Individual Assignment

2IX30 Responsible Data Science

Submission Deadlines
Preliminary Report (1): **February 8, 23:59 CET**
Preliminary Report (2): **February 22, 23:59 CET**
Final Report: **March 1, 23:59 CET**

Overview In this homework assignment, you will practice identifying risks and assessing trade-offs between different desiderata of machine learning systems.

Submission

The assignment consists of two parts: (1) *automated resume screening* and (2) *early eye disease diagnosis*. Each part is complemented by a **group discussion** during class. Before each group discussion, you have to submit a preliminary version of your report for that part of the assignment. In this way, we encourage you to think deeply about the questions beforehand and avoid "free riding".

Please submit your reports on Canvas. The deadlines are as follows:

- Preliminary Report (1): **February 8, 23:59 CET**
- Preliminary Report (2): **February 22, 23:59 CET**
- Final Report: **March 1, 23:59 CET**

Grading

The preliminary reports are ungraded and you are allowed to change as much as you like after the group discussion. However, **submitting the preliminary report and participation in the group discussion is mandatory for successful completion of the assignment**. If you do not submit your preliminary report in time or do not participate in the group discussion, you will **not** receive points for that part of the assignment. *If it is clear no serious attempt was made at writing the preliminary report, this will be treated similarly as not submitting the report at all.*

The maximum number of points for this assignment is 20. Your grade for this assignment is equal to the number of points you've acquired divided by 2. A more detailed rubric is available on Canvas.

Report Guidelines

- The maximum number of pages of the final report is **4** (excluding references). *Please note that longer is not always better.* Make sure that you can read and understand what you wrote.
- Use your own wording. If you use other sources, provide appropriate referencing or citing. **Do not copy whole sentences from websites, articles, books, or your peers.** Reports will be checked for plagiarism.

1 Automated Resume Screening (10 points)

Scenario Description. *You work as a data scientist at a Dutch bank. A startup has approached the bank's human resources (HR) department regarding an automated resume screening tool they have recently developed. The system, called AutoHire, is designed to automatically screen and rank resumes for a job opening. AutoHire works as follows. First, a recruiter can fill out the job's requirements and upload applicants' resumes. Subsequently, the system ranks the applicants based on how well the applicant's characteristics match with the job requirements.*

The startup's sales employee asserts that the HR department should have confidence in the usefulness and effectiveness of the tool, because it is based on sophisticated AI algorithms trained on large amounts of data. Over time, the startup states, the tool will learn from hiring practices which kind of applicants the bank generally interview or hire, improving the tool's effectiveness even further. Moreover, the sales employee emphasizes that the system minimizes employment discrimination compared to traditional human resume screening, because it matches applicants based on relevant data and does not take into account factors such as gender, age, ethnicity, and religion.

On average, the bank receives more than 200 applications for each job position. Currently, the initial resume screening is done by a team of recruiters. The HR director is considering adopting AutoHire, because it could significantly reduce the amount of time spent on screening resumes.

Before deciding on whether to buy AutoHire, the HR director asks the data science department for a recommendation.

In your report, please answer the following questions:

a. *(1 point)* List all stakeholders of this scenario and give a short description of each.

b. *(2 points)* What are potential benefits of using *AutoHire* in this scenario? For each benefit, describe the value (business, legal, or moral) the system brings and to which stakeholder this is beneficial.

c. *(3 points)* What are the risks of the use of this system in this scenario? In your report, include a table in which you explain for each risk: (1) the potential harm, (2) which stakeholder is impacted, (3) which value is at stake, (4) the severity of the harm, and (5) the likelihood of the harm.

d. *(1 point)* Briefly list any assumptions you have made in your risk analysis and how you would validate these assumptions in reality. In particular, consider assumptions regarding different stakeholders' interests.

e. *(3 points)* Based on your risk analysis, provide a recommendation to the human resources department regarding the use of *AutoHire*

   If you believe the HR department should use the system, argue why **and** describe the circumstances under which the system's use would be appropriate.

   If you believe the HR department should *not* use the system, argue why the risks outweigh the benefits **or** describe which conditions must be met before the system can be used.

2 Early Eye Disease Diagnosis (10 points)

Scenario Description. *You are a data scientist at a teaching hospital in the Netherlands. As part of a research collaboration with several teaching hospitals, your team is asked to develop a prototype for an early eye disease diagnosis tool, based on machine learning.*

Currently, digital retinal scans are used by ophthalmologists to diagnose and decide on treatment for several eye diseases, including diabetic retinopathy, age-related macular degeneration and glaucoma. Early diagnosis of eye diseases is critical, as it increases the likelihood that people's sight

can be saved. However, interpreting retinal scans is a complex task. As the number of experts who can interpret them is limited, diagnosis and treatment may be delayed.

The purpose of the envisioned system is to assist ophthalmologists by automatically analyzing retinal scans for common eye diseases. Ultimately, this could result in greater consistency, quality of care, and help avoid cases of preventable eye diseases.

The data will be collected at the five participating teaching hospitals. As this is a costly undertaking, your manager has asked your team to first evaluate the risks and benefits of the development and use of the system. In particular, your manager is interested in how any risks may be mitigated through responsible design of the machine learning task.

a. *(1 point)* List all stakeholders of this scenario and give a short description of each.

b. *(1 point)* What are potential benefits of the envisioned system, in this scenario? For each benefit, describe the value (business, legal, or moral) the system brings and to which stakeholder this is beneficial.

c. *(2 points)* What are the risks of the development and use of the envisioned system in this scenario? In your report, include a table in which you explain for each risk: (1) the potential harm, (2) which stakeholder is impacted, (3) which value is at stake, (4) the severity of the harm, and (5) the likelihood of the harm.

d. *(4 points)* What design choices can be made in order to mitigate (some of) the identified risks? For example, you can consider the machine learning task formulation, the chosen success metrics, the candidate machine learning algorithms, or additional methods/technologies.

   In your answer, be specific about what the design choice is supposed to achieve in this particular scenario. For example, when considering explainability, explain what is to be explained, to whom, and which approaches might be used to achieve it.

e. *(1 point)* List any assumptions you have made in your risk analysis and how you would validate these assumptions in reality. In particular, consider assumptions regarding different stakeholders' interests.

f. *(1 point)* Based on your risk analysis and mitigation strategies, provide a recommendation to the hospital director regarding the development of the early eye disease diagnosis system.

   If you believe the hospital's data science team should continue developing the system, argue why **and** describe the circumstances under which the system's development would be appropriate.

   If you believe the hospital's data science team should *not* develop the system, argue why the risks outweigh the benefits **or** describe which conditions must be met before the system can be developed.