Algorithm: DNA Analysis

Concept: **Consequences (Predictability/Discoverability)**

SUBJECT: Algorithmic Impact Assessment of the Forensic Statistical Tool (FST) for Criminal Convictions

TO: Project Manager

FROM: Data Scientist

CC: Project File

Introduction

The Forensic Statistical Tool (FST) developed by us for the purpose of analyzing complex DNA samples to determine the likelihood of the presence of a suspect's DNA in crime scenes is under scrutiny for its potential impact on the accused and on the society as a whole. In response to concerns from several defense attorneys, elected officials, and news organizations (Kirshner, 2017a), we performed an Algorithmic Impact Assessment (AIA) of this tool to investigate the ethical implications of using the algorithm (Treasury Board of Canada, 2020). The following is an overview of the FST as well as the findings of the AIA, which include an analysis of the impacts along with the suggested recommendations for mitigation.

Overview

The FST computes a statistical value known as a likelihood ratio (LR) based on hypothesis testing, which is used in the context of criminal convictions using DNA samples of suspects implicated in crime scenes from amongst a complex mixture of extraneous DNA. The LR value is a statistical measurement of the strength of support for one scenario over another. In this context, the scenario is that a known person (the suspect) contributed to a mixture versus the scenario that an unknown, unrelated person contributed instead (NYC Office of the Chief Medical Examiner, 2016; Watters & Coyle, 2016).

Seeing as how the convictions made in consultation with the LR values put forth by the tool have a severe, long-term impact on the affected parties who have had to serve long (and even life-long) sentences in prison as well as having to bear the dire impacts on their future prospects due to the accrued criminal record, an impact assessment was imperative and long overdue. The findings of the impact assessment could help in appeals and exonerations of criminal convictions if so justified.

Moreover, since the FST is proprietary and the source code has not been released to the public (Kirshner, 2017a), this involves a certain lack of transparency for those being judged by it, giving rise to public distrust. My general recommendation is that the FST and any future tools developed on top of the FST need to undergo the due diligence of a rigorous ethical impact assessment, along with the protocol of not relying solely on the tool for high-impact decisions.

Impacts and Consequences of the FST for Decision-Making

The AIA was enlightening in that the checklist-style questionnaire in the context of the FST initiated discussions about the potential impacts of decisions made from its outcomes and whether any consequences that arose could be predicted and discovered by relevant stakeholders (Sandvig, 2020a). The AIA rated the FST at Impact Level III, which denotes high impact (Treasury Board of Canada, 2020). In other words, the decisions made by the algorithm are difficult to reverse and potentially long-term, with longstanding impacts on the rights and freedom of individuals or communities (Treasury Board of Canada, 2019).

The salient impacts identified by the AIA are:

1. Consequences for the Convicted Individual: Many convictions have been made on the basis of DNA evidence in crime scenes as detected by the FST. These results have been repeatedly disputed by many defense attorneys, elected officials, and news organizations (Kirshner, 2017a). Individuals so convicted faced long sentences in prison and have permanent criminal records, which severely impacts their future prospects. In addition to those convicted using the disputed methods, many defendants may have chosen to plead guilty when they learned prosecutors had DNA evidence (courtesy of the FST) against them. Their cases face significant barriers to reconsideration (Kirshner, 2017b). Moreover, the lack of transparency due to the proprietary nature of the algorithm would negatively impact convicted individuals who cannot dispute its outputs, resulting in a model that they cannot audit (Sandvig et al., 2015; Diakopoulos, 2014). Hence, decisions made based on these heavily disputed results need to have their origins be made accessible in order to be dissected more closely.

2. Consequences for the Communities of the Convicted Individuals: There have been cases where individuals belonging to particular communities have been more likely to be implicated by the FST for conviction, such as the Hasidic Jews (Kirshner, 2017a, 2017b), indicating a certain degree of potential ethnic bias attributed to the algorithm, which could contribute to cumulative disadvantage (Sandvig, 2020b). Such potential problems should have been predicted by the analysts of our organization who created and implemented this tool (Sandvig, 2020a).

3. Consequences for the Criminal Justice System: If criminal convictions are made solely on the basis of the results given by the FST, the dynamic of the criminal justice system would be hugely overhauled. The lack of human involvement, while seemingly reassuring initially due to the assurance that there is no human bias, is actually counterproductive from an ethical standpoint as it could be argued that the algorithm itself could be built based on biased assumptions (Sandvig, 2020c), which form the foundations of the hypothesis testing. This is a testament to the fact that no matter how much AI progresses, AI will always require human input and expertise to operate at its full potential in a way that's ethical, responsible, and safe (Brydon, 2019).

