencies for a comprehensive method and instrument to assist the evaluation of various metrics in transportation system performance simultaneously (Lima et al. 2017).

Therefore, this study is aimed at using the current advanced customer-based analytical tools used for evaluating the performance of business firms, digital assets, marketing, fund raising, IT, etc. in combination with efficiency information learned from human infrastructure dynamics to develop a solution for the aforementioned complexity. The result can help inform transportation planning strategies and network design. The ultimate outcome of the framework will be a standardized approach to the analysis and evaluation of urban configurations through developing a proxy for performance of multiple urban metrics (e.g. transit accessibility, walkability, density and diversity in land use) based on the relationship between livability performance measures (LPMs) and customer satisfaction. The resulting multi-dimensional livability index will enable planners to correlate data-driven metrics with survey based public preferences and eventually optimize urban positioning and project prioritization (Lima et al. 2017).

## 1.3 Literature Review

Traditionally, livability policies have been implemented in terms of the agent perspective. Agencies have tried to understand the current conditions in the region in order to set targets. However, the process of translating the collected data into inferable metrics as a baseline for developing these targets has been challenging. For instance, The Metropolitan Transportation Commission (MTC) in the San Francisco Bay Area, has developed the targets for regional plans through extensive stakeholder engagement. Some of these approved targets are access related; for example, the share of jobs accessible within 30 minutes by auto or 45 minutes by transit (increased by 20 percent for congested conditions). Although the agency is considering the information gleaned from public involvement in adjusting the performance targets, there is no computational understanding between stakeholders' level of satisfaction in their neighborhood and their current commuting habits. Increasing job accessibility requires evaluating the performance of the existing opportunities and their positioning in the community, and measuring heterogeneous local user commuting levels of satisfaction. Most regional and statewide plans have suggested detailed policy directions but have not considered a location-based understanding of user perspectives (GICD 2017).

Up to now, there is limited academic literature on understanding livability needs and identifying alternatives for improving community quality of life. There are two major existing handbooks and research centered on understanding livable transit corridors (Ferrell et al. 2016) and an AHP model for quantification of stakeholders' perceived importance (Antognelli et al. 2016). Ferrell et al., 2016, provided a definition of transit corridor livability and set of methods, metrics, tools and strategies for transit corridor stakeholders which was developed based on the available facilities over the user-defined transit corridor area. A spreadsheet-based Transit Corridor Livability Calculator tool was developed that presents a proxy indicator of quality of life (QOL). The necessary data that was provided to the spreadsheet has been used to estimate 10 of 12 defined metrics to gauge livability for user-defined transit corridors. A developed model for validating the scores has been tested intuitively to compare the results for corridors by using the non-auto internal trip capture rate for each transit corridor.

However, relying on an intuitive inference would be based upon personal experience rather than providing an evidence-based reasoning from previous studies. The authors are hypothesizing that with more opportunities in a transit corridor in a sample, and higher metric scores, the QOL would be higher. It is suggested that the more transit, pedestrian, and bicycle trips that both start and end inside the corridor's boundaries, the more livable the corridor (X). However, the assumption that the corridor interior level of activity represents livability can be questioned with high numbers of outgoing trips from supposedly 'livable' areas.

Also, a public attitude consideration on defining the quality of life proxy and travel behavior characteristics has been neglected in this analysis. The metric considers the transportation system's available assets rather than the fact that the utilization rates of these facilities in interaction with various expectations of users would provide a closer performance appraisal. The current study asserts that public expectations and a more aligned mathematical approach should be included in order to make sure that the metric better reflects reality.

Meanwhile, Antognelli et al. 2016 confirms that stakeholder involvement is essential for this assessment. They developed a framework designing a hierarchical classification based on the Common International Classification of Ecosystem Services (CICES) for measuring both ES and US. The implemented approach, in a similar manner to Ivey et al. 2014, is used to structure a model based on Saaty’s Analytical Hierarchical Process (AHP) to quantify the stakeholder views of the importance of livability services. Despite the advantage of considering public opinion in the classification analysis, the shortcomings of the methodology are in reflecting first, the stakeholders’ perception regarding each livability indicator and second, the heterogeneity of the perceptions. Therefore, with this approach, it is not possible to consider multivariate classification analysis to be able to analyze the effects of livability indicators simultaneously and it does not address objective measures in livability perceptions.

The service industry has demonstrated the intersection of machine learning classifiers and survey domain knowledge (e.g. estimating Net Promoter Score (NPS) as a source of user feedback, using customer churn forecasting models and predicting overall user experience) for evaluating users’ quality of experiences; however, this process of inquiry-based learning has never been considered for solving the communication difficulties between community stakeholders and transportation agencies (Diaz-Aviles et al. 2015). Much of the recent research focus for machine learning is on integrating forecasting techniques into enterprise customer care. The related studies are centered on community-based preferences learning (Abbasnejad et al. 2013), approaches for identifying consumer preferences for the design of technology products (Chen et al. 2012), personalizing agent assignment to maximize customer satisfaction and prioritize conversations (Herzig et al. 2016), and understanding customer decision-making strategy from using cognitive and emotional analysis (Sylcott et al. 2011). The approach of combining service industry and urban computing frameworks to the transportation domain is entirely novel in this field; the only similar work is a non-parametric statistical analytics technique aimed at evaluating the possibility of relationships between subjective survey results and hundreds of objective measures to live data to construct a general predictable livability index representing the rank of different cities' livability at a full unit level. Subsequently, the model was supposed to be validated by machine learning methods but the numerical results have never been presented since the idea was first proposed in 2014 (Geers et al. 2014). The current study concentrates on designing similar virtual assistant solutions for transportation planning tasks. This innovation can open doors into more flexible and data-driven strategies for improving urban livability and quality of life.

According to FHWA PL0159 and Wang et al. 2015, obtaining livability performance measures is difficult due to traditional data challenges - the subjective nature and the lack of related survey data. Also, it is mentioned that LPMs include a variety of multidisciplinary elements each involving measuring a separate multidimensional metric. Based on these studies, data availability is a key limiting factor in collecting and using LPMs and there is a tension between using available data and the data that are appropriate. After a comprehensive review, PL0159 developed 12 indicator types related back to the six livability principles as follows: accessibility, aesthetic and sensory, community amenities, community engagement, economic, housing, land use, mobility, natural resources, public health, safety, and socio-cultural. The percentage of key resources addressing each livability indicator type shows that economic, accessibility and connectivity, mobility, natural resources and safety are the most available sources.

Primary studies in the public transit area of the transportation planning domain investigated the effect of service attributes on overall service quality (e.g. Eboli and Mazzulla 2007; Dell'Olio et al. 2010; Diana 2012; de Ona et al. 2013). However, most of the studies traditionally neglected objective measures and did not include attribute-based satisfaction with related objective measures. On the other hand, some studies attempted to integrate quantitative and qualitative measures based on quality of service (Eboli and Mazzulla, 2011; Cascetta and Carteni, 2014). Although, the performed approaches tried to develop quality-based methods, the relationship between attribute-based satisfaction and performance measures were not validated.

Later, de Ona et al. 2015, and Kim et al. 2017, implemented data-mining techniques to classify the sample of the users and develop models capable of analyzing the heterogeneity of customer perceptions in a transit context. De Ona et al. 2015, developed a classification and decision tree method (CART) to identify the attributes with most influence on public railway quality of service. The methodology stratifies the sample of users by four criteria according to their travel habit profiles in order to investigate the heterogeneity of passengers' perceptions. The study provided useful findings for policy makers about the passengers' preferences and needs; however, there are some limitations mentioned with this approach. The authors recommended testing other methods capable of providing a confidence interval or probability level to the splitters and predictions in the model; identifying variations in the effects some variables have on users' perceptions; and predicting service quality using regression models that forecast a value of the overall quality instead of classification predictions.

Recently, Kim et al. 2017, promoted public transit evaluation through understanding the relationship between level of service (LOS) and satisfaction. The study incorporates quantitative satisfaction measures and qualitative measure for the assessment of transport policy, using pattern recognition models. The technique presents the probability that a person belongs to each component; therefore, it can explore the heterogeneity of every group including members with different degrees of satisfaction experienced under similar LOS. In terms of objective measures, the study selected three performance measures (out-of-vehicle times, headway and access time) that can affect travel behavior of current and potential public transportation customers.

Stated preferences methods and pattern recognition techniques have been applied to multiple transportation settings such as travel behavior and mode choice studies, Hensher D. 1992 and 2008. Particularly, the Gaussian mixture model (GMM) employing the Expectation-Maximization (EM) algorithm has shown promising results regarding perception analysis (Ben-Akiva and Bruno, 1995). Recently, mixture models have been applied to transit satisfaction data for understanding the relationship between a Likert scale and the true degrees of user satisfaction (Kim & Chung 2016; Kim et al. 2017). The advantage of this technique is the ability to represent continuous densities and to allow for multimodality and asymmetry of the underlying distribution (Bishop 2006).

This approach is consistent with current service industry studies on customer loyalty prediction models in encouraging the companies to use multidimensional approaches for predicting customer behavior more efficiently. A new study of Cambridge by Zaki et. al. (2016) shows that the NPS measure does not offer an explanation of the root causes of a low score. Also, it is claimed that a single-question customer metric based on the customers' attitude on how likely is he/she to recommend a company is not enough and it needs to adopt a more nuanced multidimensional approach for predicting customers' actual behavior. Therefore, the study integrates multiple quantitative and qualitative sources of customer data to examine the combination of attitudinal, behavioral and demographic data in assessing customer loyalty. A new model transformed the transactional data into profitability scores through employing the Recency, Frequency and Monetary (RFM) technique. The complaint status of each customer is also determined based on the linguistic text-mining approach through mining the textual survey feedback. Eventually, they used a predictive analytics model to predict the customers that are likely to churn. The approach offers a new way to utilize data more efficiently and provides rich insights for improving customer experiences.

### 1.4 A Conceptual Model for a Customer-Centric Livability Framework

Considering current practice related to assessing livability, shortages in stakeholder involvement and exploring user satisfaction for livability analysis, it is proposed that a linkage between society stated preferences and quantitative measures of livability be interpreted using insight from multiple disciplines. Figure 1 describes a conceptual model for exploring objective livability performance measures and subjective livability perceptions of community residents (satisfaction) that can address the diversity of satisfaction as well. The expectation is that this approach will result in less need for and reliance on survey-based methods; however, community engagement will always be important to validate the models and to reflect changes in preferences over time. It is anticipated that richer dialogue can be developed between planners and communities through this work by developing greater understanding of influential factors and means for assessing impacts to community livability. Ultimately, the goal is to discuss and implement planning with utility-based and multivariate analysis but in order to do this, we have to examine the reverse planning process first.

Phase I of this conceptual model involves defining livability metrics through data obtained from the customers, which in this case are residents of a community. It is proposed that a survey-based approach using supervised learning methods to prioritize society stated preferences be examined (chapter 2). Next, a data-driven approach based on self-generated online datasets should be explored (chapter 3). It is expected that this combination of methodologies will enhance understanding of factors influencing perceptions of livability and will offer planners insight that will enable more data-driven dexcisions. This methodology can also extend the regular estimation approach of simple location choice modeling, implemented in current state and metropolitan travel demand models.

The Phase II approach is based on extracting activity rules by tracking commuters needs for fulfilling their daily activities through mode choices and different trip purposes. This analysis provides the necessary data for developing a quantitative livability metric that addresses multimodal and sustainable planning goals through measuring the land use efficiency and reliable mobility. Here, the aim is developing a predictive mechanism using machine learning classifiers in order to correlate the survey analysis (including residential demographics) with transportation facility utilization rates to predict deficiencies and classification scores. Therefore, for the purpose of measuring this efficiency and correlating with the perceptions, the data type that is needed should reflect infrastructure usage and user choices rather than inventory data. Eventually, the approach applies a joint distribution over observed and latent variables allowing complicated distributions to be formed from simpler components. This helps in developing a quantitative livability metric and decision-making tool at the same time. It is expected that this model can reflect the complexity of livability and better communicate the policies aimed at improving it. The remainder of the work presented in this dissertation will focus on establishing a process for the Phase I approach. The Phase II approach will be left as future work.

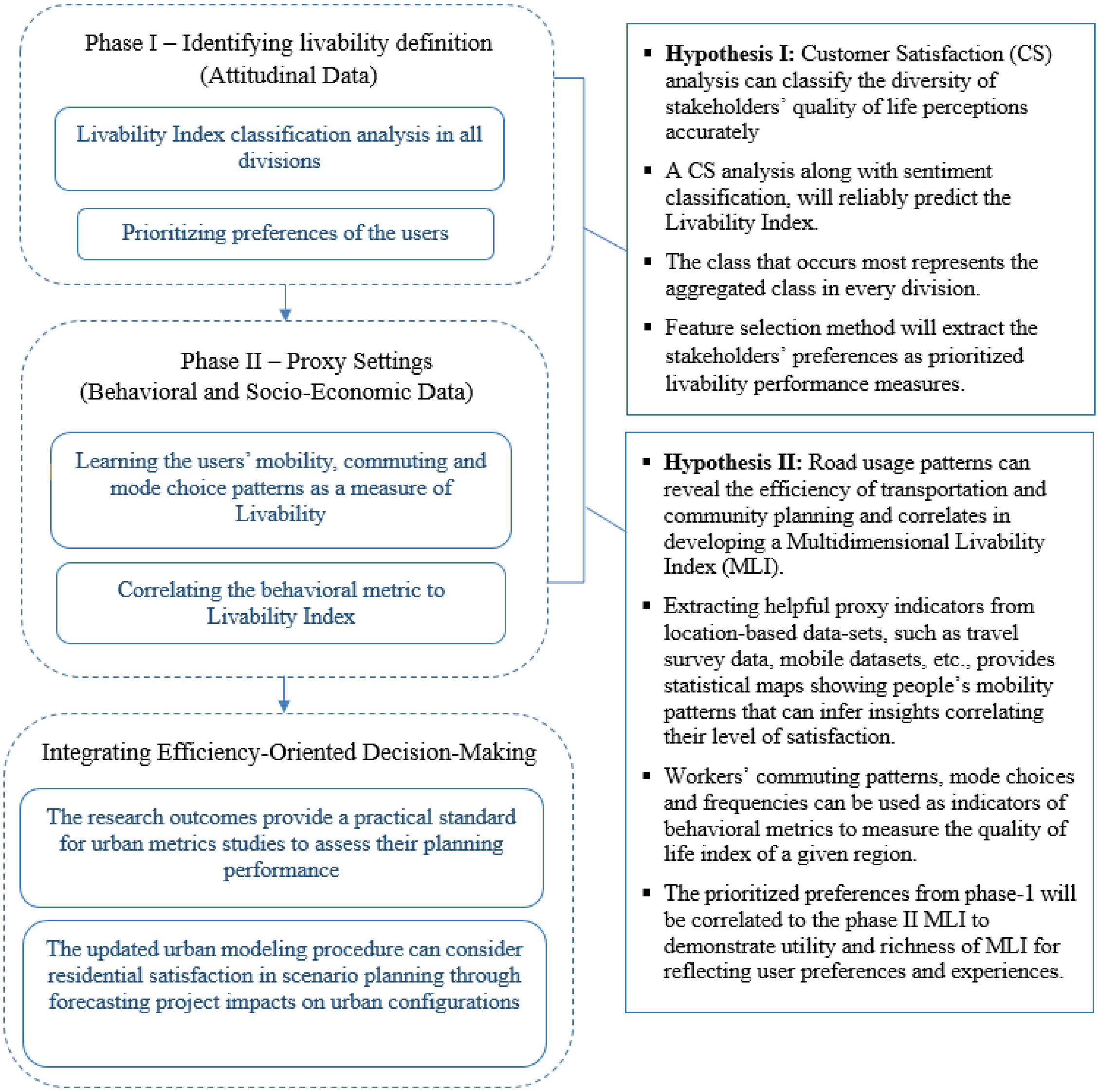


*Figure 1.1 Conceptual Model for a Customer-Centric Livability Framework*

## References

Antognelli, S. and Vizzari, M. (2016) "Ecosystem and urban services for landscape livability: A model for quantification of stakeholders' perceived importance.", Land Use Policy, 50, 277–292.

