

## **Thesis Portfolio**

**Predicting Motor Vehicle Accidents using Machine Learning Techniques**  
(Technical Report)

**Privacy Issues with Unwarranted Data Collection from Web and Mobile Applications**

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Sciences  
University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree  
Bachelor of Science in Engineering

Author  
Akanksha Alok  
May 1st, 2020

## Table of Contents

Sociotechnical Synthesis.....	3
Predicting Motor Vehicle Accidents using Machine Learning Techniques.....	6
Privacy Issues with Unwarranted Data Collection from Web and Mobile Applications.....	16
Prospectus.....	35

## **Sociotechnical Synthesis** (Executive Summary)

### *The Advantages and Disadvantages of Extensive Data Collection in our Society*

A motor vehicle accident can occur in a matter of a few seconds. However, noticing abnormalities in the speed, acceleration or location of the vehicle can alert authorities of the accident before it has even occurred, and injured victims can receive medical assistance as soon as possible without going unnoticed for long periods of time. My technical research project focused on creating a novel approach using different Machine Learning techniques to detect impending vehicle accidents, by using sensor data collected from an iOS device situated in the vehicle. My project, and almost every Machine Learning project, can only be useful if provided with the correct datasets for training and testing. GPS Spoofing and the hacking of sensors and cameras on transportation vehicles have become a prevalent problem in our society, where current GPS navigation systems used in cars can be tricked, deflecting vehicles thousands of meters from their intended destinations without the knowledge of the drivers. If this tampered data were to be fed to my Machine Learning algorithm which relies heavily on these sensor readings, then my predictions of the accident and where it has occurred would be completely off. My STS paper goes into this topic of preventing malicious parties from intercepting and tampering our personal data, such as that collected from our vehicles, and stopping them from infringing on our privacy rights. Overall, my research paper explored the ways in which a user's social media accounts, and web and mobile applications wrongfully elicit their personal information without their knowledge, and possible steps that we can take as a society to prevent this from happening.

Like mentioned before, the most difficult task was finding an appropriate time-series dataset for my Machine Learning algorithm for my technical project. In order to circumvent that

problem, I collected my own dataset by driving around Charlottesville, while using an iOS device inside the car for collecting sensor readings. The first component of the project was a Recurrent Neural Network (RNN), which was used to predict the future states of the motor vehicles' sensor readings. Once the future states of the vehicle were predicted using this neural network, I assessed whether these states were indicative of an accident, using an Autoencoder neural network in the second part of the project. The overall system developed in my project was able to pinpoint the future timestamp of when a car would be involved in an accident, along with its location.

On the other hand, in my STS paper, I was still analyzing trends and patterns with data, but instead of finding car accidents, I was exploring the common ways in which a user's social media accounts, and web and mobile applications wrongfully elicit their personal information without their knowledge. This breach of privacy can affect almost every person on this planet, regardless of race, age, locality, or economic status, and can range from a person's name and email being stored in an external database to their entire identity being stolen. After thoroughly understanding the problem, I suggested ways in which we can solve this problem from a technical, legal, and social standpoint. I have looked into the existing rules and regulations provided by governing bodies of the United States on how data from these applications can be used and distributed, and have listed improvements to the laws in place for privacy in the technological realm. Most of all, I attempted to raise awareness among users of the potential dangers of sharing too much information online or on other online applications.

My technical paper and STS paper are the perfect opposites, as they show the ways in which datasets can be rightfully used to better our society, and ways in which our data can also be misused and impede on our privacy as human consumers of technology. Both provide insight

into the advantages and disadvantages of extensive data collection, especially as we progress into an age of automation and technological advancement. As with any innovation, there is always a good and bad side to it, and these papers highlight that for the topic of user data.