Thesis Project Portfolio

Predicting Gentrification Using Machine Learning in a Post Covid-19 World

(Technical Report)

Using Actor Network Theory to Understand Covid-19's Effect on the Impact and Spread of Gentrification within Poor and At Risk Communities

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Eric Guan

Spring, 2021

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Many major metropolitan areas struggle with gentrification. That is, the process in which a poorer area is slowly changed by an influx of wealthy residents and redevelopment, causing the original residents to be displaced. Gentrification is hard to track quantitatively, and is more often identified by visually seeing an area change over a period of time, i.e. visiting Brooklyn and noticing that many of the liquor stores have been replaced by upscale coffee shops. For this reason, I focused my technical project on developing a machine learning model that could learn from data the signs of gentrification, and then apply this knowledge to other cities and formulate predictions for at risk communities. Interested in how gentrification has already taken hold in America's major cities, my STS research topic concerned the effects that gentrification had on the populations it affects. More specifically, I wanted to see how the global pandemic Covid-19 would impact gentrification moving forward.

The technical portion of my thesis produced a proof of concept for a machine learning tool to track and predict gentrification in major urban areas using 4 different machine learning model applications. These models were k-means clustering, linear regression, random forest regression, and XGBoost. Using census housing data from California, me and my partner investigated the relationship between home values and median annual income. The clustering model allows us to see concentrations where gentrification could be happening. The linear regression demonstrated a strong linear relationship between income and home value, however there was still a large variance in data that was not accounted for, so we ran a random forest regression, which uses decision trees to more accurately model the data. This regression was more accurate in predicting rising home values based on income distributions, but could still be better, which is why we used XGBoost, a gradient boosting algorithm. The XGBoost regressor showed excellent results, displaying more accuracy and accounting for more variance than the

other models. The technical work we did showed that future tools with enough data would be able to model gentrification in cities and identify at risk boroughs. Our models also showed that data sets containing more specific data about rent or demographics would enhance future models.

In my STS research, I applied Actor Network Theory to analyze the sociotechnical systems of gentrification before Covid-19, the public Covid-19 response effort, and gentrification post pandemic. For each of the aforementioned systems, I constructed a map of the various technical, cultural, and organizational factors in order to understand the complex actors and relationships involved. I saw how human migration patterns were impacted by covid; people are moving out of cities because they can work remotely, meaning gentrification can spread. Looking at how relationships between urban planners, city leaders, and workers in struggling industries evolved, I was able to see how gentrification causes poverty to expand to suburbs, meaning that poor people who were displaced from their previous place of residence stay poor and keep going to areas where they can't really receive a lot of public help causing a cycle of generational poverty. It was clear that Covid-19 exacerbated issues of lack of investment in poor areas, and negatively impacted industries where minorities work such as food and hospitality. Using Actor Network theory, I found that a delicate balance had to be found between actors who have different goals. It's not easy, but the right balance would greatly help the gentrification situation by ensuring that urban development becomes sustainable for everyone.

The projects I completed enriched each other because instead of only seeing graphs or columns of data, it felt like I was able to understand on a deeper level what it all meant and how each row was so much more than a simple data point. The STS portion of my research emphasized to me the importance of understanding how gentrification spreads and the impact it has on vulnerable communities. This research made me see the bigger picture of my technical

work. Oftentimes in engineering, we are taught to find a solution - that all that matters reaching the end of a problem. STS this year has emphasized that it is as much about the process as it is the solution - a stronger deontological methodology if you will - and I was able to incorporate both mindsets into my research. Even though the machine learning tool was technically the end goal of my technical project, it really serves as a small piece in the bigger puzzle of my sts research, which is how we can understand gentrification better in a complex, ever changing world.

Of course, this was not all possible by myself. First and foremost, I have to thank my partner in the technical research, Jonathan Wen. Jon did a lot of the heavy lifting in building out the model, and the final project would not be possible without him. Jon, I'm privileged to have worked alongside you my final year here at the university, but most of all I'm happy that we have remained good friends for 8 years now. I also need to thank my STS professor Kathryn A. Neeley. Professor Neeley honestly gave so much advice on the STS frameworks as well as the direction I wanted to take my research. She is a fantastic listener and is a major reason I was able to find what I did. Her consideration in giving me time to recover after I was in a serious car accident in the fall showed me that she cared more about me and my work than giving a final grade, and that meant the world. Beyond the named individuals, the list of people to thank goes on and on. None of my UVA experience would be possible without the friends and family I made along the way. If I learned anything from my time here to pass along to future students, it would be this quote from the iconic Disney show, *Jessey*: "Just enjoy it. It feels like a party every day."