Mitigation Actions and Recommendations

The AIA provided a general framework for assessing our tool's impacts and risks along with mitigation best-practices, including:

1. Peer review, which would enable external and diverse perspectives of impacts, risks and identify additional mitigation strategies.

2. Plain language notice about how the FST works, how it will be used, results of any audits and access to training data to all stakeholders.

3. Protocol indicating that the final decision should be made by an authorized human, such as the Judge, who would take all other evidence into consideration as well and not just the results of the FST.

4. The provision of a meaningful explanation for any decision taken with the assistance of the FST with respect to criminal proceedings.

5. Training courses and documentation on the design and functionality of the FST.

6. Contingency plans in case of the unavailability or malfunctioning of the FST.

While many of the recommendations put forth by the AIA which have not yet been implemented in the system are immensely useful for this purpose, we would additionally include the details of the fair use and interpretation of the results of the FST in our client contracts. The AIA is merely one part of an ongoing and continuous process to ensure that the FST is used responsibly and in an ethical manner. Other methods, such as the What-If Tool (Sandvig, 2020d) and collaborative audits (Sandvig, 2020e) would also be employed in the near future to assess the findings of the tool from an explanatory and ethical standpoint.

References

Brydon A. (2019). Why AI Needs Human Input (And Always Will). *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2019/10/30/why-ai-needs-human-input-and-always-will/?sh=3129d7a15ff7

Diakopoulos N. (2015). Algorithmic Accountability. *Digital Journalism*, 3 (3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Kirshner L. (2017a). ProPublica Seeks Source Code for New York City's Disputed DNA Software. *ProPublica.* https://www.propublica.org/article/propublica-seeks-source-code-for-new-york-city-disputed-dna-software

Kirshner L. (2017b). Thousands of Criminal Cases in New York Relied on Disputed DNA Testing Techniques. *ProPublica.* https://www.propublica.org/article/thousands-of-criminal-cases-in-new-york-relied-on-disputed-dna-testing-techniques

NYC Office of Chief Medical Examiner. (2016). Forensic Statistical Tool. *Forensic Biology Protocols for Forensic STR Analysis.* https://www1.nyc.gov/assets/ocme/downloads/pdf/technical-manuals/protocols-for-forensic-str-analysis/forensic-statistical-tool-fst.pdf

Sandvig C. (2020a). Who Discovers Unwanted Consequences? *Coursera*. https://www.coursera.org/learn/siads503/lecture/FjJ2S/who-discovers-unwanted-consequences

Sandvig C. (2020b). Cumulative disadvantage and protected classes. *Coursera*. https://www.coursera.org/learn/siads503/lecture/5yC8Q/cumulative-disadvantage-and-protected-classes

Sandvig C. (2020c). How Transparency Works, and Doesn't Work. *Coursera*. https://www.coursera.org/learn/siads503/lecture/Lz25w/how-transparency-works-and-doesn-t-work

Sandvig C. (2020d). Assignment introduction The What-If Tool. *Coursera*. https://www.coursera.org/learn/siads503/lecture/17SI3/assignment-introduction-the-what-if-tool

Sandvig C. (2020e). Example Technique: External Auditing. *Coursera*. https://www.coursera.org/learn/siads503/lecture/fy5a6/example-technique-external-auditing

Sandvig C., Hamilton K., Karahalios K., and Langbort C. (n.d.). Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. Computational Culture. (2014). *Data and Discrimination: Converting Critical Concerns into Productive Inquiry* (conference). http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf

Treasury Board of Canada. (2019). Directive on Automated Decision-Making. *Government of Canada*. https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

Treasury Board of Canada. (2020). Algorithmic Impact Assessment (AIA) Tool. *Government of Canada.* https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html

Watters K. B., Coyle H. M. (2016). FORENSIC STATISTICAL TOOL (FST): A PROBABILISTIC GENOTYPING SOFTWARE PROGRAM FOR HUMAN IDENTIFICATION. *Jurimetrics*, 56(2), 183–195. http://www.jstor.org/stable/26322658

Algorithmic Impact Assessment Results

Version: 0.9

Name of Respondent

XYZ

Job Title

Data Scientist

Department
Justice (Department of)

Project Title

DNA Analysis using the Forensic Statistical Tool

Project Phase
Implementation [Points: 0]

Please provide a project description:

DNA Analysis, also known as probabilistic genotyping, these systems interpret forensic DNA samples by performing statistical analysis on a mixture of DNA from different people to determine the probability that a sample is from a potential suspect.

