

# A Machine Learning Framework for Exploring Mortality in Developing Countries with Verbal Autopsies

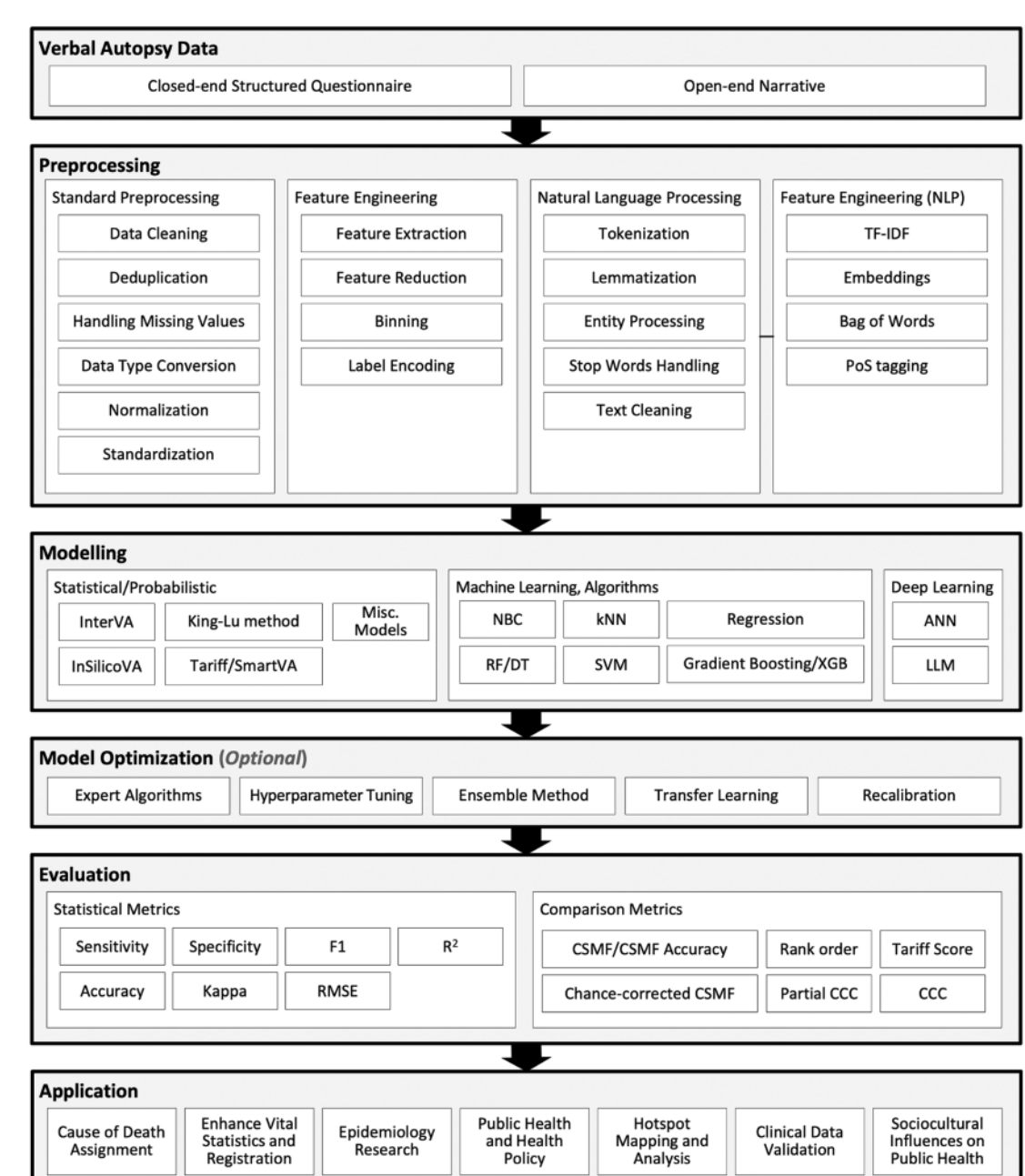
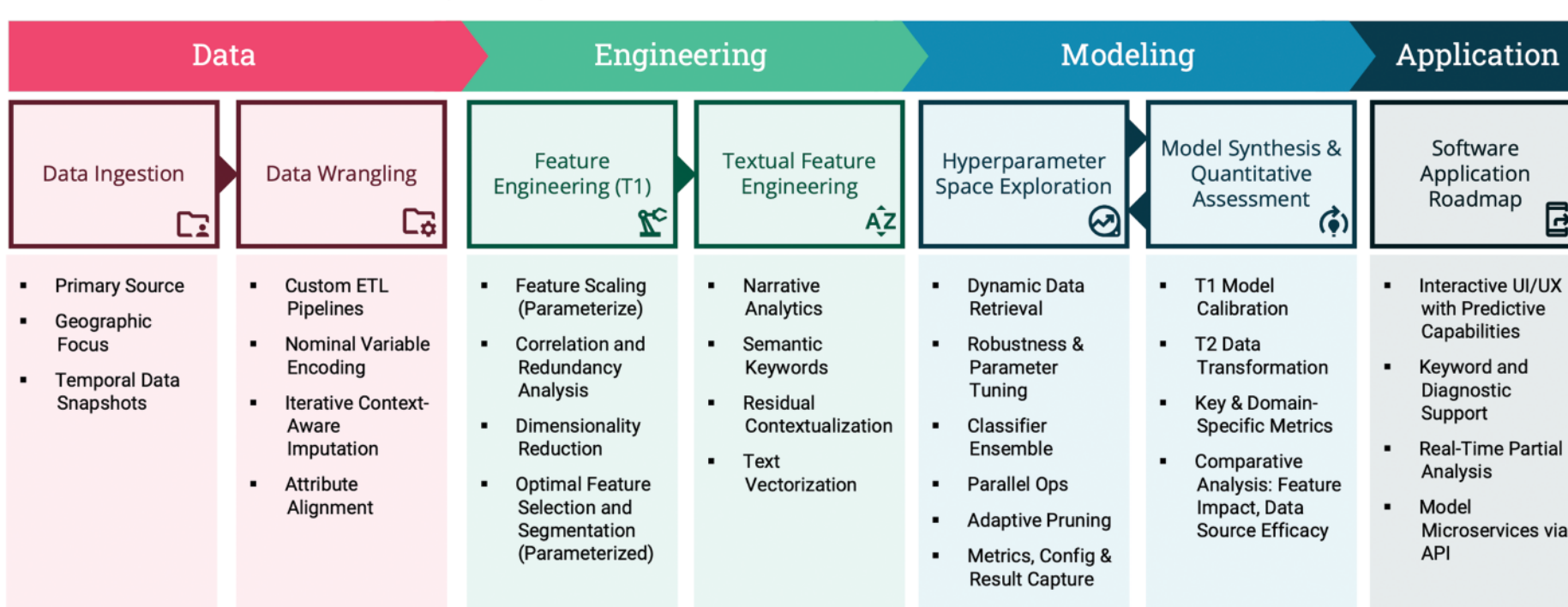
Harmonizing verbal autopsy research into developing a conceptual framework which provides a roadmap to automating cause-of-death prediction using machine learning