Abbasnejad E., Sanner S., Bonilla E. V. and Poupart. P. (2013) “Learning Community-based Preferences via Dirichlet Process Mixtures of Gaussian Processes.” *In Proceedings of the 23rd International Joint Conference on Artificial Intelligence* (IJCAI-13).

Ben-Akiva, M., & Bruno, B. (1995). “Discrete choice models with latent sets.” *International Journal of Marketing*, 12(1), 9–24.

Ben-Akiva, M., and Lerman, S. R., (1985) “Discrete Choice Analysis: Theory and Application to Travel Demand”, *MIT Press.*

Bishop, C. M. (2006). “Pattern Recognition and Machine Learning.” New York: *Springer*.

Cascetta, E., & Carteni, A. (2014). “A quality-based approach to public transportation planning: Theory and a case study.” *International Journal of Sustainable Transportation*, 8(1), 84–106.

Chen, HQ., Honda, T., Yang, MC. (2012). “Approaches for identifying consumer preferences for the design of technology products: a case study of residential solar panels.” MIT DSpace.

Chou J-S, Kim C, Kuo Y-C, Ou N-C. 2011. “Deploying effective service strategy in the operations stage of high-speed rail.” *Transportation Research Part E* 47:507–519.

de Ona, J., de O~na, R., Eboli, L., & Mazzulla, G. (2013). “Perceived service quality in bus transit service: A structural equation approach.” *Transport Policy*, 29, 219–226.

Dell’Olio, L., Ibeas, A., & Cec_ın, P. (2010). “Modelling user perception of bus transit quality.” *Transport Policy*, 17(6), 388–397.

Diana, M. (2012). "Measuring the satisfaction of multimodal travelers for local transit services in different urban contexts." *Transportation Research Part A: Policy and Practice*, 46(1), 1–11.

Diaz-Aviles, E., Pinelli, F., Lynch, K., Nabi, Z., Gkoufas, Y., Bouillet, E., Calabrese, F., Coughlan, E., Holland, P., and Salzwedel, J., (2015). "Towards real-time customer experience prediction for telecommunication operators," *in proceeding of IEEE International Conference on Big Data (Big Data)*, pp. 1063–1072.

Eboli, L., & Mazzulla, G. (2007). "Service quality attributes affecting customer satisfaction for bus transit." *Journal of public transportation*, 10 (3), 2.

Eboli, L., & Mazzulla, G. (2011). "A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view." *Transport Policy*, 18(1), 172–181.

Federal Highway Administration (FHWA). (2011). "The Role of FHWA Programs in Livability: State of the Practice Summary."

FHWA PL0159. Community Vision Metrics, (2016) "Methods for Gauging Livability Improvements: Literature and railway".

Ferrell, C.E., Appleyard, B.S., Taecker , M., Allen, C., Armusewicz, C., & Schroder, C. (2016). "Livable Transit Corridors: Methods, Metrics, and Strategies." Transit Cooperative Research Program (TCRP). Research Report No. 187, 142pp. *Published by Transportation Research Board of the National Academies*, Washington, DC.

Geers D. Glenn, Economou. D. (2014). “Computable Livability.” *National Science Foundation Workshop on Big Data and Urban Informatics,* At Chicago.

Governors’ Institute on Community Design (GICD). (2017). “The Why and How of Measuring Access to Opportunity.”

Hensher, D., (1992). “Stated-preference analysis of travel choices: The state of the practice.” *Transportation* 21, 107—133.

Hensher, D.A.(2008) “Empirical approaches to combining revealed and stated preference data: some recent developments with reference to urban mode choice.” *Res. Transp. Econ*. 23, 23–29.

Herzig J. , Feigenblat G., Shmueli-Scheuer M.l, Konopnicki D., Rafaeli A.. (2016) “Predicting customer satisfaction in customer support conversations in social media using affective features” *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, pp. 115—119.

ICF International. FHWA (2011) “Creating Livable Communities.”, Department of Transportation. Washington, D.C.

Ivey, S., Hart, M. (2014) “Making Freight-Centric Communities More Livable: Examining Residential Perceptions and Priorities for Livable Communities”. *Mid-Continent Transportation Research Symposium*. Madison, WI.

Kim S. H., Chung Jin-Hyuk, Park S., Choi. K., (2017) “Analysis of user satisfaction to promote public transportation: A pattern-recognition approach focusing on out-of-vehicle time” *International Journal of Sustainable Transportation*. 11:8, 582-592.

Kim, S., & Chung, J. (2016). “Reinterpretation of the Likert scale for public transportation user satisfaction: Pattern recognition approach.” *Transportation Research Record. Journal of the Transportation Research Board*, 2541, 90–99.

Lima F., Montenegro N., Paraizo R., Kós J. (2017) Urbanmetrics: An Algorithmic-(Para)Metric Methodology for Analysis and Optimization of Urban Configurations. In: Geertman S., Allan A., Pettit C., Stillwell J. (eds) Planning Support Science for Smarter Urban Futures. CUPUM 2017. Lecture Notes in Geoinformation and Cartography. Springer, Cham.

Rapalo, V., Ivey, S., Hart, M., Schroeckenthaler, K., and Adams, T. (2016). “Evaluation of Factors Affecting Livability in a Freight-Centric Community in Memphis, Tennessee.” *Proceedings of the Transportation Research Board 95th Annual Meeting,* Washington, DC.

Sarram G., Ivey S. S., (2017) “Evaluating a Survey of Public Livability Perceptions and Quality-of-Life Indicators: Considering Freight-Traffic Impact.” *ASCE International Conference on Sustainable Infrastructure 2017.*

Sylcott, B., J. Cagan, and G. Tabibnia,. (2011). "Understanding of Emotions and Reasoning During Consumer Tradeoff Between Function and Aesthetics in Product Design," *DETC2011-48173, ASME IDETC - Design Theory and Methodology Conference*, Washington D.C.

Zaki, M., Kandeil, D., Neely, A. & McColl-Kennedy, J. (2016). “The Fallacy of the Net Promoter Score: Customer Loyalty Predictive Model,” *Cambridge Service Alliance.* Cambridge: Cambridge University.

Zheng Yu, Capra Licia, Wolfson Ouri, and Yang Hai. (2014). “Urban computing: Concepts, methodologies, and applications.”, *ACM Trans. Intell. Syst. Technol*. 5, 3, Article 38, 55 pages. https://dl.acm.org/citation.cfm?doid=2648782.2629592.

# Chapter 2:
# Evaluating a Survey of Public Livability Perceptions and Quality-of-Life Indicators

The concept of livability is often used for integrating community quality of life, transportation facility access and neighborhood characteristics while supporting sustainability goals. Quality of life can be very difficult to measure because of the abstract nature of the concept and the varied factors that influence an individual's perceptions. Therefore, the current pilot-scale study examines whether an application of forecasting techniques using machine learning models on data from a previous stakeholder perception survey is possible and will aid in extracting quality of life patterns not only through selecting adequate prediction models but also by providing the most relevant subset of features (indicators) in each related model. This methodology has been used for similar problems in other domains, but has not been applied to livability. The results of this study provide evidence that there is a rule relating neighborhood perceptions to participants' livability scoring systems that can be revealed through machine learning techniques. The results are also consistent with a previous Analytical Hierarchy Process (AHP) approach; however, the current methodology is able to uncover more apparent impact of freight on livability perceptions than was revealed through the previous AHP. Although the pilot results are promising, it is believed that with additional research, larger datasets, and data from multiple settings, more efficient livability indicators can be identified, adopted, and employed for planning purposes.

## 2.1 Introduction and Background

Transportation planning officials are constantly trying to define a set of methods, metrics, and strategies for identifying corridors' livability deficiencies and quality of life improvements. The process of determining the communities' residential and industrial stakeholder's needs, values, priorities and behavior requires collection of stakeholder input, which can be very time consuming and difficult to obtain and even more problematic to analyze. The abstract nature of the concept of livability and the varying ways in which stakeholders perceive and define it can make it difficult to assess and translate beyond community boundaries, and can lead to difficulties in communicating between community stakeholders and transportation agencies.

The U.S. Department of Transportation (DOT), the U.S. Environmental Protection Agency (EPA), and the U.S. Department of Housing and Urban Development (HUD) entered into an interagency "Partnership for Sustainable Communities" in 2009 outlining an approach to improving quality of life in communities through increasing transportation mode choices while reducing transportation costs and supporting the environment through the incorporation of six principals of livability (FHWA, 2011). These principals encompass concepts of the availability of many transportation options, equitable and affordable housing, enhanced economic competiveness, support of existing communities, coordinated federal policies and investment, and an increased value for communities and neighborhoods. This helps transportation agencies prioritize decisions building mobility choice as a part of equalized multimodal transportation networks. This improvement in turn helps to develop sustainable patterns either in an urban, suburban, or rural context. A focus on improving livability via transportation systems leverages the economy of communities, businesses and consumers.

Addressing the related goals of livability can also help integrate planning processes between different agencies and levels of government. Transportation investment requires a decision-making process that includes planning, programming, implementation, and evaluation, accompanied by federal transportation funds. The process requires decision tools that evaluate up to date and immense information on land use, housing, people's mobility, economic development, and many other factors. Meaningful information can be inferred from this data that allows support and testing of demographic and economic theories, and can lead to sustainable development of smart cities.

Meanwhile, urban growth and return of residential focus in urban cores necessitates study on the dynamics of and connections between citizens' mobility and their associated quality of life. However, the state of practice on measuring this relationship is limited through existing survey methods. These methods carry large burdens in both data collection and interpretation. The most prevalent problems are inaccuracies due to sparse or uneven data coverage or missing responses, non-responsiveness challenges and translating participants' qualitative responses. Promoting understanding of residential perceptions of livability requires an approach that provides rich information to create a reliable framework that can infer the residential quality-of-life level of satisfaction through utilizing novel alternative quantitative measurements.

A variety of studies have attempted to address livability assessment. Ferrell and Appleyard, 2016, provided a definition of transit corridor livability and set of methods, metrics, tools and strategies for transit corridor stakeholders. A spreadsheet-based Transit Corridor Livability Calculator tool was developed that presents a proxy indicator of quality of life (QOL). The authors hypothesize that with more livability opportunities in a transit corridor in a sample, and higher metric scores, the QOL would be higher. It is suggested that the more transit, pedestrian, and bicycle trips that both start and end inside the corridor's boundaries, the more livable the corridor.

Meanwhile, Antognelli et al. 2016 confirms that stakeholder involvement is essential for livability assessment. They developed a framework designing a hierarchical classification based on the Common International Classification of Ecosystem Services (CICES) for measuring both ES and US. The approach is used to structure a model based on Saaty's Analytical Hierarchical Process (AHP) to quantify the stakeholder views of the importance of livability services. However, this methodology does not consider multivariate classification analysis for analyzing the livability indicators simultaneously and it does not address freight traffic impact on livability perceptions.

A previous study, on which the current pilot-scale work is based, tested livability in freight centric and non-freight centric communities from a variety of stakeholder perspectives using a mixed-methods approach for a case study in Memphis, Tennessee (Rapalo et al. 2016; Doherty et al. 2013). First, a series of residential stakeholder focus groups were conducted to inform survey instrument design for a residential livability questionnaire. Second, an online survey was developed from this livability questionnaire and was deployed in Memphis, Tennessee to determine factors influencing livability perceptions of neighborhood residents. This survey included 31 items related to demographics, defining livability, perceived barriers to livability, personal commuting patterns, and transportation infrastructure and policy (Rapalo et al. 2016; Doherty et al. 2013). This survey also included a scoring item where residents were asked to rate their neighborhood's livability on a scale of 1-10. The results of this survey led to development of an Analytical Hierarchy Process (AHP) methodology for prioritizing factors influencing livability (Ivey et al. 2014).

With this approach, however, there are challenges of integration, validation and prediction using the current model. Though the responses received for the survey were at relatively low cost, the sample size was limited, and such surveys often have noise and biases that must be considered. Due to this limitation, the livability quantification in the Memphis study was conducted at the neighborhood level. A census-tract approach may reveal heterogeneity that impacts metric scores of the study and was identified as an area of future research. An approach with the capability of increasing translation from metric to policy by identifying the relationship between factors and perceptions of livability based upon different demographic data (e.g. income characterizations), addressing the overlapping nature of the livability principle definitions, and incorporating additional factors (e.g. health indicators) could be more valuable. Also, more research was necessary for generalization of the findings related to the case study in Memphis, TN to other communities.

Considering these works with the shortages on stakeholder involvement, this pilot-scale study is focused on interpreting a linkage between society stated preferences and quantitative measures of livability. The hope is that this approach will result in less need for and reliance on survey-based methods; however, community engagement will always be important to validate the models and to reflect changes in preferences over time. It is anticipated that richer dialogue can be developed between planners and communities through this work by developing greater understanding of influential factors and means for assessing impacts to community livability.