What is motivating your team to introduce automation into this decision-making process? (Check all that apply)

Improve overall quality of decisions
The system is performing tasks that humans could not accomplish in a reasonable period of time
Use innovative approaches
Please check which of the following capabilities apply to your system.
Risk assessment: Analyzing very large data sets to identify patterns and recommend courses of action and in some cases trigger specific actions [Points: 0]
Other (please specify) [Points: +1]

Please describe

Statistical hypothesis testing.

Section 1: Impact Level : 3

Current Score: 61

Raw Impact Score: 61

Mitigation Score: 19

Section 2: Requirements Specific to Impact Level 3

Peer Review

At least one of: Qualified expert from a federal, provincial, territorial or municipal government institution. Qualified members of faculty of a post-secondary institution. Qualified researchers from a relevant non-governmental organization. Contracted third-party vendor with a related specialization. Publishing specifications of the Automated Decision System in a peer-reviewed journal. A data and automation advisory board specified by Treasury Board Secretariat.

Notice

Plain language notice through all service delivery channels in use (Internet, in person, mail or telephone). In addition, publish documentation on relevant websites about the automated decision system, in plain language, describing:How the components work;How it supports the administrative decision; andResults of any reviews or audits; andA description of the training data, or a link to the anonymized training data if this data is publicly available.

Human-in-the-loop for decisions

Decisions cannot be made without having specific human intervention points during the decision-making process; and the final decision must be made by a human.

Explanation Requirement

In addition to any applicable legal requirement, ensuring that a meaningful explanation is provided with any decision that resulted in the denial of a benefit, a service, or other regulatory action.

Training

Documentation on the design and functionality of the system.Training courses must be completed.

Contingency Planning

Ensure that contingency plans and/or backup systems are available should the Automated Decision System be unavailable.

Approval for the system to operate

Deputy Head

Other Requirements

The Directive on Automated Decision-Making also includes other requirements that must be met for all impact levels.

Link to the Directive on Automated Decision-Making

Contact your institution's ATIP office to discuss the requirement for a Privacy Impact Assessment as per the Directive on Privacy Impact Assessment.

Section 3: Questions and Answers

Section 3.1: Impact Questions and Answers

Is the project within an area of intense public scrutiny (e.g. because of privacy concerns) and/or frequent litigation?
Yes [Points: +3]

Are clients in this line of business particularly vulnerable?
Yes [Points: +3]

Are stakes of the decisions very high?
Yes [Points: +4]

Will this project have major impacts on staff, either in terms of their numbers or their roles?
Yes [Points: +3]

Will you require new policy authority for this project?
Yes [Points: +2]

The algorithm used will be a (trade) secret
Yes [Points: +3]

The algorithmic process will be difficult to interpret or to explain
Yes [Points: +3]

Does the decision pertain to any of the categories below (check all that apply):
Other (please specify) [Points: +1]

Please describe

Public safety and law enforcement.

Will the system only be used to assist a decision-maker?
Yes [Points: +1]

Will the system be replacing a decision that would otherwise be made by a human?
No [Points: +0]

Will the system be replacing human decisions that require judgement or discretion?
No [Points: +0]

Is the system used by a different part of the organization than the ones who developed it?
No [Points: +0]

Are the impacts resulting from the decision reversible?
Difficult to reverse [Points: +3]

How long will impacts from the decision last?
Impacts can last years [Points: +3]

Please describe why the impacts resulting from the decision are as per selected option above.

Impacts involve prison sentencing.

The impacts that the decision will have on the rights or freedoms of individuals will likely be:
Very high impact [Points: +4]

Please describe why the impacts resulting from the decision are (as per selected option above).

Impacts involve prison sentencing.

The impacts that the decision will have on the health and well-being of individuals will likely be:
Moderate impact [Points: +2]

Please describe why the impacts resulting from the decision are (as per selected option above)

Impacts involve prison sentencing.

The impacts that the decision will have on the economic interests of individuals will likely be:
High impact [Points: +3]

Please describe why the impacts resulting from the decision are (as per selected option above)

Impacts involve prison sentencing.

