# The Use of Machine learning in Database Systems
(Technical Report)

# The Struggle over Predictive Analytics in U.S. Urban Policing
(STS Research Paper)

An Undergraduate Thesis Portfolio
Presented to the Faculty of the
School of Engineering and Applied Science
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

by

Grady Roberts

May 9, 2021

**Preface**

The proliferation of data and machine learning presents opportunities and hazards.

The total volume of enterprise data generated each year is expected to reach 175.8 zettabytes (ZB, equivalent to one trillion gigabytes) by 2025, up from 18.2 ZB in 2015. A survey found that only 32% of this data is available in a state ready to be analyzed. With data volume growing faster than companies can keep up with, it is important to find efficient ways to store, manage, and analyze data. One approach is the application of machine learning (ML) to aspects of database design and query optimization. To this end, a literature review of the use of ML to address the problems of managing data at scale is completed. These problems include creating an efficient index of the data, the expense of joining tables, and the computational cost of copying data into another repository for analysis. We contribute a taxonomy of different methods to solving these problems as well as present open challenges in this area. This taxonomy describes the benefits and drawbacks of each method.

Police forces in many major U.S. cities deploy predictive analytics—tools for predicting crime on the basis of historical data. In the absence of regulation, predictive policing's proponents and opponents compete to influence its perception. This struggle is defined by the scale at which each group characterizes the technology. Proponents of predictive analytics cite its benefits for the community at large, while opponents warn of threats to individuals. Understanding the difference in how participants view the scale of the technology's impact suggests that a bottom-up, community-driven implementation may be more effective.

# List of Contents

1. Technical Report:  The Use of Machine Learning in Database Systems

2. Sociotechnical Research Paper:  The Struggle over Predictive Analytics in U.S. Urban Policing

3. Prospectus