The current pilot-scale study examines whether an application of forecasting techniques using machine learning models on the preceding stakeholder perception survey data is possible and will aid in extracting quality of life patterns not only through selecting adequate prediction models but also by providing the most relevant subset of features (indicators) in each related model. This approach would allow future investigations to center on identifying quantifiable data that can be used as a proxy for indicating community livability, which could lead to a more robust approach that avoids the pitfalls of a purely qualitative assessment. This methodology has been used for similar problems in other domains (Diaz-Aviles et al. 2015), but has not been applied to livability.

### 2.2 Methodology and Data Overview

This research introduces data mining exploration procedures to the domain of livability for discovering neighborhood residents' perceptions and rules based on machine learning methodologies. This technique has been very successful among diverse areas for decades (Wolf et al. 2014) and has recently resulted in scientific investigations of human response data related to opinion mining and sentiment analysis (Clavel et al. 2015, Ravi et al. 2015). Understanding that the livability score decisions made by residential stakeholders can be appraised based on daily life choices and preferences allows consideration of a quantitative metric using a customer service satisfaction framework (or similar, such as predicting customer brand loyalty), to determine which pertinent indicators are important in this decision (Chen et al. 2012).

An open-source Java application, "Waikato Environment for Knowledge Analysis" (Weka) is tested on our survey data that provides an intuitive interface to view and analyze the results generated through preprocessing, learning algorithm employment and predictions of classifiers. A number of algorithms have been chosen and applied for recognizing meaningful information and patterns that may be helpful for livability indicator decision support. Although the adoption of new indicators can be useful for developing a generalized model, this study focuses on the feasibility of the methodology in this setting. Therefore, the success of the technique can be judged by the analysis accuracy on predicting the test data and its consistency with previous work.

Eliciting qualitative information from stakeholder perceptions is crucial and in many decision-making problems, it is not possible to identify metrics with confidence from qualitative data. While different methods attempt to measure relative importance of the objective, this study used the survey data (n=427) from Rapalo, et al. (2016) to attempt to predict respondents' scoring of their neighborhood's livability classification (scored on a rating scale of 1-10) from a set of responses to other items in the survey.

This in turn allows identification of factors most important to determining livability from the viewpoint of a residential stakeholder. These items, termed indicators, must be isolated from the full set of survey responses and determined to have predictive power through variable selection techniques. For this research, only the items related to perceived contributors to livability were incorporated in the final set of models developed. The items considered were related to neighborhood characteristics and transportation system experiences, and relevant items to the models are listed in the Results section and full details are provided in Rapalo, et al., 2016.

The following sections outline considerations necessary for defining algorithms appropriate for the problem considered in this research as well as assumptions and limitations of each.

***Data Preparation and Mining:***

Data preprocessing techniques are designed to clean, impute, standardize, transform and resample the data. This procedure also contains feature extraction and selection for improving data quality. Noise in data may come from nonresponses or missing values. The method selected to deal with noisy data is important as it can influence the results inappropriately if not carefully considered. Missing data for particular survey items was the key challenge in dealing with the dataset for this study. This is not unexpected, as survey data is frequently plagued with incomplete responses (Narayanan et al. 2014).

***Multi-Class Problem:***

Because we have a series of discrete responses to the 'How do you rate your neighborhood for livability?' (scale of 1-10) rating question, this establishes a multi-class problem scenario. Due to the limited data set, it was necessary to collapse these ratings into a smaller set of classifications so that more accurate models could be developed.

Selecting proper classification algorithms for a multi-class problem can be evaluated by performance metrics. Besides accuracy performance measures (Precision, Recall, and F – measure) that reflect correct classification of labels between different classes, a comprehensive visualizing and organizing graph, known as the Receiver Operating Characteristics (ROC) is commonly used to represent binary decision problem results in machine learning (Fawcett et al. 2006; Sokolova et al. 2006; Davis et al. 2006). Therefore, the algorithm selection process in this work was based on accuracy, misclassification rate and related model comparison measurements.

One limitation of the dataset on achieving an accurate classifier was the relative nature of the livability scores. Since respondents from diverse areas with varied backgrounds most likely have different expectations and opinions on how a livability score is defined, it is concluded that there should be a categorization of the livability scores to provide the highest model performance. Therefore, the machine learning classification was initiated by a manual process of grouping different subsets of the livability scores to determine an appropriate classification. As a result, this repeated subset evaluation analysis identified two types of categorization schemes, a two class (Class I) and a three class (Class II) arrangement. Specifically, respectively grouped scores include: {Class I (Medium: 4, 5, 6, 7) and (High: 8, 9 10)} and {Class II (Low: 4, 5), (Medium: 6, 7) and (High: 8, 9 10)}.

Another limitation with this dataset was its uneven multi-class distribution. This type of problem refers to the situation where, when comparing with other classes, some classes are highly underrepresented. For example, the three first classes of 1-3 in the livability score barely had a frequency of 5 between all classes, whereas class 8 had a frequency of 83. Having a skewed distribution of classes can cause less predicting efficiency in machine learning algorithms mainly on low frequency class examples. This problem has been addressed based on the several helpful resources (Finlay et al. 2014; Marujo L. et al. 2013; Chawla et al. 2002; Agrawal et al. 2015).

For this study, the SMOTE filter of Weka was used to insert newly created synthetic instances into the dataset under automatically clarified minority (low frequency) classes. The resampling percentage (200%) is evaluated based on the classifier's performance. Depending on the ratio between minority and majority classes and after several iterations, using different percentages, 100% for class I and 200% for class II provided the best results. After this application the classification accuracy for the learning algorithms steadily improved to more than 70%.

***Training and Testing the Classifier Models:***

Having the predicted classes on each test set, the classifier's performance can be appraised by the error measurements. The error rate represents the proportion of incorrectly forecasted class of instances over total number of instances. Therefore, the resulting N accuracy measurements delivered by the N test datasets will be averaged and introduced as a normal accuracy rate for future test data. In every classification different types of classifiers can be applied.

For this project, 5 numbers of folds were chosen for testing the classifiers performance. Additionally, other numbers of folds were tested (e.g. 10 and 20) for checking the consistency of results. Regarding the selection of appropriate multi-class classifiers, this study has used three of the most well-known and robust models: Multi-Class Support Vector Machine (SVMmulticlass) (Crammer et al. 2001), Random Forest (RF) (Biau et al. 2012), and Logistic Regression (RF) (Omrani et al. 2015). All of these multivariate techniques were chosen based on their robustness reputation, capability of group assessments and multi-class classification analysis (Garcia et al. 2015; Wang et al. 2012). These three methods are combined with attribute selection methods in order to evaluate the relevance of groups of features simultaneously.

## 2.3 Results and Discussion

### *Performance Measurements:*

There are different metrics available for evaluating a classification problem. Accuracy, G-mean, F-measure and Receiver Operating Characteristic (ROC) Area provides informative performance measurement metrics for the case of multi-class imbalance problems that are demonstrated in this section in order to discuss the classifier systems' efficiency. Classification accuracy simply represents the ratio of correct predictions versus all number of predictions that are made during the process but not showing class distinction. The study of a confusion matrix can provide a visualization of correctly versus incorrectly classified responses. This matrix is the most common evaluation method that aids in multi-class problem interpretation and choosing an appropriate evaluation metric; it is also more suitable for balanced class distributions. More detailed discussion of these performance metrics is available from a variety of sources (Sun et al. 2006; Labatut et al. 2008).

Table 1 shows the three selected metrics of accuracy rate, F-measure and ROC curve area for comparing the classification algorithms' efficiency. Despite the fact that logistic regression classifiers were providing reasonable results, because the results are much lower in performance than SVM and Random Forest (RF) algorithms, the results for this classifier are not presented here.

Overall the accuracy and F-measure of both machine learning experiments by SVM and RF are close and competitive in both class types. Therefore, these two measures did not give enough information to support an algorithm selection and there was a need for another measurement called Receiver Operating Characteristic (ROC). This graphical plot is created through illustrating true positive rate (sensitivity) against false positive rate (1-specifity) for all possible classification threshold settings. As the percentage underneath of this plot is simply represented by a single summary number of the Area Under the Curve (AUC) in results (presented as weighted average ROC area in Table 2), we used this measurement to compare the classifiers' performance.

In Table 1 (a), before implementing the SMOTE function, both the accuracy and AUC results are notably close and one might choose the SVM model as the winner of the classification. However, after SMOTE, it is easier to discern the better performing model and it is demonstrated that the accuracy rate is not the most appropriate performance measure for imbalanced problems, and G-mean is more representative of the dominant classifier.

Table 2.1 Performance of classifiers before and after feature selection (Classification I: Medium, High)

| Indicators set | Datasets | Metrics | Classifiers | |
|---|---|---|---|---|
| | | | SVM (Radial Basis) | Random forest |
| Full features | (a) Pre-SMOTE | Accuracy Rate | 64.54% | 63.03% |
| | | Weighted Avg. F-measure | 0.633 | 0.63 |
| | | Weighted Avg. ROC-Area | 0.616 | 0.659 |
| | | G-mean | 0.621 | 0.593 |
| | (b) Post-SMOTE | Accuracy Rate | 73.93% | 75.21% |
| | | Weighted Avg. F-measure | 0.708 | 0.751 |
| | | Weighted Avg. ROC-Area | 0.708 | 0.831 |
| | | G-mean | 0.7 | 0.735 |
| | | | Wrapper | Ranker |
| Reduced features | (c) Post-SMOTE | Accuracy Rate | 74.57% | 73.30% |
| | | Weighted Avg. F-measure | 0.737 | 0.734 |
| | | Weighted Avg. ROC-Area | 0.716 | 0.8 |
| | | G-mean | 0.697 | 0.73 |

Table 2.2 Performance of classifiers before and after feature selection (Classification II: Low, Medium, High)

| Indicators set | Datasets | Metrics | Classifiers | |
|---|---|---|---|---|
| | | | SVM (Radial Basis) | Random forest |
| Full features | (d) Pre-SMOTE | Accuracy Rate | 57.88% | 56.96% |
| | | Weighted Avg. F-measure | 0.468 | 0.528 |
| | | Weighted Avg. ROC-Area | 0.518 | 0.661 |
| | | G-mean | 0.19 | 0.29 |
| | (e) Post-SMOTE | Accuracy Rate | 64.81% | 65.74% |
| | | Weighted Avg. F-measure | 0.603 | 0.641 |
| | | Weighted Avg. ROC-Area | 0.706 | 0.829 |
| | | G-mean | 0.336 | 0.5277 |
| | | | Wrapper | Wrapper |
| Reduced features | (f) Post-SMOTE | Accuracy Rate | 66.43% | 65.27% |
| | | Weighted Avg. F-measure | 0.626 | 0.646 |
| | | Weighted Avg. ROC-Area | 0.72 | 0.809 |
| | | G-mean | 0.4065 | 0.53 |

As we can see in both Table 1 and 2 before synthetic resampling, the classifier correctly predicts the instances at a lower rate; whereas, the performance after implementing SMOTE is higher due to a better distribution of classifications. This distinctive property of ROC curve area and confusion matrix evaluators can be explained as the insensitivity to changes in class distribution. Since other performance metrics such as accuracy and F-measure use values from pair columns of the confusion matrices, they are inherently sensitive to class skew (Fawcett et al. 2006).

Similarly, in all other analyses (Table 1 and 2 (a) through (f)) in a similar manner we can infer that close accuracy rates are not good metrics for performance judgment and the AUC and G-mean values can give a better representation of the confusion matrices. Based on these results, Random Forest performed more reliably after resampling.

***Feature Selection:***

The process of selecting a subset of features (in this case indicators of livability) that simplifies the classification models is called feature selection. The advantages of this process are an easier identification of model patterns, less training time and increased generalization by having less overfitting in the analysis. The related algorithm contains two parts: a search approach and an evaluation measure that scores all possible subsets of the features. Our analysis determined that various feature selection algorithms provided by Weka including the filter and wrapper methods can be selected for this study based on the features' effectiveness algorithm (Janecek et al. 2008). Table 3 presents the Ranker method analysis results for each class type, in respective order from top to bottom. Based on Table 1, the winner performances are introduced as the Random Forest model implemented on the Ranker method and the SVM on the Wrapper chosen subset in order for class I and class II.

However, as we can see in Table 1 (a) and (b), the results show that the feature selection process had a negative impact on the classifier accuracy results. Based on other studies, it is shown that apparently the Random Forest model is sensitive to the dimensionality of the data and by removing some features some part of the information for this ensemble method will be discarded leading to a weaker performance (White et al. 1994). The wrapper method selected subset in Table 3 (c) shows that for achieving higher performance only three low impact features need to be eliminated among 27 features. This means that the SVM results are improved when 24 of the 27 features are use. In reviewing the results table, we can infer that the Ranker method represents a qualified mechanism for prioritizing indicators and the following meaningful information will be helpful towards developing a general approach to extracting livability patterns:

- Table 3 (a) and (b) address seven mutual features of 'Having a park in my neighborhood'; 'Feeling safe in my neighborhood'; 'Having alternative transportation options'; 'Living in an economically thriving neighborhood'; 'Having a sense of community'; 'Minimal road congestion'; 'Experiencing presence of freight or heavy trucks traffic'; and 'Experiencing negative environmental issues' that are all well aligned with expectations and results from the previous AHP (Ivey et al. 2014) as the most prominent livability indicators.

Table 2.3 List of extracted features in final training data set ((a) Classification I, (b) Classification II, (c) Wrapper method (SVM analysis).

| (a) No. | Top ranked Features | Score |
|---|---|---|
| 26 | Experiencing negative environmental issues (smog, air pollution, noise, etc.) | 0.1144 |
| 1 | Having a park in my neighborhood | 0.089 |
| 27 | Taking alternative routes avoiding roadways with rail crossings or high volumes of trucks | 0.0753 |
| 7 | Having alternative transportation options (walk, bike, public transit) | 0.072 |
| 12 | Minimal road congestion | 0.0616 |
| 17 | Experiencing presence of freight or heavy trucks traffic (Response: Never) | 0.0328 |
| 15 | Good bus service | 0.0182 |
| 9 | Having a sense of community | 0.0093 |
| 8 | Living in an economically thriving neighborhood | 0.0031 |
| 6 | Feeling safe in my neighborhood | 0.0002 |
| (b) No. | Top ranked Features | Score |
| 26 | Experiencing negative environmental issues (smog, air pollution, noise, etc.) | 0.1941 |
| 8 | Living in an economically thriving neighborhood | 0.1557 |
| 5 | Knowing my neighbors | 0.1495 |
| 12 | Minimal road congestion | 0.1471 |
| 11 | Quality affordable housing | 0.1252 |
| 1 | Having a park in my neighborhood | 0.115 |
| 24 | How often stucking in traffic due to trains (Response: Occasionally) | 0.0873 |
| 23 | How often stucking in traffic due to trains (Response: Never) | 0.0748 |
| 6 | Feeling safe in my neighborhood | 0.0656 |
| 17 | Experiencing presence of freight or heavy trucks traffic (Response: Never) | 0.0565 |
| 2 | Living close to school/work | 0.0313 |
| 7 | Having alternative transportation options (walk, bike, public transit) | 0.014 |
| (c) No. | Removed features from top subset | |
| 17 | Traffic experiences in the Mid-South region trucks (presence of freight or heavy trucks (Response: Never) | |
| 20 | How often stucking in traffic due to freight presence (Response: Never) | |
| 21 | How often stucking in traffic due to freight presence (Response: Occasionally) | |

- The highest information gain attribute in both lists is 26- ‘Experiencing negative environmental issues’ and including the three other related ranked features of 17, 23, 24 all can be considered as freight impact indicators that have direct or reverse influences on livability classification (Note: classification contains three different levels of Low through High which can be influenced positively/negatively by positive/negative indicators). This reflects the importance of this subject in residents' quality of life and means it should be taken into account as an important priority.

