

## **Thesis Project Portfolio**

### **Developing Airbnb Price Matching Tool Using Various Machine Learning Models**

(Technical Report)

### **The Socio-Political Implications of Facial Recognition Technology**

(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

**Vaidic Naik**

Spring, 2023

Department of Computer Science

## **Contents of Portfolio**

Executive Summary

Developing Airbnb Price Matching Tool Using Various Machine Learning Models  
(Technical Report)

The Socio-Political Implications of Facial Recognition Technology  
(STS Research Paper)

Prospectus

## Executive Summary

Machine learning is a rapidly expanding field in computer science that has been growing in popularity. There are many use cases for the technology because of its ability to make meaningful predictions given that it is trained on a large data set. Some of the applications of machine learning include image and speech recognition, natural language processing, fraud detection, and recommendation systems. However, despite its advantages, machine learning has some potential drawbacks, such as the risk of overfitting, which can lead to poor generalization performance, and the possibility of introducing bias into the model. It is important to carefully evaluate the data used to train machine learning models and ensure that they are representative and unbiased. Additionally, transparency and interpretability of the models are critical, especially when dealing with sensitive information or making decisions that could have a significant impact on individuals or society. As machine learning continues to evolve, it is important to consider its ethical implications and ensure that it is used responsibly to benefit society.

During my internship at Capital One's Capital One Shopping Department, I was a part of the Travel team that developed a price matching tool to help users find better deals for hotels through Capital One's partners. Our team wanted to expand this product to vacation rental properties, specifically Airbnb properties. Our task was to find rental properties from Capital One's partners that were similar enough to the Airbnb properties that users were viewing so that they would choose our recommendations. To solve this problem, we employed machine learning models to compare a variety of features such as size, amenities, name similarity, description similarity, and distance. Our data was acquired through extensive web scraping of Airbnb properties and properties from Capital One's partners which we labeled as canonical properties. We decided to use k-nearest neighbors to compare size and amenities. We generated two separate scores for these features since we wanted to weigh them differently. For amenities, we filled in the null data assuming that the property probably didn't have the amenities in question. We used Doc2Vec to compare the descriptions of the properties and a pre-trained open-source NLP model called spacy to compare the names of the properties. We scaled the distance between properties based on the density of properties in the area. Finally, we made our final score by taking a weighted average of scores. Our final pipeline was able to make 5x more accurate predictions when compared to our labeled data that was scored by a variety of users. The team consisted of another intern and two software engineers who acted as mentors for the project. The initial codebase was minimal, allowing us to decide which features we would compare and which machine learning models we would use to rate their similarities.

Facial recognition technology is widely used by law enforcement and federal agencies, but its reliance on this technology raises ethical questions, especially concerning its potential racial biases. Studies have shown that facial recognition technologies are more likely to falsely identify certain racial groups, with black women being disproportionately misidentified. This bias can negatively impact groups that are already subject to racial discrimination within the law enforcement system. The source of bias in facial recognition technologies lies in the training of deep learning convolutional neural networks, which may have a disproportionate representation of certain marginalized groups in the training data. The problem of misidentification can be

minimized by using these technologies alongside human forensic examiners, but accountability is also essential. Most federal agencies use a third-party facial recognition system without properly considering the issue of bias. It is the government's responsibility to ensure that they understand the shortcomings of these systems and take better precautions to mitigate bias. As the technology improves, the rate of inaccuracies decreases, but it is still important to consider racial biases when dealing with facial recognition technology. There has been a long-lasting distrust in the law enforcement and judiciary system, especially when talking about racial bias. Recent events of police brutality have heightened tensions between the public and law enforcement, and the rise of social media has made these atrocities known around the nation, creating a call to reform the law enforcement agencies. The use of facial recognition technology only adds to the distrust of the system, especially when it comes to racial biases. It is crucial to address these issues to regain public trust and ensure that the technology is used ethically and accurately.

In conclusion, machine learning has shown great potential in various industries, from recommending products to predicting outcomes. However, it is important to be aware of the potential drawbacks of these models, such as the risk of overfitting and bias. Transparency and interpretability are critical, especially when the outcomes could significantly impact individuals or society. As technology continues to advance, it is important for researchers and developers to consider the ethical implications of their work and ensure that it benefits society. When it comes to facial recognition technology, it is essential to recognize the potential for bias and take necessary precautions to minimize it, such as using human examiners alongside the technology and ensuring accountability. Researchers should continue to study the issue of bias in facial recognition technology and work towards creating fair and unbiased systems. It is crucial to address these issues and work towards creating responsible and ethical use of machine learning and AI.