WHARTON
UNIVERSITY of PENNSYLVANIA
GLOBAL YOUTH PROGRAM

Video Segment 2: Machine Learning as a Methodology and Using Diverse Data Sources

"We pull in all kinds of additional data sources, including even using police body cam videos, to audit the records of the officer's narrative of what took place." -Dr. Dean Knox

NOTE-TAKING GUIDE

How has machine learning and the use of unconventional data helped Professors Hamsa Bastani and Dean Knox uncover patterns and extract insights for the following projects? What are some of their diverse data sources?

Deep Web Activity and Deceptive Recruiting:

Greek COVID-19 Project:

Bias in Policing:

Computer Vision Audio Analysis:

WHARTON
UNIVERSITY of PENNSYLVANIA
GLOBAL YOUTH PROGRAM

Explore Business: Using Data to Protect and Service | 1
©2023 The Wharton School, The University of Pennsylvania
globalyouth.wharton.upenn.edu

Professor **Dean Knox** says, "If you think about discrimination in a police encounter, it's a question about whether from the moment an officer lays eyes on a civilian, what chain of events or enforcement actions take place that wouldn't otherwise take place if that civilian had been white." Dr. Knox believes that sunlight is the best disinfectant – and data transparency is his sunlight in tackling police reform. Explain his quote and how he is addressing bias in policing through data and statistical analysis.

Reflection: While machine learning requires datasets to draw conclusions, clearly the work of Dr. Bastani and Dr. Knox also depends on the human perspective for a more complete analysis. Consider this balance of big numerical data and personal quallitative data in their research contexts of deep-web human trafficking and bias in policing. What are your thoughts on the combination of quantitative and qualitative data, especially as artificial intelligence takes a more central role? Will humanity always have a place among the data? Why is qualitative analysis essential? Draw parallels to other contexts, like investing.

THE LANGUAGE OF BUSINESS
(Linked terms are defined in the Wharton Global Youth Video Glossary)

Data   PhD

Demographics   Supply Chain

Academic Research: Research conducted by faculty at a university like Penn that focuses on creating new ideas and perspectives. Academic researchers read all the existing studies and papers examining a specific topic or theory and then see where they can add value with new experimentation and insight.

Active-learning Approach: Engaging in learning by thinking, discussing, investigating and creating.

Artificial Intelligence or AI: Intelligence demonstrated by machines (computers), rather than humans.

Beyond Business: An ongoing conversation at the Wharton School of the University of Pennsylvania that explores the most complex and pressing issues impacting individuals and organizations across the world. Each year Wharton Global Youth creates Explore Business mini-sites around different Beyond Business themes.

Commercial Sex Sales Websites: Websites that sell sex services online.

Computer Vision: A field of artificial intelligence that enables computers to derive information from images, videos and other inputs. Audio analysis, for instance, looks at and interprets audio signals recorded by digital devices.

Data Analyst: A professional who reviews data to identify key insights into a business's customers and how the data can be used to solve problems.

Data Analytics (or just Analytics): The systematic computational analysis of data or statistics. It is used for the discovery, interpretation, and communication of meaningful patterns in data.

Dean Knox: An assistant professor of operations, information and decisions at the Wharton School of the University of Pennsylvania and co-founder of Research on Policing Reform and Accountability.

Dataset: A collection of data.

Deceptive Recruiting: Practices that deceive job seekers, for example, to entice them to accept a position that is not what they expected, such as forced labor. They might be deceived about the nature of the employment, the location or the employer.

Deep Learning: A machine learning technique that teaches computers to do what comes naturally to humans. Deep learning uses Deep Neural Networks (DNNs) to learn levels of representation and abstraction that make sense of data, such as images, sound and text.

Deep Web: The part of the internet, often where drug sales or other illicit activity occurs, that can only be accessed with specialized software, not through a regular search engine. It's also known as the Dark Web.

Deep Web Ads: Advertisements on the deep web that provide information on the sale of illegal products.

Erika James: Dean of the Wharton School with research interests in crisis leadership, workplace diversity and management strategy.

Forced Labor: When individuals are compelled against their will to provide work or service.

Greek COVID-19 Project: A study conducted by Wharton Professor Hamsa Bastani in the summer of 2020 that determined that the application of machine learning (namely Eva, a machine learning tool designed to track data on travelers' COVID-19 screening results along with demographic information and region of origin) made COVID-19 screening at the Greece's borders more efficient.

Hamsa Bastani: An assistant professor of operations, information and decisions at the Wharton School of the University of Pennsylvania, who has been analyzing dark web data related to human trafficking to provide new insights into online sex-trafficking recruiting practices and, more broadly, illicit supply chains on online platforms.

Law Enforcement: The agencies and employees responsible for enforcing laws and maintaining public order, like police departments.

Machine Learning: Using and developing computer systems that learn and adapt by using algorithms and statistical models to analyze and understand patterns in data.

Mobile Location Data: Data collected from interactions between smart devices and other technology; typically geographical information that points to where a specific device is located. This might help us navigate to a new location using our smartphone.

Multi-modal Learning: Machine learning that aims to build models that can process and relate information from different modalities, such as various data types like image, text, speech and numerical data.

Police Body Camera: A camera worn by police that records interactions between community members and law-enforcement officers.

Social Work: An area of study and a profession in which trained professionals help vulnerable people and communities work through challenges.

TellFinder Alliance: A web platform that connects its users to the data of the deep web, where human trafficking and other crimes take place.

Triangulating: Validating data through cross-verification from more than two sources, as a way to test results.

Unconventional Data: Data that might traditionally have been overlooked or was not traditionally used in machine learning and might provide valuable insights.

The Wharton School: The Wharton School, also referred to as just Wharton, is the business school of the University of Pennsylvania, a private Ivy League university based in Philadelphia, Pennsylvania. Wharton is a world leader in business education and a hub for academic research in all areas of business.

EXPLORE MORE IN OUR ONLINE BUSINESS JOURNAL

A Data Project Explores Gun Ownership in New York City