- Although the feature selection productivity depends on the class type division (I or II), the classifier modeling approach (SVM or RF), and survey design efficacy on reflecting respondents' points of view, the Ranker method results are consistent with each other for both classes.
- The results are consistent with previous study results of the AHP technique (4) that derived a weighted hierarchy of similar indicators without considering the interrelationship between the features and livability scores classification. For instance, the important features of 'Knowing your neighbors' and 'Feeling safe in my neighborhood' and 'Living close to school/work' and 'Negative environmental impacts' had the highest AHP weights, which is compatible with the top ranked features from this study. Interestingly, the impact of freight on livability perceptions is more apparent through the current approach than with the AHP.

## 2.4 Conclusion and Future Work

The goal of this research was to investigate the application of forecasting models on quality of life pattern recognition first through understanding stakeholder's priorities and perceptions and second by correctly classifying the livability of residential neighborhoods according to stakeholder scoring. This practical approach can be translated to a large-scale study on the urban livability network domain as well as providing a computable metric and instrument for livability principles and fund prioritization.

A considerable amount of previous work on survey design and data collection based on available and self-constructed livability metric frameworks was available for use in this pilot-scale study. Using Weka software, several machine learning classifiers were utilized to try to predict residents' livability scores based upon developed models. The data utilization was limited at the beginning due to some inaccuracies from missing values and imbalanced class distribution; therefore, there was a need for preprocessing the data using a SMOTE filter in order to increase the lowest frequency class population synthetically. This study evaluated models suitable for this approach including Support Vector Machine (SVM), Random Forest (RF) and Logistic Regression (LR) on two different versions of the data with different class categorizations. For class type I, where respondents' neighborhood livability scores were divided into two ranges, households were classified with 75% accuracy using the Random Forest classifier. For class type II, where livability scores were divided into three ranges, an accuracy of 66% was achieved using both SVM and Random Forest. Although the Logistic Regression classifier passed the chance level for both class types, the results were not accurate enough for use.

The accuracy enhancement following this filtering process provides evidence that there is a rule relating neighborhood perceptions to participants' livability scoring systems. The results are also consistent with previous AHP results; however, the current methodology is able to uncover more apparent impact of freight on livability perceptions than was revealed through the AHP. However, it is believed that with additional research, larger datasets, and data from multiple settings, more efficient livability indicators can be identified, adopted, and employed for planning purposes. The feature selection analysis presented in this study helped in understanding the currently developed metric's strengths and weaknesses and the extracted indicators motivate further research on exploring a more representative and efficient mechanism for developing a systematic indicator set.

This study has demonstrated that prioritizing livability indicators through data mining techniques is plausible and shows promise for developing more robust analysis and informed decision-making. Transportation development and decision-making depends on the associated comparisons for clarifying the differences among various sets of strategies within long-term financial constraints. The evaluation process requires hybrid scenarios from a combination of strategies, indicating the natural interconnected relationship of livability factors and complex outcomes of these decisions.

**References**

Agrawal, A., Viktor, HL., and Paquet, E. (2015), " SCUT: Multi-class Imbalanced Data Classification Using SMOTE and Cluster-based Undersampling.", 7th ACM International Joint Conference on Knowledge Discover (KDIR).

Biau, G. (2012). Analysis of a random forests model. JMLR, 13:1063–1095.

Chawla, NV., Bowyer, KW., Hall, LO., Kegelmeyer, WP. (2002). "SMOTE: synthetic minority over-sampling technique". J Artif Intell Res., 16: 341-378.

Chen, HQ., Honda, T., Yang, MC. (2012). “Approaches for identifying consumer preferences for the design of technology products: a case study of residential solar panels.” MIT DSpace.

Clavel, C., Callejas, Z. (2015). “Sentiment analysis: from opinion mining to human-agent interaction”. IEEE Trans. Affect. Comput. 7, 74-93.

Crammer, K., and Singer, Y. (2001). “On the Algorithmic Implementation of Multi-class SVMs”, Journal of Machine Learning Research (JMLR).

Davis, J., and Goadrich, M. (2006). “The relationship between precision-recall and ROC curves. The 23rd International Conference on Machine Learning, pages 233–240, New York, NY, USA. ACM.

Diaz-Aviles, E., Pinelli, F., Lynch, K., Nabi, Z., Gkoufas, Y., Bouillet, E., and Calabrese, F., (2015). “Towards real-time customer experience prediction for telecommunication operators”. The IEEE International Conference on Big Data, pp. 1063-1072.

Doherty, M., Hart, M., Adams, T., Wise, A., and Ivey, S. (2013) “Defining Livability for Freight-Centric Communities: Identifying Priorities of Residents of the Lamar Avenue Corridor in Memphis, TN.” Proceedings of the ASCE 2nd T&DI Green Streets, Highways and Development Conference, Austin, TX.

Fawcett, T. (2006). “An introduction to ROC analysis”. Journal of Pattern Recognition Letters, special issue on ROC analysis 27, 8., 861–874.

Finlay J., Pears R., Connor A.M. (2014). “Synthetic Minority Over-sampling Technique (SMOTE) for Predicting Software Build Outcomes”. The 26th International

Conference on Software Engineering and Knowledge Engineering, BC, Canada, 546-551.

Garcia, L.P.F., Sáez, J.A., Luengo, J., Lorena, A.C., de Carvalho, A.C.P.L.F., Herrera, F. (2015). “Using the One-vs-One decomposition to improve the performance of class noise filters via an aggregation strategy in multi-class classification problems”. Journal of Knowledge-Based Systems.

ICF International. (2011). “Creating Livable Communities”. FHWA, Department of Transportation. Washington, D.C.

Ivey, S., Hart, M. (2014) “Making Freight-Centric Communities More Livable: Examining Residential Perceptions and Priorities for Livable Communities”. Mid-Continent Transportation Research Symposium. Madison, WI.

Janecek, A., Gansterer, W., Damel, M. and Ecker, G. (2008). “On the Relationship between Feature Selection and Classification Accuracy”. JMLR.

Labatut, V., Cherifi, H. (2008). “Accuracy measures for the comparison of classifiers”. In Proceedings of The 5th International Conference on Information Technology, Las Vegas, NV, USA, 7–9.

Luengo, J., Garca, S., Herrera, F. (2012). “On the choice of the best imputation methods for missing values considering three groups of classification methods”. Journal of Knowledge and Information Systems. 32: 77-108.

Marujo, l., Gershman, A., Carbonell, J., Matos, D., Neto, J., (2013). “Co-Multistage of Multiple Classifiers for Imbalanced Multiclass Learning.” CoRR abs.

Narayanan, S., Jaiswal, A., Chiang, Y., Geng, Y., Knoblock, C. A., and Szekely, P. (2014). "Integration and Automation of Data Preparation and Data Mining". The First Workshop on Data Integration and Application at the IEEE International Conference on Data Mining.

Omrani, H. (2015). "Predicting Travel Mode of Individuals by Machine Learning", Journal of Transportation Research Procedia.

Rapalo, V., Ivey, S., Hart, M., Schroeckenthaler, K., and Adams, T. (2016). "Evaluation of Factors Affecting Livability in a Freight-Centric Community in Memphis, Tennessee." TRB 95th Annual Meeting.

Ravi, K., Ravi, V. (2015). "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowl. Based Syst. 89.C, 14-46.

Sokolova, M.; Japkowicz, N.; and Szpakowicz, S. (2006). "Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation". The AAAI'06 workshop on Evaluation Methods for Machine Learning, 24–29.

Sun, Y., Kamel M, Wang Y. (2006), "Boosting for learning multiple classes with imbalanced class distribution." Proceedings of the 6th IEEE International Conference on Data Mining, USA: IEEE Computer Society, 592-602.

Wang, S., and Yao, X. (2012). "Multi-class Imbalance problems: Analysis and Potential Solutions". The Journal of IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics.

White, A. P., and Liu, W. Z. (1994). "Technical note: bias in information based measures in decision tree induction". Machine Learning, vol. 15, no. 3.

Wolf, J., Bachman, W., Oliveira, M. S., Auld, J., Mohammadian, A. K., Vovsha, P. (2014). “NCHRP: Applying GPS Data to Understand Travel Behavior”, Volume I: Background, Methods, and 120 Tests, No. Project 8-89.

# Chapter 3:

## Evaluating the Potential of Online Customer Data for Augmenting Traditional Transportation Planning Practice

This study investigates the potential of social media review mining for bridging the gap between understanding of neighborhood livability and conventional planning decision-making practices. The novelty of the approach is in interpreting an automated user-defined translation of qualitative measures of livability by evaluating users' satisfaction of the neighborhoods through social media and enhancing the traditional approaches to defining livability planning measures (through a sentiment analysis and visualization package). The provided data-driven neighborhood satisfaction package draws the marketing structure of transportation network products with residential nodes as the center of the structure. In this study, the FastText algorithm is tested to employ a social review analysis, common neighborhood aspects are annotated, neighborhood satisfaction is visualized and compared to the associated demographic side information. This approach has the potential to capitalize on residents' interests in social media outlets and to increase public engagement in the planning process by encouraging users to participate in online neighborhood satisfaction reporting. This study demonstrates a proof of concept for a social review analysis and specifies the methodology for agencies to augment planning practice.

### 3.1 Introduction

Planning agencies struggle with achieving public involvement goals to appropriately inform decision making processes. The topic of livability creates even greater difficulty in this regard in that perceptions of livability are abstract and difficult to quantitatively assess. Understanding patterns of public perceptions related to livability can provide richer information for planning decisions. By first understanding public quality of life perceptions and expectations, we can design performance measures and project choices for effective and efficient investments. Incorporating state-of-the-art performance measures requires defining measurable livability metrics reflecting stakeholders' needs, wants, and behaviors. Much of the work surrounding livability to date has focused on residential surveys and traditional measures of congestion or other attributes to indicate livability. The connections between livability perceptions and consumer preference theory/demand modeling have only recently entered transportation literature (Sarram and Ivey, 2018). Integrating such approaches from other domains in transportation settings, particularly in the case of livability studies, can provide greater insight and can lead to more informed decision-making.

One popular approach in consumer product demand modeling for guiding product design decisions is Discrete Choice Analysis (DCA), adopted from market research and transportation planning (Ben-Akiva and Larman, 1985). Through establishing the relationships between key customer attributes (product features, brand, price and warranty) and engineering attributes (a function of the design options), richer information is integrated into the decision-making process. The application of DCA for demand estimation is known for addressing unobserved taste variations, unobserved attributes, and modeling deficiencies. The approach considers the preference structure of customers through identifying the key customer attributes for the Kano (a widely-used consumer preference theory model) selected customer utility function of the DCA demand analysis (Wassennar et al. 2005, Michalek et al. 2005). The manufacturing and the design domains have also developed frameworks of data-driven analytic processes for decision makers to integrate traditional data sources with subjective and objective impact factors (Chien et al. 2016, Lin et al. 2016).

Meanwhile, the artificial intelligence (AI) domain has provided a simulated framework of human cognitive behavior that has rarely been studied in the transportation planning field. One recommended practice for extracting valuable customer experience and preference information from large market and product data is data mining. The approach assists the quantification of some parameters in design models. Tucker and Kim (2008 and 2011) first implemented data mining techniques merged with multilevel optimization to predict the maximum company profit for a particular product architecture, and later, predicted the trend of customer preferences for generating a demand model that reflects changes in the preferences. In the case of mass customization, the technique has been used to balance the heterogeneity of customer preferences and the complexity of products. Clustering analysis and association rule mining have been applied widely to learn the association between customer features and product requirements (Agard and Kusiak, 2004, Morency et al., 2007, Jiao and Zhang, 2005).

Investigators have extended the study to the domain of product configuration (Shao et al., 2006, Song and Kusiak, 2009, Moon et al., 2010), however, the synchrony of transportation planning decisions with public preferences and expectations has been overlooked. Due to the competitiveness of the market, the manufacturing domain for new products has developed beyond use of limited survey information and now includes input from technology and retailer websites (e.g. Amazon, Google and Yelp) through applying text or opinion mining to online reviews. Li et al. 2015, performed association rule mining to analyze customer opinions in online reviews and to investigate the product attributes that are desired by specific user types. Zhang and Narayanan 2010, presented a feature-based product ranking technique instead of only using the existing automated ranking mechanisms for overall product quality. They identified features within a product category and analyzed their frequencies based on the subjective and comparative sentences in reviews and modeled the relationships between the products. From another perspective, Proserpio et al. 2017, studied user interaction in the sharing economy and market demand incorporating reciprocity. The study evaluates the tendency to increase effort in response to other's increased effort based on online review rating. They tested data from Airbnb and proved that reciprocity affects market price equilibrium through impacting ratings and changing expected demand of listings. These approaches have not yet been applied in the urban planning domain for informing neighborhood planning decisions. However, there is significant potential to improve public involvement outcomes through attracting residents to provide data in a different and more efficient way that reflects individual choice preferences. The current study focuses on evaluating the potential for using text mining and existing online platforms for informing transportation planning practice.

## 3.2 Literature Review and Research Plan

### *Approaches for Understanding Quality of Life:*

As discussed in Sarram and Ivey 2018, there is limited academic literature in urban planning on developing rich understanding of livability needs and community quality of life. The current studies mostly represent aggregated opportunity proxies or hierarchical classification processes. These methods are limited for quantifying residential satisfaction and identifying the potential alternatives for improving communities based on users' preferences and differentiated levels of satisfaction.