Andy Lee

**Dr. Frank Rudzicz**  
ACADEMIC SUPERVISOR

**Richard Wen, PhD**  
INDUSTRY SUPERVISOR

ML Pipeline for Verbal Autopsy Analysis in Sierra Leone



## PROJECT SUMMARY

In low- and middle-income countries (LMICs), a significant number of deaths remain medically unrecorded, resulting in gaps in Civil Registration and Vital Statistics (CRVS) systems. These deficiencies undermine the efficacy of public health policies by hampering efficient resource allocation and timely responses to health crises. The World Health Organization endorses Verbal Autopsy (VA) as an interim measure to ascertain causes of death (CoD). Nonetheless, this process is labor-intensive and often lacks standardization.

While automated methods show promise in delivering scalable CoD predictions and shaping public health policies, they face inherent challenges. Inconsistencies in methodological approaches and reference standards across studies make performance comparison evaluations difficult. Furthermore, there are risks of misclassification and a noticeable absence of a universally accepted ground truth. This project aims to consolidate a decade of research in automated VA methods to establish a unified conceptual framework, equipping stakeholders like epidemiologists and health officials with insights for the effective implementation and validation of automated VA methodologies.

In the practical application of this framework, a case study is being conducted using verbal autopsy data from Sierra Leone, provided by Centre for Global Health Research (CGHR) through openmortality.org. This application explores various machine learning models to predict causes of death, laying groundwork for a future tool designed to assist medical professionals. By proposing a standardized approach, this research aims to enhance the reliability and scalability of automated VA techniques, thereby strengthening the implementation of CRVS systems.