On the Ethics of Data Collection, Analysis, and Modelling in UK Benefit Fraud Detection

UCL Statistical Design and Data Ethics (Spring 2025)

J. Tang

1 Background

In the UK, housing benefit can help you pay your rent if you are unemployed or on a low income. If you receive benefit, then you need to report a change of circumstances[1] for you and anyone else in your house. Examples of housing benefit fraud are not reporting all income or not reporting a change of income.

In a large city in the UK, the manager who deals with housing benefit in the city wants to use data science to help identity fraud. The manager's idea is to use data from past receivers of housing benefit who were investigated for fraud. Assume that for these people individual information is available on whether or not fraud was detected. The manager envisages using the data to define a statistical prediction model and next to use this model to identify current receivers of housing benefit who are likely to commit fraud.

You are asked to lead this project. The main statistical parts of the project are: collecting relevant data, data analysis, defining a model that can be used for prediction, and using the model to make a prediction for current receivers of housing benefit. Assume that the chosen prediction model is a logistic regression model for a binary response variable with value 1 for fraud and value 0 otherwise.

In this report, we will discuss the moral dimensions in data collection, analysis, and decision-making based on predictive models in the detection of benefit fraud. We will discuss the importance of observing and implementing data ethics guidelines at various stages of the project.

[1] GOV.UK. (n.d.). Housing Benefit: Report a change of circumstances. https://www.gov.uk/housing-benefit/report-a-change-of-circumstances.

2 Data Collection

The data subjects in this scenario are the past receivers of housing benefit who were investigated for fraud, and whose data will be used for data analysis and predictive modelling. The model subjects are the current recipients of housing benefit, on whom the logistic regression model will be applied.

First, we must decide what data to collect from data subjects. We should always only collect data that are strictly necessary for the prediction of the likelihood of misreporting or falsification of personal details in benefit claims. Explanatory variables of interest might include age, household income, tenancy information (e.g. address or type of accommodation), pension history and marital status. We should be mindful of what should not be collected. For instance, should we collect data on race and ethnicity, gender, or criminal records? If past investigations on the suspicion of fraud disproportionately targeted certain groups, the data will reflect those biases and the model will inherit and reinforce them. Biased data will lead to biased output (garbage in/garbage out).

Another crucial aspect of data collection is privacy protection. In the UK, data protection is governed by the UK General Data Protection Regulation[2] (UK GDPR) and the Data Protection Act 2018[3]. To ensure **Fairness 2**: acceptable treatments of privacy aspects, we must obtain consent from data subjects before collecting and using their personal information.

Article 13 of the UK GDPR lays out the "right to be informed" requirements when you collect personal data directly from the individual it relates to[4]. The data subject have the right to to be informed if their data is being collected and have the right to refuse our data collection (or the right to have their data erased). The data subject must be informed how their personal data will be used, stored, or shared. The purpose and scope of data usage should be explicitly stated in the consent agreement. A person applying for a housing benefit claim should not later discover that her data is being used to develop a fraud prediction model without her knowledge.

3 Data Processing and Analysis

Article 5 of the UK GDPR outlines seven core principles of data protection, one of which is purpose limitation. This principle allows for the reuse of existing personal data for research-

[2] European Parliament. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council*. https://www.legislation.gov.uk/eur/2016/679/contents.

[3] UK Parliament. (2018). *Data Protection Act 2018*. https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted.

[4] Information Commissioner's Office. (n.d.). Collecting personal data. https://ico.org.uk/for-organisations/direct-marketing-and-privacy-and-electronic-communications/guidance-for-the-use-of-personal-data-in-political-campaigning-1/collecting-personal-data/.

related purposes, provided that appropriate safeguards are in place[5]. In this project, our lawful basis for data processing is public task, i.e. the processing is necessary for us to perform a task in the public interest or for official functions.

Safeguards must be implemented in data storage and processing. We can set up physical barriers to protect personal and sensitive information from data breaches. Protection by perturbation can effectively remove personal identifying information to reduce the risk of misuse. We should control sharing and grant only data scientists responsible for modelling the access to sensitive data. This is particularly important if data such as criminal records are involved.

Transparency 1 states that we should aim clarity in the data science process. One important question to consider is whether the data collected for analysis remains relevant to draw valid statistical conclusions today. There may be a cohort effect. Fraud patterns change over time and correlations observed in past benefit receiver data may no longer hold.

4 Modelling and Deployment

As stated in *Fairness 1*, modelling should avoid discrimination against sensitive groups. When selecting explanatory covariates for the model, once again we should consider carefully whether to include variables such as gender, education, and race. While there may be correlations, including such variables could lead to making an unjust or prejudicial discrimination against sensitive groups. We should be aware of such potential discrimination in the process of modelling.

How about a benefit receiver who was convicted of criminal offence(s) in the past? Should past offences impact our predictions? This is an ethical dilemma that is worth further consideration.

Even if sensitive variables are dropped out, the model may still indirectly discriminate by making statistical inference that disproportionately assigns certain groups with higher weights, which are otherwise untrue or mischaracterised.

Regression models come with some measure of uncertainty, such as estimated standard errors or confidence intervals. This raises another series of questions:

- Does the model output directly trigger the decision to suspend a benefit claim? Or,
- does it provide a reason for officials to conduct further investigation and make a case-by-case argument for approval/suspension of individual claims?

[5] Information Commissioner's Office. (n.d.). Principles and grounds for processing. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/the-research-provisions/principles-and-grounds-for-processing/.

Using model predictions as the only basis for decisions in a penal system risks penalising innocent people and sensitive groups unfairly as a result of false classification.

Logistic regression produces binary outcome. The nature of logistic regression model makes the effect of existing bias on output more prominent. The model may be systematically biased towards/against a model subject based on their marital status (e.g. single or married), demographic variables (e.g. race and ethnicity), immigration status and etc. By assigning 1 or 0 to the model subject, it leaves little wiggle room for human intervention to improve the process fairness.

Transparency 2 states that the ability to explain decisions made by data science models is important to ensure fairness. While the logistic regression does technically provide interpretable results, it is nonetheless difficult for people responsible for the screening process of benefit claims to justify why one particular model subject is flagged as “risky” in an unbiased and legally compliant way. Unable to show proofs of implementation, upon request, that reflects principles of data ethics and compliance to GDPR will invite litigation. For example, the model subject may sue his local council for unjust decision based on the prediction of a statistical model.

Furthermore, as we deploy the algorithmic decision-making based on logistic regression and gather new data to feed back into the model to update the estimated parameters, we may reinforce existing biases and cause over-fitting to the regression model.

5 Summary

This report provides only a limited discussion of the ethical considerations on the use of data science in benefit fraud detection. We stress the importance of fairness, transparency and accountability at each stage of the project, especially when dealing with personal data. The conclusions we draw from the statistical model can have real impacts on people’s lives. Whether those impacts are good and just hinges on how we approach the use of data. This does not mean we should do nothing with data, but rather that we must engage with it carefully, critically, and with good intentions.

6 Appendix: Declaration

Use of AI Tools. ChatGPT is used for studying the UK GDPR, the Data Protection Act 2018 and other cited documents and for proofreading this report.