From an urban economic standpoint, Albouy and Lue, 2015, estimated a behavioral model to measure willingness-to-pay to access local amenities based on combined costs, rent, wage, and commuting differentials for inferring changes in quality of life. Household preference is defined as a utility function based on aggregated amenities in a quality of life index. The model assumes each mobile and informed household chooses place-of-residence and place-of-work combinations (j,k) with similar satisfaction. Therefore, with homogeneous households, all observed combinations (j,k) must drive the same level of utility. The overall study results show that neighborhood quality within metro areas varies substantially due to the artificial amenities created through use of a homogenous approach. Artificial amenities arise from a homogenous approach to assessing 'willingness to pay' for a set of determinants, where restricted feasible options result in the appearance of choice (for example because of higher rent, a person lives further away from desired amenities, making it artificially appear they are willing to accept a longer commute to live in an outlying area). The limitation inherent in this approach may be addressed through considering customer satisfaction patterns in the case of transportation policy considerations, where more information is required about individual preferences.

Similarly, Couture and Handbury, 2017, have been conducting a within-city analysis on understanding urban revival using 10 years of data from the Consumer Expenditure Survey (CEX) and the National Household Travel Survey (NHTS). The applied location choice model used residential-only data due to the limitation on the age and education group for the workplace location data. The authors discussed the importance of various explanations for the location decisions of the young and college-educated. The results show that varied preferences of these groups for non-tradable services (restaurants, bars, gyms, and personal services) in addition to their travel and expenditure shares can explain the diverging location decisions. The decisions depend on the presence of other young college graduates due to the higher demand of networking and socializing. Considering these results, it can be inferred that the nature of satisfaction is interconnected with the neighborhood quality. This means that the demand for social activities, depending on the individual priorities and preferences, affects residential location choices. As people's choices regarding residential locations are made by best satisfying their priorities rather than finding a perfect place to live, it is necessary to invest in and plan neighborhoods efficiently and thoughtfully through deep understanding of related factors.

In 2012, the Nashville Metropolitan Planning Organization (MPO) conducted a regional household travel survey that included a health component (Nashville Area MPO, 2013). The travel behavior and general health of 6000 households were surveyed including a subset of 1000 households recruited into a GPS and health component of the study. The study compared the summary statistics in a fine geographic resolution, used health status data to identify health priority areas instead of using common demographic approaches alone, and modeled the potential health impacts of active transportation in the region. The results illustrated the relationship between transportation, physical activity and overall health. The study goal was to gather baseline data for the MPO to quantify the effect of health and safety criteria in the evaluation and ranking of future transportation projects on the overall health of Nashville residents. The study results have been used in policy approaches for enhancing the livability of Tennessee communities. Later, Meehan and Whitfield, 2017, described the multifaceted approach and projects that the Nashville MPO has undertaken to improve the transportation system, quality of life and health through using the aforementioned dataset. The study discussed that the MPO has impacted public and environmental health through the objective of providing mobility for workers and residents. The model results were subject to several limitations such as over-valued benefits in behavioral change and prevention of chronic diseases. However, the study efforts were focused mostly on the transportation effects on public health which overlooks other lifestyle factors (e.g. work stress). While the study is a demonstration of the interplay between the measure of public health and providing mobility for increasing community livability, it only concentrated on analyzing one component of livability (health). A more comprehensive correlation between residential travel behavior and quality of life components is neglected here.

***Assessing a Customer Centric Framework:***

A primary understanding of the association between the mobility impedances and users' stated satisfaction along with correlation to neighborhood attributes can identify the variables that are more important in order to directly measure neighborhood livability. Therefore, in planning practices, there is a need for a joint metric that considers the trade-off between community livability and network mobility. Caltrans has developed a mobility planning guide, Smart Mobility 2010. This report has structured a set of network mobility performance indicators around six smart mobility issues of location efficiency, reliable mobility, health and safety, environmental stewardship, social equity, and robust economy. These principles have been addressed through major measurable components of Vehicle Miles of Travel (VMT) for emissions reduction; travel times and costs per trip by mode for multi-modal travel mobility, and connectivity; volume/capacity for congestion judgment; multi-modal Level of Service (LOS) which represents system service quality (on-time performance); Average Vehicle Occupancy (AVO); road condition; highway accident rates; park and ride lot utilization; and bicycle and pedestrian facilities, safety and interaction conflicts. This guide served as a basis for the framework developed for the current research, and the adopted mobility metrics are described later in the methodology section. For a long time, user satisfaction surveys in transportation planning have considered satisfaction intervals and cardinal scale applications; however, the complexity of users' innate perceptions has been disregarded.

In the marketing domain, Kamakura and Russel, 1989, applied a flexible choice model as a unified description of market structure that links the pattern of brand switching to the magnitudes of price elasticities. The authors characterized each segment by a vector of mean preferences and a single price sensitivity parameter. Also, Mittal and Kamakura, 2001, presented a conceptual model capturing the relationship among satisfaction ratings, latent satisfaction, repurchase behavior and consumer characteristics. The results showed the relation is highly non-linear and also presents that consumers at the same level of rated satisfaction have different repurchase rates due to the different characteristics they have. The study implies that customer segments with higher intrinsic loyalty coefficients represent a consistent customer base. Attribute-performance management is also recommended in order to optimize overall satisfaction. The approach maximizes performance based on attributes that have the largest weight in determining overall satisfaction and considers each customer segment individually.

In transportation studies, initially Ben-Akiva and Boccara, 1995, developed a framework of choice set generation using a special case of finite mixture as discrete choice models with latent class to evaluate the influence of attitudes and perceptions on the choice set generation process. The estimation approach jointly incorporated information on the individual's perceived choice set from alternative availability questions and the revealed preference based on the observed choice. Later, Herrmann et al. 2002, proposed a modified finite-mixture-model to capture unobserved customer heterogeneity based on partial least squares. In transportation behavioral data analysis, Lo 2013, used Gaussian mixture models (GMM) to classify users' driving behavior into multiclass traffic flow. In transit planning, Kim and Chung 2016, applied the GMM approach to public transportation data to address the heterogeneity of the human complex cognitive system and to evaluate the Likert scale accuracy on this matter. The study presented new implications in analyzing public transportation satisfaction surveys.

***A New Approach Enabled by Social Media:***

The increasing digital social connectiveness of our society provides an opportunity to connect these social ties on the geospatial level and improve consideration of human dynamics on quality of life. The analysis behind such an approach has to consider not only the decision sets of transportation planning but also user preferences. An approach is needed that is able to: 1) reveal the users' preferences, from online review information to direct inquiries of their requirements, 2) interpret them as a complex, through comparing the results with similar residents' satisfactions and 3) integrate the performance of the transportation network surrounding the neighborhood complexes. This type of information about how users make decisions and the ultimate criteria considered in decision-making is very important for planning agencies to deeply understand livability considerations through a customer-centered approach.

Machine learning models maximize predictive accuracy on unseen data, rather than observed data which is used for training. They are mostly non-linear, have different bias-variance tradeoffs and regularization mechanisms capable of controlling model complexity (Kleinberg et al. 2015; Paredes et al. 2017). Covariates may also be simultaneously selected with model derivation. These techniques outperform traditional econometric models on prediction; however, this comes at the expense of interpretability. Since discrete choice models place more emphasis on using features that are easily mapped to human behavior and well understood causal relationships and correlations, this study aims to exploit the power of pattern recognition models in prediction accuracy and appraising the potential of interpreting the causal relationship among the livability features through an interactive system. As a breakthrough, users are not required to fill out surveys for manifesting their preferences. Although, the result could also be adopted for use as a pre-survey method to determine expectations, gaps, and services that need to be targeted in actual survey and action planning.

While there is a large body of work related to preferences of residents extracted based on their activities, this is the first to develop a review analysis package based on the users stated preferences in the transportation planning domain. This framework provides an efficient online text classification and visualization approach for automating subjective performance measures used in planning. Developing this data-driven neighborhood satisfaction package draws the marketing structure of transportation network products with residential nodes as the center of the structure.

Hu et al. 2019, conducted a semantic and sentiment analysis on online neighborhood reviews. They compared latent Dirichlet allocation (LDA) and multi-grain LDA models to identify the semantic topics (aspect) in Niche neighborhood reviews dataset. The authors extracted the probabilistic distribution of topics, for instance safety, and modeled these as a probabilistic distribution of words, such as safe, crime, and police. In order to quantify the sentiments of the identified aspects, they proposed to leverage an aspect rating analysis (LARA) model to derive aspect-specific ratings from online neighborhood reviews. The purpose has been resolving the requirement for deriving sentiment lexicons from extreme opinions in a certain domain. At the end, they corelated the subjective perceptions extracted from the dataset and the objective socioeconomic attributes of NYC neighborhoods. However, the models have their own assumptions and limitations of LDA ignoring the order of words in texts, and LARA assuming that the overall rating of a reviewer follows a Gaussian distribution with its mean as the weighted sum of the aspect-specific ratings. The correlation results also did not sufficiently reflect the study’s objective measures.

Meanwhile, estimating neighborhood satisfaction of individual users based on their reviews on Twitter platform has been the focus of limited research works: Saeidi et al. 2017 investigated whether sentiment analysis can be used to predict demographic attributes for neighborhoods of London. The correlation results between the two datasets, Yahoo! Answers and Twitter, shows differences in the nature of information. The former contained more encyclopedic information, while the later provided current sociocultural aspects. In another multimedia study, Flaes et al. 2016 analyzed both visual and text content posted on Twitter and Flickr (a multimodal content analysis on multimedia data). The result from microblogging platform Twitter (64 thousand geolocated tweets within a 10-mile radius of the city center of Amsterdam) appeared more suitable for sentiment analysis than the content-sharing website Flicker (64 thousand images taken in and around the city of Amsterdam). The sentiment scores detected from Twitter show significant relationships with the demographics of inhabitants, ethnic background, income and education level. For both platforms the economic indicators (i.e. people living on welfare checks and income) show significant relationship with the sentiment scores, however no relationship with the livability or social index of a neighborhood. The authors, however, have not yet identified the adjustment factors to the sentiment scores that are important to the residents of a city. Twitter microblogs have also been used to discover the correlation (0:35) between collective sentiments of tweets and the deprivation index of the involved communities (Quercia et al. 2012).

Gibbons et al. 2019 evaluated Twitter sentiment analysis to predict self-rated local health data from the American Community Survey. They computed the hedonometer grand mean score for each census tract based on the health sentiment lexicons from previous health studies and used a multivariate generalized linear model to identify how neighborhood attributes like aggregated sentiment affect population-level screening behaviors. The major deficiency with lexicon-based sentiment analysis metrics such as Hedonometer is the inability to properly take context into account (Pang and Lee, 2008). The study result shows some limitations in profanity usage (as an indicator of mood) and local inconsistencies due to the aggregation in a smaller geographic scale.

Zivanovic et al. 2018, collected four years of tweets in the area of Bristol between January 2012 to September 2016. The study explored the use of Twitter data in quality-of-life research and the approach was a semi-manual content analysis using computer-assisted qualitative data analysis and GIS software. The study concluded that Twitter data may be useful to indicate the emergence of concerns not identified by traditional quality of life surveys and complement them. On the other hand, the study did not find the Twitter data representative enough and discussed limitations with using it (e.g. migration and demographic bias) that may render invisible certain segments of the population.

However despite these limitations, the current study will evaluate the use of Twitter datasets rather than Niche, Nextdoor, Yahoo! Answers and other sources of neighborhood reviews due to the following reasons: first, abundant available sources of training datasets that are labeled and have shown reasonable performances in state-of-the-art NLP research, and second, open access to the trending real-time information. These features make Twitter data more accessible and potentially more valuable for use in transportation planning applications for public agencies.

Thus, the current research implements insights from three disciplines, transportation planning, marketing and Natural Language Processing (NLP), to design an innovative data-driven and efficiency-based planning sentiment analysis (SA) package to present how neighborhood satisfaction can estimate residential quality-of-life, and identify users' preferences and priorities.

### 3.3 Methodology

The focus of this research is to:

- Evaluate the potential of Twitter datasets for providing insight useful for transportation planning related to livability assessment in the US and
- Develop a review analysis package based on users' stated preferences in the transportation planning domain.

The intent is for this framework to provide an efficient online text classification and visualization approach for automating extraction of user-defined performance measures that can be used in planning decision-making. Developing this data-driven neighborhood satisfaction package draws the marketing structure of transportation network products with residential nodes as the center of the structure.

Recent NLP works, including the winners of SemEval Aspect-Based Sentiment Analysis (ABSA)[1] tasks (Baziotis et al. 2017, Cliché 2017), have been using deep learning methods such as Concurrent Neural Network (CNN), and Recurrent Neural Network (RNN). However, the problems with using ABSA-DL methods are first, the ABSA methods need the availability of domain specific human annotated training dataset, including the identified topics of the text (called the Aspect) and second, the DL algorithms can be very expensive timewise. Despite having great performances in laboratory practices, deep learning models are very slow to train and test on real-time and large-scale datasets. For practicality in transportation planning applications, the current study evaluates a fast and efficient language processing approach (including the dataset sources and sentiment analysis approach).

To this aim, neighborhood reviews from the Twitter platform in Memphis, San Francisco and New York City (NYC) were appraised to determine if the social media outlet can be used to indicate public satisfaction of neighborhoods and determine the key user-defined performance measures related to this subjective metric. These three case studies were selected to provide a range of city characteristics in terms of population size, geographic scale, and community type.  The methodology for this study included:

---

[1] ABSA refers to determining the opinions expressed on different features of entities, e.g., of a transit service, or a neighborhood (Hu et al. 2004). An aspect is an attribute of an entity, e.g. the timing of a metro, or the safety of a neighborhood. This level of analysis incurs an additional annotation task to achieve a domain-related training dataset that is also labeled regarding polarity of the aspect (Cataldi et al. 2013).

- Extraction of neighborhood-relevant data from Twitter for three cities in the US over a six months period
- Testing of a more efficient language processing approach (using fastText) for the planning domain
- Applying the fastText model to develop sentiment profiles for neighborhoods included in case study cities
- Developing sentiment maps to 1) discover existence of possible satisfaction patterns in each city, 2) compare the public expectation and online expression cultures in these three case studies, and 3) examine relationships to Census tract demographic profiles.
- Evaluating the potential of Twitter data for informing transportation planning decisions related to livability.

