1 2 9 0 FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE D COIMBRA
dei25 departamento de engenharia informática 1995 – 2020

Case Study #2 - Segurança e Privacidade

Mariana Lopes Paulino, 2020190448

In the 21st century, Artificial Intelligence has grown and is becoming better each day that goes by. One of its fields, **Machine Learning,** has a **big impact in today's activities** since it can help in the most various areas from being able to diagnose people to translating languages in real time.
With this growth, comes various concerns, such as **fairness** and **privacy**. These two concepts are very important in ethical aspects when leading with data and its processing.

Fairness, in a Machine Learning context, is an issue that's gaining a lot of attention from the Artificial Intelligence world. This is due to the inherent potential for machine learning systems to exaggerate existing societal biases, as these **biases may be embedded in the data** used to train these models. In order to address this challenge, several approaches have been proposed to **mitigate bias** and **promote equity for all people**, regardless of their characteristics or background.
An example of how to achieve fairness in Machine Learning is to pay close attention to features that can cause bias in the models, and to be aware and cautious when continuously reviewing and refining the feature selection process. These precautions are needed because some **features can be impacting the models incorrectly and leading to cloudiness** around the results. Allied to this suggestion, there are some more things that can be done.

Simultaneously, there needs to be **Privacy** when talking about Machine Learning and its models, As well as fairness this concept is also critical in this field because there's an ethical and legal obligation to protect and handle people's data responsibly.
To achieve this important goal in Machine Learning, there are quite a number of solutions that can be applied. One is **developing Secure Multiparty Computation techniques**, this allows multiple parties such as A and B to jointly compute a function over their inputs while keeping them private, **enabling collaboration while keeping data private**. Apart from this, we can also do Data Minimization, that is when we only **collect the minimum amount of data possible** avoiding collecting sensitive information.

There are several situations where these two concepts correlate:

- **Explainability**: Crucial to ensure fairness and privacy. When models are not interpretable, it can be really difficult to understand and interpret the bias, which is a problem that raises fairness and privacy issues,
- **Data Anonymization**: This technique is known to be a privacy preserving technique which removes identifiable information which protects privacy. In terms of fairness, this process needs to not mask any possible occurring bias,

These are just two examples of many that relate to **Fairness** and **Privacy**.

These two topics are very highly correlated in today's world because Machine Learning is evolving and is now capable of doing almost a bit of everything. Taking **facial recognition** as a concrete example. Supposedly, **Facial Recognition Systems** are used to identify people from their faces. However, these systems have shown to be biased against certain races, or even gender, this leads to the necessity of evaluating fairness as a metric when developing their systems and testing them.

As a conclusion, with all of these topics being mentioned and thoroughly, there is a clear and noticeable relationship between fairness and privacy that is applied when talking about Machine Learning. These topics can be really sensitive as they can affect loads of people, specially if Machine Learning systems don't maintain privacy and fairness with their users.

References:
[1] - Sushant Agarwal: Trade-Offs between Fairness and Privacy in Machine Learning
[2] - Jun Yu, Xinlong Hao, Haonian Xie, and Ye Yu: Fair Face Recognition Using Data Balancing, Enhancement and Fusion
[3] - Cleo Matzken, Steffen Eger, Ivan Habernal: Trade-Offs Between Fairness and Privacy in Language Modeling
[4] - Alex Serban, Koen van der Blom, Joost Visser: Enforce Fairness and Privacy, https://se-ml.github.io/best_practices/06-responsible_ml_ai/, 16 December 2023