DISCRIMINATION AND BIAS IN THE WORLD OF MACHINE LEARNING

APPLIED MACHINE LEARNING

SHREYA BANERJEE

PROFESSOR BARNY SANCHEZ

ABSTRACT

In the world we live in, machine learning models have slowly taken over majority of the jobs of people and cars and will tend to take on more than we have estimated yet. The importance of biases in a machine learning model plays a vital role however it has its own limitations. In this paper the problem of discrimination and bias in machine learning and AI will be discussed in detail with ways to solve the issue of biases in machine learning models.

INTRODUCTION

Machine Learning is field in which artificial intelligence related experience of computer algorithms are fed with data and learned through involving repetition of data manipulation-based outcomes. There are various techniques which can be used to develop machine learning models. Biases simply means weight which in term of machine learning is placed on advantageous group for assumptions in the system over some disadvantageous group. Biases play a very important role in a machine learning model, as it aids into model generalization making it less prone to sensitivity towards lower data group points. The presence of biases in the training data i.e., over or under sampling or loss in labels make it the most important concern for unwanted biases in a machine learning model which also fail to capture essential irregularities present in a dataset, making it less efficient for execution with reflections of biases.

PROBLEM OF DISCRIMINATION AND BIAS IN ML

The existence of biases in a model can happen due to following reasons:

- Bias is occurred due to lack of appropriate feature sets which is also commonly known as underfitting of model. In other terms, such models can be found to show low variance and high bias. Appropriate features have major possibility to lead to bias detections. In case of larger volume of data covering different scenarios, the biasing issue can be resolved. Precaution should be taken to avoid overfitting such as high variance problems affecting model performance eventually.
- The biases also generally come in picture when a model is designed or trained which can result into cognitive biases or biases can occur by faulty datasets which are used for training the model. The attributes such as race, gender, color, education background, age, marital status, religion, etc. could result in bias

EXAMPLES

The intentional or unintentional discrimination could arise in many cases such as:

- **Hiring Employees:** Because of underlying biases in many large tech companies, it has been found that machine learning model or algorithm used for selection of candidates have been biased towards the gender category of males and against females resulting into resume selection of males predominantly giving unfair opportunities for females to be selected for fair chance of employment.
- **Housing:** There are scenarios where models with a high bias making incorrect predictions of house pricing. The outcome impact both the landlords and prospective buyers with reference to lacking sales opportunities.

COMBAT BIAS

The best practices to reduce bias includes the following steps:

- Selecting larger data which sufficient representative can be used to counteract common types of bias, which includes simple and prejudice bias. For cases when data is relatively small, weights can be used but it must be dealt with caution as it could create unnecessary biases.
- Use additional resources, such as 1-FairML used to check bias in predictive modelling and IBM's AI Fairness 360 Open Source should be used to examine and inspect models.
- To ensure that biases do not leech into the systems, hence close monitoring is required while systems run algorithms.

REFERENCES

1. Kumar, A. (2018, October 9). AI / Machine Learning Bias Explained with Examples.https://vitalflux.com/Ai-Machine-Learning-Bias-Explained-Examples/
2. Rouse, M. (2020, July). Machine learning Bias. https://searchenterpriseai.techtarget.com/definition/machine-learning-bias-algorithm-bias-or-AI-bias
3. DeBrusk, C. (2018, March 26). The Risk of Machine Learning Bias. Https://Www.Oliverwyman.Com/Content/Dam/Oliver-Wyman/v2/Publications/2018/November/Risks-of-Machine-Learning.Pdf. https://www.oliverwyman.com/our-expertise/insights/2018/dec/risk-journal-vol-8/rethinking-tactics/the-risk-of-machine-learning-bias-and-how-to-prevent-it.html