### 3.3.1 Data Collection

For the current study, the Twitter API platform was used to extract data for three cities to examine its potential for use in neighborhood satisfaction analysis by transportation planners. The approach to extracting tweets and the selected training dataset are described in the following sections.

***Neighborhood Tweets:***

The Twitter API platform offers three tiers of search APIs: standard, premium, and enterprise, which provide free and paid access to either the last 30 days of the Tweets or access to Tweets as early as 2006 with full data fidelity. However, the private company shuts searches down in cases of detecting individuals running the queries and overloading the system. According to the Twitter guidance, the analysts and developers are allowed to accommodate different searches of 1,500 tweets and Twitter does not archive the searches for more than three to seven days' query on the API, depending on their numbers (Twitter, 2018). Text from tweets with neighborhood topics were collected during a roughly 6 months period, between May 2019 to November 2019, during this study to obtain a representative number of data samples. As collected Tweets span a 10-day time period, the inquiries were continuously submitted every 10 days.

The goal of this data collection activity was to extract a variety of neighborhood characteristics following a word choice strategy. Therefore, the neighborhood Tweets were queried from Twitter public API[2] using the neighborhood hashtag and first-person pronouns (me/my). This experimental approach of extracting the data was identified in order to reduce the number of irrelevant Tweets and real estate advertisements. Entities in our text datasets are locations and aspects are the topics of Tweets. The TwitterR package was used for the open source software R (Schweitzer, 2014).

***Training Dataset:***

According to Bojanowski et al. 2017, fastText exploits character-level similarities between words, therefore the model is supposed to have a good performance modeling

[2] https://developer.twitter.com/en/use-cases/academic-researchers

infrequent words or even using restricted size datasets. However, in order to reach an accurate result the current study tested fastText with different training datasets including: 1) SemEval-2017 Task 4 [3] (subtask A-sentiment analysis in Twitter) accumulated datasets, which was developed from around 60,000 tweets collected since 2013. This task also provides topic-based labeled datasets (subtask B-C) but the topics were not applicable for the current research; and 2) Sentiment 140[4], developed from a collection of tweets using keyword searches by Stanford University (Go et al. 2009).

The Sentiment 140 dataset has been used in other NLP research and was created automatically as opposed to manually annotated tweets (Rosenthal et al. 2017). The approach assumed that any tweet with positive emoticons were positive and vice versa. Among tested datasets for the current study, fastText models were able to provide accurate results only using the large training dataset of Sentiment 140.

### 3.3.2 Shallow Neural Network

This study implemented fastText, an NLP library for learning of word embeddings and text classification created by Facebook AI Research (FAIR) lab. As per Joulin et al. 2017, fastText classifier is often on par with deep learning classifiers in terms of accuracy while being faster in many orders of magnitude. The model allows creation of an unsupervised or supervised algorithm that relies on the Continuous Bag of Words

[3] http://alt.qcri.org/semeval2017/task4/

[4] http://help.sentiment140.com/for-students

(CBOW), for vector representations of words and a hierarchical classifier accelerating training. CBOW is a shallow neural network that predicts a category after being trained in a single-layer model. Moreover, the softmax layer in deep learning processes is replaced with a hierarchical softmax and each node represents a label. Therefore, the model only predicts the probabilities for a limited number of labels and parameters which reduces the training time. FastText also considers the imbalanced nature of the classes through using Huffman algorithm to represent categories by building the category tree. The depth in the tree is smaller and leads to further computational efficiency without sacrificing accuracy (Bojanowski et al. 2017).

This open-source library also works on generic standard hardware and is able to fit on portable devices. It requires remarkably lower training time and has better scalability than deep learning models. This makes an efficient tool to build NLP models for the classification of neighborhood reviews and generating real-time predictions. To enable the application of these models in the planning decision-making process, a simple pre-processing step is introduced, as described in the following section.

***Preprocessing, Reverse Geocoding and Spatial Aggregation:***

Sentiment 140 contains 1,600,000 tweets in three categories of Negative, Neutral and Positive. However, fastText is able to only open around 1,048,576 in csv format. Due to its binary prediction design regarding the sentiment categories, Neutral groups were removed from the Sentiment 140 dataset for the current study (discussed more in the result and discussion section). Approximately 10% of the dataset was used as a validation set to compare the accuracy of the trained model. Therefore, in total around 948,576 samples were used as training and 100,000 tweets as validation sets.

FastText removes hashtags, URLs, HTMLs, punctuations, numbers, emojis and special characters as they are not included in the sentiment analysis; converts the text into lowercase; and applies the same cleaning process on the validation dataset as well. A manual inspection revealed that both training and validation datasets were evenly balanced between positive and negative labels, thus the upsampling function was not required.

FastText is able to produce vectors for any words, even made up words. These word vectors are built from vectors of substrings of characters contained in it. Therefore, it allows recognition even for misspelled words or concatenation of words[5]. During the training process, the Tweets will be converted to tokens (split text into pieces) which will be used as an input to the model[6].

***Location Categorization and Spatial Unit of Analysis:***

Sometimes users discuss several characteristics of different neighborhoods in one Tweet, therefore the same Tweet was assigned for every mentioned neighborhood. Each of the datasets from the case study cities was manually reviewed to examine how residents were indicating neighborhood location in the Tweets. In two of the case studies, Memphis and San Francisco, neighborhood reviews are more precisely articulated and addressed the

[5] https://fasttext.cc/docs/en/faqs.html

[6] https://fasttext.cc/docs/en/supervised-tutorial.html

neighborhood names (NNs) directly. A manual aspect annotation process was applied only to meaningful samples of data from Memphis and San Francisco because the samples contained non-geocoded tweets that mentioned the neighborhood names. This phase was performed based on the SemEval 2016 task 5 (ABSA-16) annotation guideline. The annotator was familiar with the problem definition and the task.

While the appropriate focus of analysis for planning applications in a livability context may be the neighborhood unit, this study collects neighborhood-related Tweets and evaluates the results at the zipcode level, due to the availability of census demographic data at this level. This is intentional since the current study is focused on evaluation of Twitter data for transportation planning applications, and it is important to assess demographic trends found in the study datasets.

***Aspects:***

Based on ABSA-16 annotation guidelines (SemEval 2016, Task 5)[7], the annotation process included three steps:

1) Identifying the entity E and the attribute A pair (E#A) towards which an opinion is expressed. An initial inventory of entity types (e.g. safety, mobility, environmental) was defined and lists of their related attribute labels (e.g. crime, bikeability, park) are extracted. These entity categories are defined based on livability performance measures from Sarram and Ivey, 2017. For the purpose of

[7] http://alt.qcri.org/semeval2016/task5/data/uploads/absa2016_annotationguidelines.pdf

the current study, the Sarram and Ivey (2017) livability performance measures were categorized into five types: environmental, socioeconomic, safety, mobility and amenity, as shown in Table 3.1. The term ‘amenity’ was defined based upon amenity types defined by Couture and Handbury, 2017.

2) Assigning polarity to each E#A pair of a sentence, from a modified set of P={positive, negative}. The neutral label has not been considered in the current study due to the increased performance of the fastText classifier (discussed more in the result and discussion section).

3) Extracting the opinion target expression (OTE) as an explicit reference (mention) to the reviewed entity E of the E#A pair. OTE can be a named entity, a common noun or a multi-word term. It is uniquely identified by its starting and ending offsets.

The aim was to evaluate if the social media review data are applicable for planning project decision-making purposes while comparing the results with the extracted features from the previous study (Sarram and Ivey, 2017), integrating the entity types list, and developing their paired attribute labels for future studies.

***Code Availability:***

Jupyter notebooks for fastText training, validation and testing, as well as descriptive and sentiment visualization codes are customized in one package towards developing a data-driven planning decision-making tool. To view this package, visit this Github page, NextUrban.

Table 3.1. Summary list of extracted features in final training data set (Sarram and Ivey, 2017)

| **Priority** | **Top Ranked Features in Order** | **Performance measure** |
|---|---|---|
| 1 | Experiencing negative environmental issues | Environmental |
| 2 | Living in an economically thriving neighborhood | Socioeconomic |
| 3 | Knowing my neighbors | Safety |
| 4 | Minimal road congestion | Mobility |
| 5 | Quality affordable housing | Socioeconomic |
| 6 | Having a park in my neighborhood | Amenity |
| 7 | How often stuck in traffic due to trains | Mobility |
| 8 | Feeling safe in my neighborhood | Safety |
| 9 | Experiencing presence of freight or heavy trucks traffic | Mobility |
| 10 | Living close to school/work | Mobility |
| 11 | Having alternative transportation options | Mobility |

## 3.4 Result and Discussion

Twitter data were collected for 6 months and the total number of relevant Tweets extracted for each case study area was: Memphis, X; San Francisco, X, and NYC, X. However, for the data to be useful for livability applications it is necessary that Tweets be associated with a specific geolocation. When just the geocoded Tweets were extracted, the total number remaining in each case study was: Memphis, 110 ; San Francisco 650; NYC 1800. The aggregated datasets were used separately as a test set during the fastText training. The stored spreadsheet in each case study includes several types of user information, retweets, place names, latitudes and longitudes. Note that even though a substantial number of meaningful tweets were collected in each case study, only the geocoded tweets could be considered as applicable data for visualization and decision-making purposes. Therefore, the focus of current study was on using the location-based information and evaluating its potential for use in transportation planning studies, where agencies could access much larger datasets using paid options within the Twitter API. Because of the relative size of the case study datasets, only that for NYC was used in visualization evaluations. The other two case studies were used to inform the aspect extraction.

The histogram of Figure 1 visualizes the relative frequencies of Tweets in NYC neighborhoods. In all case studies the majority of neighborhoods have less than 100 related Tweets. Saeidi et. al 2017 reported similar results in London, UK. This suggests that there is a lack of available direct residential expression in random online neighborhood reviews, indicating outreach activities will be needed by planning agencies to encourage residents to use Twitter to express their neighborhood satisfaction**.** Potential strategies for addressing this limitation will be further discussed in next section.

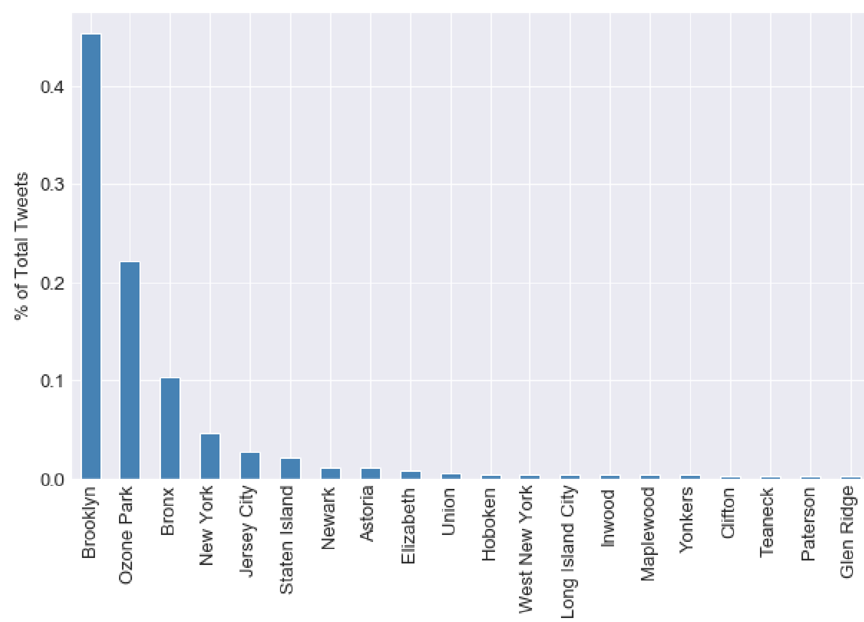


*Figure 3.2 Histogram of percentage of Tweets per zipcodes in New York*

***Aspect Category Annotation:***

During the training procedure, it was noticed that the majority of neighborhood Tweets are either critical or complementary and a neutral opinion was infrequently observed. This observation triggered the idea of reconsidering the SemEval 2016 annotation guideline and removing the opinion polarity of Neutral from the training labels. This helped notably in developing more accurate fastText models and to provide more accurate sentiment classification results. A similar conclusion was reached by Shah et al. 2018, who determined that varying characteristics of datasets will require adjustment of annotation guidelines.

A summary of the results of aspect category annotation for meaningful neighborhood Tweets from San Francisco areas are presented in Table 2. An example set of Tweets is also provided in Table 3. A review of the aspect entities and attributes for the total dataset indicates:

1. Besides the *Crime* and *Crash* attributes in San Francisco, *Verbal Assault* is a third mentioned **Safety** concern.
2. *Bikeability* and *Transit* are indicated besides *Walkability* in the neighborhood **Mobility** aspect category of the San Francisco area.
3. *Affordability* is not mentioned frequently as a **Socioeconomic** concern in the Memphis area.
4. In San Francisco, *Parking availability* is Tweeted as a trending Traffic issue effecting neighborhood **Mobility**, however in Memphis the effect of a public place and related *Events* on the Traffic and the neighborhood quality are more important.

5. *Ambiance* in both cities is considered as a key **Environmental** attribute (e.g. having a park, great view) while *Noise* and *Sustainability* are reported in Memphis and San Francisco respectively.
6. Having a supportive and active neighborhood *Community* are addressed as important **Cultural** vibes while *Racism* is also reflected in the language of both cities' Tweets.
7. Proxy to the neighborhood **amenities** (e.g. restaurants and groceries) was important in the residential satisfaction.

Table 3.2. Summary of extracted neighborhood Entities and Attributes

| **Entity** | **Attributes** | | | | | |
|---|---|---|---|---|---|---|
| | **Memphis** | | **San Francisco** | | | |
| **Safety** | *Crash* | *Crime* | *Crash* | *Crime* | *Verbal Assault* | |
| **Mobility** | *Walkability* | *Traffic* | *Bikeablity* | *Walkability* | *Transit* | *Traffic* |
| **Cultural** | *Community* | *Racism* | *Community* | *Racism* | | |
| **Environmental** | *Ambiance* | *Noise* | *Ambiance* | *Sustainability* | | |
| **Amenities** | *Quality* | | *Quality* | | | |
| **Socioeconomic** | | | *Affordability* | | | |

The aspect-based results are compatible with the extracted features from Sarram and Ivey (2017), while adding a new entity of neighborhood **Culture or lifestyle.** This new neighborhood attribute has not been previously suggested by the neighborhood livability survey results which could also relate to the socioeconomic characteristics of a neighborhood as well.

