Ministry of Justice – Bias Analysis in Probabilistic Data Linkage

TIN-MOJ-002

About the Organisation

The Ministry of Justice (MoJ) is a ministerial government department responsible for running over 100 prisons, 500 courts in England and Wales, as well as probation services. The department’s priorities are to protect the public from serious offenders, to improve the safety and security of our prisons, reduce reoffending, and deliver swift access to justice.

The MoJ works with many other government agencies including HM Prison and Probation Service (HMPPS) and Youth Justice Board (YJB). The department conducts an extensive programme of analytical and research-based work, helping to deliver a world class justice system. Diverse teams of experts from different fields work together to conduct research reviews, commission research, provide justice statistics and secondary data analysis, cost benefit analysis, forecasting, advanced analytics and data engineering.

Role Description and Responsibilities

Background:

The Ministry of Justice faces the challenge of linking numerous large datasets related to courts, probation, and prisons. These datasets lack a single unique identifier, making it difficult to accurately identify individuals across different databases. Without this fuller picture, MoJ is unable to produce analysis on people that spans data collection systems (prisons, probation, courts). This allows us to tackle key priorities, such understanding what leads to reoffending, shape new policy, and ensure those entering the judicial system have access to the support they require as individuals. Splink was developed by our internal data linking team to address this issue. Existing projects using Splink include the ADR UK’s DataFirst and the BOLD programme.

By generating common linked IDs through probabilistic matching, we have successfully overcome the problem of siloed datasets and the limited understanding of individuals within the judicial system. However, we recognise the importance of ensuring that our data linkage process remains unbiased and does not disproportionately affect any particular group.

Responsibilities:

As the researcher specialising in bias analysis within probabilistic data linkage, your primary responsibilities will include:
1) Conducting an in-depth investigation into potential sources of bias that may arise during the probabilistic data linkage process.

2) Developing an analytical framework to assess the fairness and accuracy of the data linkage process, considering different demographic groups and any potential disparities in matching outcomes.

3) Collaborating closely with the internal data linking team to understand the nuances of Splink's implementation and how bias may affect the overall data linkage results.

4) Producing a comprehensive report outlining your findings, methodologies, and recommendations for mitigating any identified bias within the probabilistic data linkage process.

Expected Outcomes

The successful completion of this research project will lead to the generation of impactful outcomes that address bias within internal linked datasets. The key deliverable will be a comprehensive report, which will be drafted in close collaboration between the intern and their supervisor throughout the project. The report will serve as a crucial resource for promoting fairness, accuracy, and improved operational efficiency within the criminal justice system. The expected outcomes include:

1. **Detailed Bias Investigation Report**: The primary deliverable will be a thorough and well-documented report that systematically investigates bias within the datasets linked using Splink. This report will provide a clear justification for the analytical methods employed and will encompass the entire process of bias evaluation. It will detail the application of these methods, highlighting their strengths and limitations, and present the results of bias assessments. By understanding the nature and extent of bias, the report will shed light on potential disparities in the representation of different demographic groups. This critical insight will enable policymakers, researchers, and operational teams to make informed decisions and implement appropriate strategies to mitigate bias.

2. **Plan for Bias Mitigation**: The report will include a comprehensive plan for addressing the identified bias, which the data linking team can readily implement. This plan will outline actionable steps, guidelines, and best practices to achieve fairness and accuracy in the data linkage process.

Whilst these are our intended outcomes of the project, we appreciate this is a large task. We will work directly with the successful applicant to scope out the project and further shape it such that it fits their interests and what they believe to be feasible within the timescale.

In addition, we believe this project will cause the following secondary outcomes:

- **Enhanced Data Linkage across MoJ**: The evidence produced in the report will not only address bias in these datasets, but will also create a foundation for enhancing data linkage across MoJ.

- **Advancing the Development of Splink**: The insights gained from investigating bias will inform the creation of features and functionalities that empower Splink’s 7 million users to gain a deeper understanding of bias within their data linkage processes. This, in turn, will enhance the impact and accessibility of Splink, reinforcing its position as a leader in probabilistic data linkage.
The above details our aim for the project, but the exact scope of the project will be tailored to the unique skills of the successful applicant and the duration of their internship.

Overall, the expected outcomes of this research will have far-reaching implications for promoting fairness, accuracy, and efficiency in data linkage within the criminal justice system. By producing a comprehensive report detailing the bias investigation and actionable strategies for mitigation, the research will significantly contribute to the public good, ensuring that decision-making processes are grounded in transparent, unbiased, and ethically sound principles. Moreover, the research will have a profound impact on the ongoing efforts to develop and enhance Splink, making it a vital tool for addressing complex data linkage challenges worldwide.

Supervision and Mentorship
Samuel Atkin – Lead Data Scientist
Samuel works as a lead data scientist across the BOLD programme and the internal Ministry of Justice data linking team. He will provide project direction and management of the intern throughout the project.

Person Specification
- A strong academic background in a relevant field, such as computer science, statistics, data science, or a related area.
- Experience in conducting research and data analysis, preferably in the context of data linkage, machine learning, or fairness in AI/ML algorithms.
- Proficiency in Python programming and a solid understanding of data manipulation and statistical modelling.
- Familiarity with probabilistic data linkage methodologies and algorithms is desired.
- Excellent communication skills, both written and verbal, with the ability to present complex technical concepts to diverse audiences.
- A passion for addressing societal challenges and an interest in promoting fairness and ethics in data science and AI.

Internship Logistics
Start Date: January 2023
Salary: £35,000 per annum pro rata.
Location: London or national hubs

Security Clearance: SC clearance is required as well as training and signing of internal agreements with respect to each data source.

Internship Duration: Full-time for either 3 or 6 months. Part-time for 6 months may be considered.

The intern would join as a civil servant. MoJ policy is to attend the headquarters in London or national hub twice a week. Fully remote working is also an option.
There are strict nationality rules for civil servants. Only the nationals represented in this list can be employed as civil servants:

- UK nationals
- Nationals of Commonwealth countries who have the right to work in the UK
- Nationals of the Republic of Ireland
- Nationals from the EU, EEA or Switzerland with settled or pre-settled status or who apply for either status by the deadline of the European Union Settlement Scheme
- Relevant EU, EEA, Swiss or Turkish nationals working in the Civil Service
- Relevant EU, EEA Swiss or Turkish nationals who have built up the right to work in the Civil Service
- Certain family members of the relevant EU, EEA, Swiss or Turkish nationals