The impacts that the decision will have on the ongoing sustainability of an environmental ecosystem, will likely be:
Little to no impact [Points: +1]

Please describe why the impacts resulting from the decision are (as per selected option above)

Impacts involve prison sentencing.

Will the Automated Decision System use personal information as input data?
Yes [Points: +4]

Have you verified that the use of personal information is limited to only what is directly related to delivering a program or service?
Yes [Points: +0]

Is the personal information of individuals being used in a decision-making process that directly affects those individuals?
Yes [Points: +2]

Have you verified if the system is using personal information in a way that is consistent with: (a) the current Personal Information Banks (PIBs) and Privacy Impact Assessments (PIAs) of your programs or (b) planned or implemented modifications to the PIBs or PIAs that take new uses and processes into account?
No [Points: +1]

What is the highest security classification of the input data used by the system? (Select one)
Classified / Confidential [Points: +2]

Who controls the data?

Federal government [Points: +1]

Will the system use data from multiple different sources?
Yes [Points: +4]

Will the system require input data from an Internet- or telephony-connected device? (e.g. Internet of Things, sensor)
No [Points: +0]

Will the system interface with other IT systems?
Yes [Points: +4]

Who collected the data used for training the system?
Another federal institution [Points: +2]

Who collected the input data used by the system?
Another federal institution [Points: +2]

Will the system require the analysis of unstructured data to render a recommendation or a decision?
No [Points: 0]

Section 3.2: Mitigation Questions and Answers

Internal Stakeholders (Strategic policy and planning, Data Governance, Program Policy, etc.)
No [Points: +0]

External Stakeholders (Civil Society, Academia, Industry, etc.)
No [Points: +0]

Do you have documented processes in place to test datasets against biases and other unexpected outcomes? This could include experience in applying frameworks, methods, guidelines or other assessment tools.
No [Points: +0]

Is this information publicly available?
No [Points: +0]

Have you developed a process to document how data quality issues were resolved during the design process?
No [Points: +0]

Is this information publicly available?
No [Points: +0]

Have you undertaken a Gender Based Analysis Plus of the data?
No [Points: +0]

Is this information publicly available?
No [Points: +0]

Have you assigned accountability in your institution for the design, development, maintenance, and improvement of the system?

No [Points: +0]

Do you have a documented process to manage the risk that outdated or unreliable data is used to make an automated decision?
No [Points: +0]

Is this information publicly available?
No [Points: +0]

Is the data used for this system posted on the Open Government Portal?
No [Points: +0]

Does the audit trail identify the authority or delegated authority identified in legislation?
Yes [Points: +1]

Does the system provide an audit trail that records all the recommendations or decisions made by the system?
Yes [Points: +2]

Are all key decision points identifiable in the audit trail?
Yes [Points: +2]

Are all key decision points within the automated system's logic linked to the relevant legislation, policy or procedures?
Yes [Points: +1]

Do you maintain a current and up to date log detailing all of the changes made to the model and the system?
Yes [Points: +2]

Does the system's audit trail indicate all of the decision points made by the system?
Yes [Points: +1]

Can the audit trail generated by the system be used to help generate a notification of the decision (including a statement of reasons or other notifications) where required?
Yes [Points: +1]

Does the audit trail identify precisely which version of the system was used for each decision it supports?
Yes [Points: +2]

Does the audit trail show who an authorized decision-maker is?
Yes [Points: +1]

Is the system able to produce reasons for its decisions or recommendations when required?
No [Points: +0]

Is there a process in place to grant, monitor, and revoke access permission to the system?
Yes [Points: +1]

Is there a mechanism to capture feedback by users of the system?
Yes [Points: +1]

Is there a recourse process established for clients that wish to challenge the decision?
No [Points: +0]

Does the system enable human override of system decisions?
No [Points: +0]

Is there a process in place to log the instances when overrides were performed?
No [Points: +0]

Does the system's audit trail include change control processes to record modifications to the system's operation or performance?
Yes [Points: +2]

Have you prepared a concept case to the Government of Canada Enterprise Architecture Review Board?
No [Points: +0]

If your system involves the use of personal information, have you undertaken a Privacy Impact Assessment, or updated an existing one?
No [Points: +0]

Have you designed and built security and privacy into your systems from the concept stage of the project?
No [Points: +0]

Is the information used within a closed system (i.e. no connections to the Internet, Intranet or any other system)?
Yes [Points: +1]

If the sharing of personal information is involved, has an agreement or arrangement with appropriate safeguards been established?
Yes [Points: +1]