Table 3.3. Example Tweets from San Francisco with annotated aspect categories. Subjective phrases are italicized. Topics are **bold** and aspect phrases are ***bold italic***.

| **Tweets** | **{Entity#Attribute, "*OTE*", Polarity}** |
|---|---|
| Another Mission restaurant closes due to ***skyrocketing rents***. Another sad ***loss*** for my neighborhood.... | **1. {**SOCIOECONOMIC#AFFORDABILITY, "***rents***", negative} |
| Just spent a speedy lunchtime in the **Castro** — my old neighborhood! — and am reminded of how this place has the best ***sunshine***, the best ***movie palace***, and the best ***vibe*** around. How lucky I was to call this ***place*** my last stop every day | **1.** {AMENITIES#QUALITY, "***place***", positive} **2.** {ENVIRONMENTAL#AMBIANCE, "***vibe***", positive} |
| This is about 3 blocks from where I live and right by where my wife walks every weekday. SF has been incredibly tone deaf on the ***safety*** concerns of the residents of this neighborhood. I can't understood how the judge released this man after he ***violently assaults*** this woman. | **1.** {SAFETY# VERBALASSAULT, "***assaults***", negative} |
| Big news for **Oceanview**. My food desert of a neighborhood finally opened a ***coffee shop***. No longer ***walking*** two miles for Arco gas station coffee. @ Blue House Café | **1.** {AMENITIES#QUALITY, "***coffee shop***", positive} **2.** {MOBILITY#WALKABILITY, "***walking***", positive} |
| **Neptune Gardens** is a neighborhood on the West end of **Alameda** that combines the accessibility of ***shopping*** and ***restaurants*** with the advantages of being largely ***residential.*** | **1.** {MOBILITY#WALKABILITY, "***accessibility***", positive} **2.**{CULTURAL#COMMUNITY, "***residential***", positive} |
| @JesseArreguin @DavidChiu @davidying @WatsonLadd My former ***affordable*** housing boss: "I live in **NB** and the neighborhood has already changed. The ***diversity*** is disappearing. Prioritize housing over parking. More ***bikes. Shuttles*** and ***buses***. We need 100% affordable" #berkmtg | **1.** {MOBILITY#BIKEABLITY, "**bikes**", negative} **2.** {MOBILITY#TRANSIT, "***buses***", negative} **3.** {CULTURAL#COMMUNITY, "***diversity***", negative } **4.** {SOCIOECONOMIC#AFFORDABILITY, "***affordable***", negative} |
| @prinzrob @akronisticlotor My own neighborhood is my least ***favorite area*** there's no ***bike lanes***; everyone drives like they're on a rural back road (ie, carelessly). Though ***biking up*** san pablo through west **Berkeley** has been the most ***terrifying*** of all my ***bike routes***; the closest I've come to being ***hit***. | **1.** {MOBILITY#BIKEABLITY, "**bike**", negative} **2.** {SAFETY#CRASH, "***hit***", negative} |

***fastText Sentiment Analysis:***

Like most machine learning algorithms, performance of a fastText model varies depending on the preprocessing of data and the choices of hyperparameters. The optimal values of these parameters tend to vary depending on the dataset or task. While fastText has recently developed an autotuning hyperparameter feature[8], the current study searched for the best set manually to be able to directly monitor the relationship between performance of the text classifiers and the choices of training datasets.

The model components consist of epoch, learning rate (lr), word n-grams and dimension (dim), with the standard ranges respectively between 5-50, 0.1-1.0, 1-5 and 100-300. The most important hyperparameters of the model are the range of size for the subwords and its dimension. Word n-grams relates to sequencing of the tokens in the text. As mentioned in the preprocessing section, word vectors capture hidden information about a language (e.g. word analogies and semantic) and the subwords are all the substrings contained in a word between the minimum (minn) and the maximum (maxn) size. Subword-level information is in particular interesting to build vectors for unknown words. Meanwhile, the dimension controls the size of these vectors. The larger the dimension, the more information can be captured. However, it requires learning more data and it is also harder and slower to train. Therefore, the optimality of the training parameters is dependent on the quantity of the dataset.

[8] https://fasttext.cc/docs/en/autotune.html

The epoch parameter determines the number of times a model will loop over the data. Adding more epoch would increase the training time, therefore in case of massive datasets, it is recommended to loop over it less often. On the other hand, the higher the learning rate hyperparameter, the faster the model converges to a solution. This is while the risk of overfitting to the dataset would increase. The fastText default values for epoch and lr are 5 times and 0.05 [9].

An overview of the results of fastText models compared between two different training datasets is shown in table 4. The predicted results from each model were tested manually in order to confirm the reported accuracy of fastText models for transportation planning/livability contexts.

In all final models, the same hyperparameters lr, epoch, and wordNgrams are interplayed while the dimension parameter is changed between 20 to 100. This suggests that the extended training size in this study yields an improvement in the fastText validation process, therefore the model needs less dimension in order to train.

In order to evaluate the performance of the learned classifier, fastText requires splitting the data into training and validation sets and similar to other classification algorithms, the accuracy of the model is calculated based on the fraction of correctly predicted Tweets in each process. Here, two validation sets are used to test the learned classifier (see table 4). The fastText model developed on the Sentiment 140 dataset in both training and validation processes provided a result on the neighborhood Tweets test set that

[9] https://fasttext.cc/docs/en/unsupervised-tutorial.html

outperforms other tried datasets including SemEval 17. This means, for the smaller size dataset (SemEval 17), fastText validation results are slightly worse than those of the Sentiment 140 model (0.65 vs 0.83). This is while retaining a short training time in each scenario (less than 5 minutes), far less than performance time of the state-of-the-art deep learning algorithms.

Table 3.4 Training accuracy - comparison between two labeled Twitter datasets

| **Model data** | | **Hyperparameters** | | | | **Accuracy** | |
|---|---|---|---|---|---|---|---|
| **training** | **validation** | **lr** | **epoch** | **wordNgrams** | **dim** | **training** | **validation** |
| Sent. 140 | SemEval 17 | 0.05 | 15 | 2 | 20 | 0.943476 | 0.65021 |
| | | 0.05 | 15 | 2 | 100 | 0.941563 | 0.618878 |
| Sent. 140 | Sent. 140 | 0.05 | 15 | 2 | 20 | 0.94473 | 0.83298 |

***Sentiment Maps:***

An online Jupiter visualization package was developed to translate the neighborhood Twitter reviews from qualitative to quantitative data (refer to the code availability section). This visualization tool receives the extracted neighborhood Tweets, preprocesses the data, builds the fastText models and visualizes the aggregated classified sentiments at the zipcode level. The goal is to gain new insights about the neighborhood satisfaction to facilitate the project decision-making process.

To consider more reliable sentiments, only neighborhoods with more than 20 Tweets were taken into account. Next, the binary predictions were replaced with the related labels, 0-negative and 1-positive, to calculate the total neighborhood sentiments using the polar aggregation. The research results in figure 2 reveals available satisfaction maps of

Twitter neighborhood data in NYC for a short data collection span of six months. In the New York region the neighborhoods of Brooklyn (zipcode 11226), Ozone Park (zipcode 11417) and Bronx (zipcode 10461) have the most active inhabitants on the Twitter platform related to neighborhood perceptions.

These limited detected sentiments are compatible with the national Twitter penetration rate in the U.S. According to a new Pew Research Center survey[10] conducted in Feb. 2019, around one-in-five U.S. adults (22%) say they use Twitter. Among this population, in a late 2018 survey[11], it is also revealed that the Twitter users tend to be younger than 50 (mainly between 30-49), have at least a college degree, and are wealthier than the general public (earn at least $75,000 a year). Additionally, 13% of the adult users in U.S. keep their accounts private. The major portion of users who opt not to post tweets publicly are women, who are also more active when it comes to creating or favoriting tweets. Thus, it can be expected that a large proportion of the tweets would be shared without a specific reference of location (either in the user profile or added to the tweet). Mechanisms to address these limitations are needed to make Twitter more valuable for the transportation planning domain. If steps are taken to create more balanced expression via Twitter, then sentiment analysis and visualization can provide a very valuable tool for transportation planners to quickly identify areas where further investigation and investment may be warranted, as demonstrated in figure 4.

---

[10] https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

[11] https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/#fn-22366-1


*Figure 3.3 Sentiment-derived satisfaction of neighborhoods in New York*

As previously mentioned in the data collection section, while a portion of Tweets are geotagged or mention the location, there are certain limitations with online reviews that are important to consider. First, as pointed out by Zivanovic et al. 2018, there are risks of 'migration bias' and representability with Twitter data. This means that the statement about a specific location could be shared from a completely different location and different time; in addition to the problem that the use of Twitter is very uneven among the population (e.g. by age group, by income group, and languages they use). In another study in 2017, Blank and Lutz investigated the representativeness of different social media platforms and concluded that Twitter users in UK are greatly different from the total population only in terms of age and income. Second, the main challenge of online review data (Twitter, Niche, and Flickr) in comparison with predefined sources of extracting neighborhood information (e.g. surveys, focus groups comments, public meetings) is that they unfortunately do not provide the demographic information of the users who wrote the reviews. When the demographic information is available along with the online reviews, it is possible to develop customer profiling and translate the derived sentiments more accurately into the decision-making components/measures. As people can have different expectations of their living environment, depending on their cultures, backgrounds, and socioeconomic statuses (Hu et al, 2019). Therefore, in order to detect these limitations in the current study, a comparative analysis was necessary at the neighborhood level.

While there is an uneven representation of neighborhoods, the visual content contains some complementary information to the texts making it valuable to employ text and visualization jointly. One popular metadata for explaining these two sets of information is residential demographics that has been commonly used to translate relational information into project decision making components. Thus, this study assumes the differences in neighborhood satisfaction patterns are driven by demographic differences and describes trends that are most likely to shape these patterns, starting with appraising the related Census demographic trends and concluding with possible impacts of those trends on users' satisfaction and expectations.

Demographic data from US Census Bureau was used as an objective dataset with which the neighborhood residential characteristics were compared. Three demographic characteristics were more distinctive among the most reported neighborhoods: age, mean household income, and educational attainment. According to these information (presented in figure 3) and from aligning it with sentiment maps in figure 2 and the zipcodes, it is inferred that in NYC:

1. Residents younger than 50 and mainly in their 30s are more likely to share Tweets about the quality of their neighborhood Figure 5.
2. Most of the residents in the active neighborhood areas have pursued higher education at least at the bachelor level.
3. The mean household income in the active neighborhood areas is not the highest level and mainly reflects middle class households (between $50-70K).

While this study found some significant limitations with Twitter data currently, there is much potential for it to enhance transportation planning strategies. First, the nature of Tweets provides more rich information that may not be obtained through traditional mechanisms because it is not biased by a survey developer. In the example presented in this study, an additional feature influencing livability perceptions was identified through the Twitter analysis that was not uncovered using traditional approaches. Second, Twitter provides an opportunity to identify factors influencing livability that are highly positive or negative (and therefore may be more important to address) as shown in the lack of neutral tweets in the case study datasets. With large amounts of data, high frequency issues can be identified and addressed. Given the current limitations with Twitter data, it is recommended that it be used in combination with other sources of datasets and that data be collected in a longer time span. Transportation agencies can do this more rapidly through use of the premium platform privileges via paid accounts, and can also track changes in sentiment over time.

This study also demonstrated the utility of fastText and the Sentiment 140 training dataset for transportation planning contexts. The combination provides an accurate and efficient tool that shows promise for enhancing decision-making. The approach demonstrated in this study can be applied to online data alongside or before developing customer surveys, for example, for the purpose of identifying additional features that need to be examined in transportation planning surveys.

There are numerous strategies that transportation agencies may employ to increase the utility of Twitter for sentiment analysis for livability or other purposes. First, targeted outreach efforts can enhance participation and sharing of meaningful data, without the need for people to share geocoded location data. For example, agencies can develop hashtags and public outreach campaigns to encourage people to share thoughts related to specific topics (such as #TDOTlivability or hashtags incorporating specific neighborhood names and keywords). Intentionally developing Twitter as a means of communication requires not only outreach but also sharing by the agency of how these comments will be used to enhance planning practice and investments, and developing two-way communication.

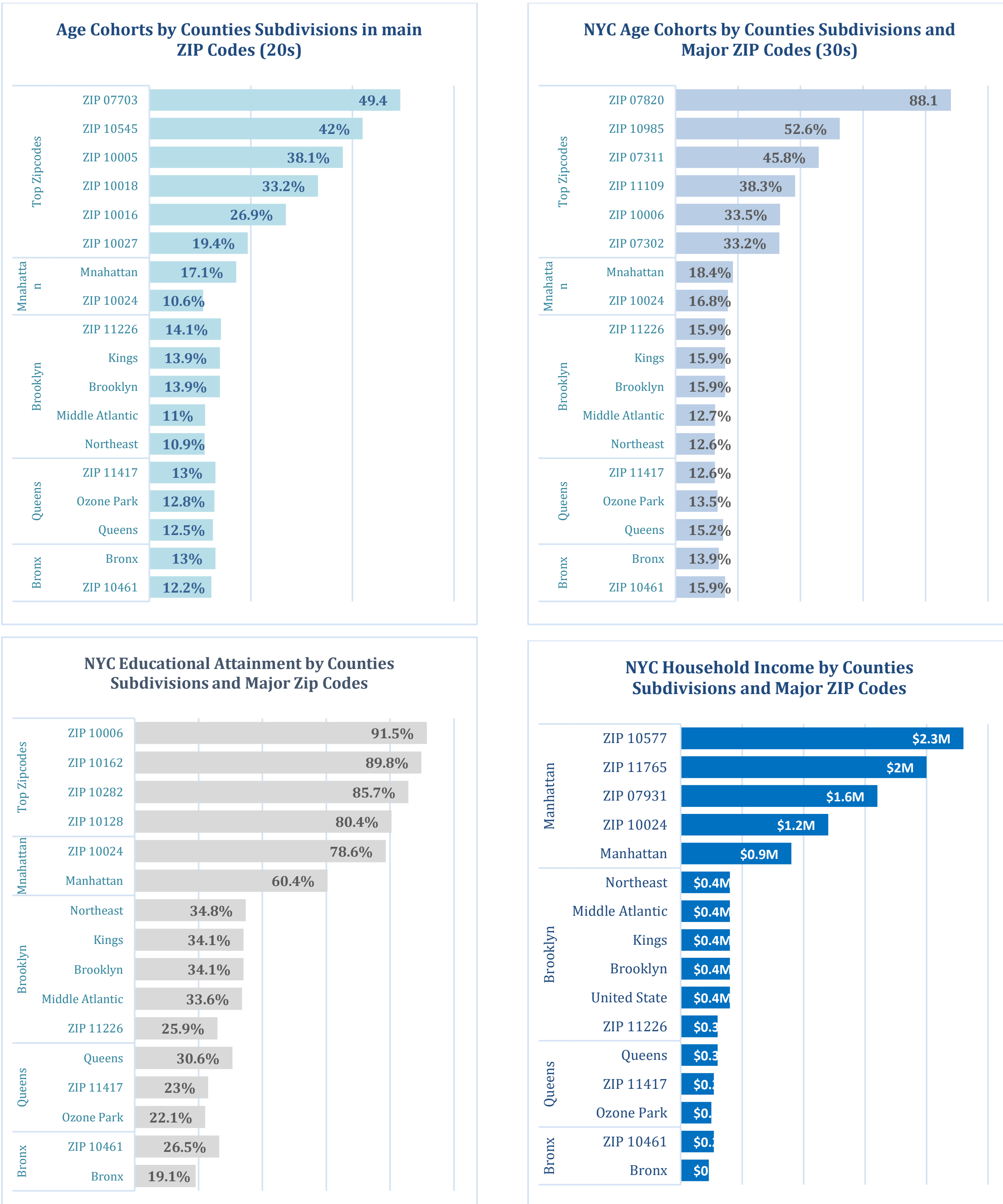


*Figure 3.4 Census demographics by Counties Subdivisions and major zip codes in the New York area, (a) age (20s), (b) age (30s), (c) mean household income, and (d) Educational Attainment.*

