

Literature review structure

Introduction (500 words)

Predicting recidivism (1,000 words)

This section will provide an overview of what recidivism is, why it is important to be able to predict it, and some current methods used. Key topics:

- Defining recidivism
- Key methods and variables used to predict recidivism (unrelated to machine learning)
- Tools used in England and Wales

Machine learning to predict recidivism (1,000 words)

This section will give a broad overview of machine learning solutions to predict recidivism. Key topics:

- Various models and approaches tested with results
- Features used (and how they compare with the key variables from the previous section)

Ethical challenges (2,500 words)

This section will discuss the ethical challenges with using machine learning to predict recidivism. Re-visiting many of the papers from the previous section, it will address:

- COMPAS – many studies on potential racial bias but it's still in use
- Issues with fairness arising from other solutions/papers e.g.:
 - Racial bias
 - Gender bias
 - Age discrimination
- Methods for assessing fairness
- Methods for increasing fairness, including:
 - Ethical use of data to train and test the models
 - Performance measures used (the trade-offs for victims and defendants between false positive and false negatives)
 - Explainability increases fairness (explainable models preferred over deep learning)

Ontology to increase fairness (500 words)

This section will examine how ontologies have been used or proposed to address ethical challenges with machine learning. It will be relatively short because there is not much literature on the subject, hence this dissertation will be new and innovative.