

Yujin Hwang

CS 3600

12/8/23

Mark Riedl

## Project 4 Wrapper

### Question 1:

Since the stakes are high and decisions can have significant consequences, a decision tree presents a better alternative to a deep neural network with 50 hidden layers. First and foremost, decision trees are inherently interpretable, providing transparency in the decision-making process. This transparency is crucial in financial institutions, as it allows stakeholders to understand and trust the model's output. Moreover, decision trees excel in capturing complex decision boundaries, especially when there are non-linear relationships between features. In contrast, a deep neural network with numerous hidden layers may be prone to overfitting and can be challenging to interpret, raising concerns about its reliability and accountability. Additionally, decision trees are computationally efficient during inference, making them well-suited for real-time decision-making scenarios, such as loan approval processes. In a situations where the speed of decision-making is important, a decision tree strikes a balance between accuracy, interpretability, and efficiency, making it the better choice for sensitive tasks like loan risk assessment.

### Question 2:

While removing the sex and race columns from the dataset may mitigate the direct discrimination of them in the machine learning model, it does not guarantee the prevention of discrimination based on race or sex. Discrimination can still occur indirectly through correlations with other features that act as proxies for these removed attributes. Machine learning models can learn intricate patterns and relationships within the data, and if historical biases exist in the dataset. Therefore, the model may still capture and perpetuate those biases even in the absence of explicit race and sex information.

### Question 3:

An instance that I researched is Amazon's AI-driven hiring tool which was developed with the aim of automating the recruitment process. The purpose of the model was to analyze resumes and rank job applicants based on their suitability for various positions within the company. However, the tool ended up causing harm by perpetuating gender biases present in the historical hiring data it was trained on, disadvantaging female candidates by systematically downgrading their resumes based on gender-associated keywords. The algorithm demonstrated a preference for male candidates, penalizing resumes that contained terms associated with women, such as references to women's colleges or certain women's activities. This discriminatory behavior resulted from the AI combing through predominantly male resumes submitted to Amazon over a 10-year period to accrue data about whom to hire.

<https://www.businessinsider.com/amazon-built-ai-to-hire-people-discriminated-against-women-2018-10>

#### Question 4:

The reason for separation of a dataset into training and test sets is to evaluate a model's ability to generalize to new, unseen data. By using distinct subsets for training and testing, we ensure that the model does not overfit the training data and can perform well on previously unseen instances. If all data were used for both training and testing, the model might memorize the dataset instead of learning underlying patterns, leading to poor generalization. The separation allows for a more realistic assessment of a model's performance and its applications for the real world.