Additionally, transportation agencies struggle to get input from a variety of populations, including younger adults. Twitter already captures a younger audience, and because adults in their 20s-30s are digital natives, they are much more likely to be motivated to engage via a social media platform rather than traditional venues like public meetings. With carefully planned communication strategies, Twitter can be leveraged to balance participation and increase engagement and perceived relevance of transportation agencies. Low income populations are another demographic traditionally difficult to engage in the transportation planning process. Many communities have programs specifically designed to provide internet access and mobile devices to disadvantaged populations. According to a survey of teen social media use in 2018 by Pew Research Center[12], young adults from low income communities are using social media at higher rates (70%) than their higher income counterparts (36%). This also points to the opportunity to create education and outreach programs designed to encourage input via Twitter from this population to create more balanced engagement. Expression via social media may also be more attractive to these groups, who may be less likely to express opinions in public forums where they may feel hesitant because of the presence of people from higher income or education demographics who may dominate discussions (Jacobs et al. 2009). And, in fact, social media may be a more convenient and less threatening venue for expressing opinions about neighborhood quality for many people, as evident from this Tweet from a user in San Francisco:

[12] https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/

*“after being inspired I found a site that allows me to report any issues I see in my neighborhood and Oakland in general check it out and see if it is available for your area”*

The main contribution of this study is in preparing the path for transportation planners to be able to actively identify project performance measures and their impacts on residential quality of life while collectively detecting key decision-making components for residents. In this age of increasing IoT and digital interconnectedness, agencies who want to engage stakeholders must begin thinking about engagement strategies in new ways. The use of social media for text mining and sentiment analysis provides a new approach that can not only enhance understanding of factors affecting livability or other transportation considerations, but also has the potential to engage stakeholders that previously have not participated in planning decisions.

### 3.5 Conclusion

This study has provided a review analysis package based on users’ stated preferences in social media for transportation planning decision-making purposes. The Github instruction page provides guidance for agencies in developing an efficient online text classification tool combined with a visualization approach for automating the performance measure extraction.

In addition to examining the application of fastText algorithms in this domain, a guided annotation process was established towards the extraction of user-defined neighborhood expectations, and a Jupyter notebook approach was developed. The Twitter dataset reveals meaningful information and the results show promise for engaging the public in expressing their neighborhood satisfaction and expectations via a targeted social media strategy. The aspect-based annotated results from this study are compatible with the extracted features from a previous study that used traditional survey techniques. In the current study, an additional neighborhood aspect (Cultural) has been extracted and strategies are suggested for promoting the active use of social media by neighborhood residents in order to achieve more efficient input to the planning process. This approach shows particular potential for livability studies, however; the methodology could be targeted for numerous planning-related interests of transportation agencies.

Consequently, the outcome of this work demonstrates the potential of Twitter to significantly increase agencies' public involvement activity and to obtain enhanced information about the efficiency of the system and the areas that need improvements. Developing strategies that leverage this tool may provide planners the ability to decide more readily on the projects and investment distributions to best enhance the communities they serve.

## References

Agard, B., Kusiak, A. (2004) Data-Mining Based Methodology for the Design of Product Families, International Journal of Production Research, Vol. 42, No. 15, pp. 2955-2969.

Albouy, D., Lue B. (2015). Driving to Opportunity: Local Rents, Wages, Commuting Costs and Sub-Metropolitan Quality of Life. Journal of Urban Economics 89: 74-92.

Baziotis C., Pelekis N., Doulkeridis C., (2017). DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis.

Ben-Akiva, M. and B. Boccara (1995). Discrete Choice Models with Latent Choice Sets". International Journal of Research in Marketing 12. 9-24.

Ben-Akiva, M., and Lerman, S. R., (1985) "Discrete Choice Analysis: Theory and Application to Travel Demand", MIT Press.

Bojanowski P., Grave E., Joulin A., and Mikolov T. (2017). Enriching word vectors with subword information. TACL 5:135–146.

Cataldi, M., Ballatore, A., Tiddi, I., Aufaure, M.A., (2013). "Good location, terrible food: detecting feature sentiment in user-generated reviews". Social Network Analysis and Mining. 3 (4): 1149–1163. CiteSeerX 10.1.1.396.9313.

Chien C-F, Kerh R, Lin KY, et al. Data-driven innovation to capture user-experience product design: An empirical study for notebook visual aesthetics design. Comput Ind Eng. 2016;99:162–173.

Cliché M., (2017). BB twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs.

Couture V., Handbury J., (2015). Urban revival in America, 2000 to 2010.

Flaes J. B., Rudinac S., and Worring M., (2016). What multimedia sentiment analysis says about city liveability. In ECIR '16, pages 824–829.

Gibbons J., Malouf R., Spitzberg B., Martinez L., Appleyard B., Thompson C., Nara A., and Tsou M. H., (2019). Twitter-based measures of neighborhood sentiment as predictors of residential population health. PloS one, 14(7):e0219550.

Herrmann, A., Hahn, C. H., Johnson, M. D., Huber, F., (2002). Capturing Customer 5 Heterogeneity using a Finite Mixture PLS Approach. Social Science Research Network, 6 Volume 54, pp. 243-269.

Hu Y., Deng C., Zhou Z., (2019). A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. Annals of the American Association of Geographers, pages 1–21.

Hu, M., Liu, B., (2004). Mining and Summarizing Customer Reviews. Proceedings of KDD 2004.

Jacobs, L. R., Cook, F. L., Delli Carpini, M. X., (2009). Talking Together. Chicago: University of Chicago Press.

Jiao, Jianxin, & Zhang, Yiyang (2005). "Product portfolio identification based on association rule mining" Computer-Aided Design, 37, 149–172.

Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2017). Bag of tricks for efficient text classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL).

Kamakura WA., Russell GJ., (1989). A probabilistic choice model for market segmentation and elasticity structure. Journal of marketing research.

Kim, S., Chung, J. (2016). Reinterpretation of the Likert scale for public transportation user satisfaction: Pattern recognition approach. Transportation Research Record. Journal of the Transportation Research Board, 2541, 90–99.

Kleinberg, J., Mullainathan, S., Raghavan, M., (2016). Inherent trade-offs in the fair determination of risk scores.

Li, S., Nahar, K. & Fung, B.C.M., (2015). Product customization of tablet computers based on the information of online reviews by customers. J Intell Manuf 26, 97–110.

Lin KY, Chien CF, Kerh R., (2016). UNISON framework of datadriven innovation for extracting user experience of product design of wearable devices. Comput Ind Eng.; 99:487–502.

Lo, S.C., (2013). Expectation-maximization based algorithm for pattern recognition in traffic 1 speed distribution. Mathematical and Computer Modelling 58.1: 449-456.

Meehan L. A., Whitfeld G. P., (2017). Integrating health and transportation in Nashville, Tennessee, USA: From policy to projects, Journal of Transport & Health, vol. 4, pp. 325–333.

Michalek, J. J., Feinberg, F., and Papalambros, P., Linking Marketing and Engineering Product Design Decisions via Analytical Target Cascading. J. Prod. 814 Innovation Manage.: Spec. Issue Des. Market. New Product Dev., 22, pp. 42–62.

Mittal V., Kamakura WA., (2001). Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics, Journal of marketing research.

Moon, S. K., Simpson,T.W.,&Kumara, S.R.T. (2010). Amethodology for knowledge discovery to support product family design. Annals of Operations Research, 174, 201–218.

Morency C., Trépanier M., Agard B., (2007). Measuring Transit Performance using Smart Card Data, World Conference on Transport Research, San Francisco, USA, June 24-28.

Nashville Area MPO. Middle Tennessee Transportation and Health Study Final Report. Nashville TN: Nashville Area MPO; 2013.

Pang B, Lee L., (2008). Opinion mining and sentiment analysis. Found Trends Inf Retr. 2: 1–135.

Paredes M., Hemberg E., O'Reilly U., Zegras C., (2017). Machine learning or discrete choice models for car ownership demand estimation and prediction? 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, pp. 780-785.

Quercia, D.; Ellis, J.; Capra, L.; and Crowcroft, J. (2012). Tracking gross community happiness from tweets. In Proceedings of the International Conference on Computer Supported Cooperative Work and Social Computing.

Proserpio D., Zervas G., (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. Marketing Science.

Rosenthal S., Farra N., Nakov P., (2017). SemEval-2017 Task 4: Sentiment Analysis in Twitter. SemEval@ACL 2017: 502-518.

Saeidi M., Venerandi A., Capra L., Riedel S., (2017). Community Question Answering Platforms vs. Twitter for Predicting Characteristics of Urban Neighborhoods.

Sarram G., Ivey S. S., (2017). Evaluating a Survey of Public Livability Perceptions and Quality-of-Life Indicators: Considering Freight-Traffic Impact. ASCE International Conference on Sustainable Infrastructure 2017.

Sarram G., Ivey S. S., (2018). Investigating Customer Satisfaction Patterns in a Community Livability Context: An Efficiency-Oriented Decision-Making Approach. ASCE International Conference on Transportation and Development.

Schweitzer, L., 2014. Planning and social media: a case study of public transit. J. Am. Plan. Assoc. 80 (3), 218–238.

Shao, X.,Wang, Z., Li, P.,&Feng, C. J. (2006). Integrating data mining and rough set for customer group-based discovery of product configuration rules. International Journal of Production Research, 44, 2789–2811.

Shah Fa., Sirts K., Pfahl D., (2018). The impact of annotation guidelines and annotated data on extracting app features from app reviews.

Smart Mobility 2010. California Department of Transportation (Caltrans).

Song, Z., & Kusiak, A. (2009). Optimizing product configurations with a data-mining approach. International Journal of Production Research, 47, 1733–1751.

Tucker, C., Kim, H.M., (2008). Optimal product portfolio formulation by merging predictive data mining with multilevel optimization. J. Mech. Des. 130, 041103.

Tucker, C., Kim, H.M., (2011). Trend mining for predictive product design. J. Mech. Des. 133, 111008.

Wassenaar, H., Chen, W., Cheng, J., and Sudjianto, A., 2005, “Enhancing Discrete Choice Demand Modeling for Decision-Based Design,” ASME J. 811 Mech.Des., 127(4), pp. 514–523.

Zhang K., Narayanan R., Choudhary A., (2010). Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking. WOSN.

Zivanovic, S., Martinez, J., Verplanke, J., (2018). Capturing and mapping quality of life using Twitter data. GeoJournal 85, 237–255 (2020).

## Conclusion and Future Work

Decision-making in transportation planning is complex and is particularly challenging in the case of assessing community performance related to livability, as this is an abstract concept reflecting customer satisfaction and requires significant input from stakeholders. This research proposed a two-phase conceptual model for developing a customer-centric approach to assessing livability. An integrated approach to Phase I of the model is described that examines the potential of machine learning for extracting information and factor relationships from traditional survey measures, and of sentiment analysis of online review data to further this understanding.

This study has demonstrated that prioritizing livability indicators through data mining techniques is plausible and shows promise for developing more robust analysis and informed decision-making. The results of the machine learning study provide evidence that there is a rule relating neighborhood perceptions to participants' livability scoring systems. The methodology, while consistent with previous results, is able to uncover additional impact of key features on livability perceptions. It is expected that with additional research, larger datasets, and data from multiple settings, more efficient livability indicators can be identified, adopted, and employed for planning purposes. The feature selection analysis presented in this study helped in understanding the livability metric's strengths and weaknesses and the extracted indicators motivate further research on exploring a more representative and efficient mechanism for developing a systematic indicator set.

The innovative application of online information from social media regarding human mobility, safety, activity, and other factors provides a new approach that can not only enhance understanding of factors affecting livability or other transportation considerations, but also has the potential to engage stakeholders that previously have not participated in planning decisions, including younger generations. This study also resulted in a sentiment analysis package that allows extraction and application of users' stated preferences in social media for transportation planning decision-making purposes. This procedure is also able to reshape the consistency of proxy settings with reality, while reflecting the changes in community preferences over time. The package includes an efficient online text classification tool combined with a visualization approach that automates performance measure extraction.

Although the Twitter dataset revealed meaningful information and the results showed promise for engaging the public in expressing their neighborhood satisfaction and expectations via a targeted social media strategy, it is recommended that it be used in combination with other sources of data and that data be collected in a longer time span, due to the representativity limitation. The compatibility of the aspect-based annotated results from chapter 3 with the extracted features from the previous study in chapter 2 that used traditional survey techniques, suggests that in future studies it would be beneficial to conduct a comprehensive aspect-based annotation task in order to develop a specific training dataset for the domain of neighborhood satisfaction (Livability). Additionally, developing strategies and approaches that leverage this tool to provide planners the ability to decide more readily on projects and investment distributions can be a valuable addition.

Developing a predictive mechanism in order to correlate the residential demographics with transportation facility utilization rates to predict deficiencies and classification scores (phase II) is suggested as future work. This would help in developing a quantitative livability metric and decision-making tool at the same time. Forecasting the livability score accurately through correlating the Phase I and Phase II subjective and objective measures could provide a combination of validated coefficients. This could extend to a Multidimensional Livability Index (MLI) incorporating a range of indicators. This approach can reflect the complexity of livability and better communicate the policies aimed at improving it.