Assignment questions

Question 1

PARiS’ lists of suggested applicants closely resembled the lists that would have been drafted by Strategeion’s human HR team. To the extent that PARiS was biased towards a particular kind of applicant, this suggests that the human HR workers were as well. Indeed, it can be argued that PARIS is merely an extension of the human biases already in existence at Strategeion. Are computational biases necessarily worse than human biases in a recruitment context or are humans just as bad, what are advantages and disadvantages for each?

Question 2

Biased data pose a problem for ensuring fairness in AI systems. Given the company's demographics, what could Strategeion's engineers have done to counteract the skewed employee data? To what extent do you think such proactive efforts are the responsibility of individual engineers or engineering teams?

Question 3

When it comes to hiring decisions, do you think there should always be a “human in the loop” to make sure that machine decisions don’t bypass human qualities? Why/why not? What would this entail, considering that we might be dealing with very large numbers of applications when making hiring decisions?

Question 4

Job interview companies Pymetrics and Hirevue implement games from psychology research to more accurately determine the qualities of the applicants. They can also use facial recognition software to detect and judge emotional reactions during interviews. Both companies have statements around their focus on fairness in the interview process and work to eliminate human bias. What do you see as the main potential advantages and disadvantages with hiring interviews conducted this way and do you think such companies will improve the hiring process overall?

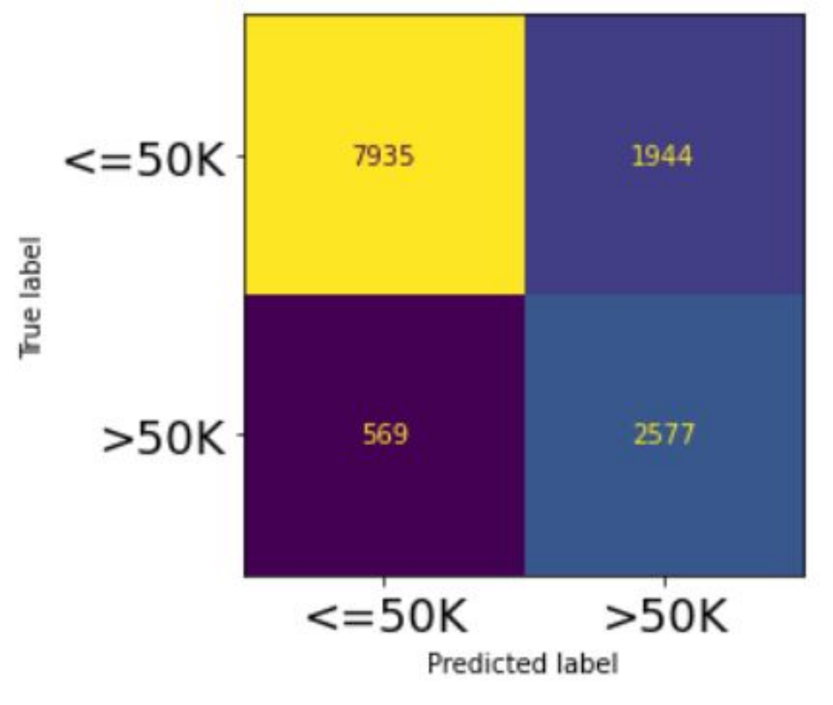
Question 5

We are building a model in order to determine who to approve for a loan based on their predicted income. We are using US Census data that includes information about education, marital status, sex, race, etc. We want to predict who makes makes >50k / year and ensure that our model doesn’t have a bias towards either men or women in the prediction.

On the next few slides you will find the model accuracy and confusion matrix, both overall and for women and men separately. Do you think this model is fairly treating the two groups? Why/why not?

Q5 - Accuracy & confusion matrix

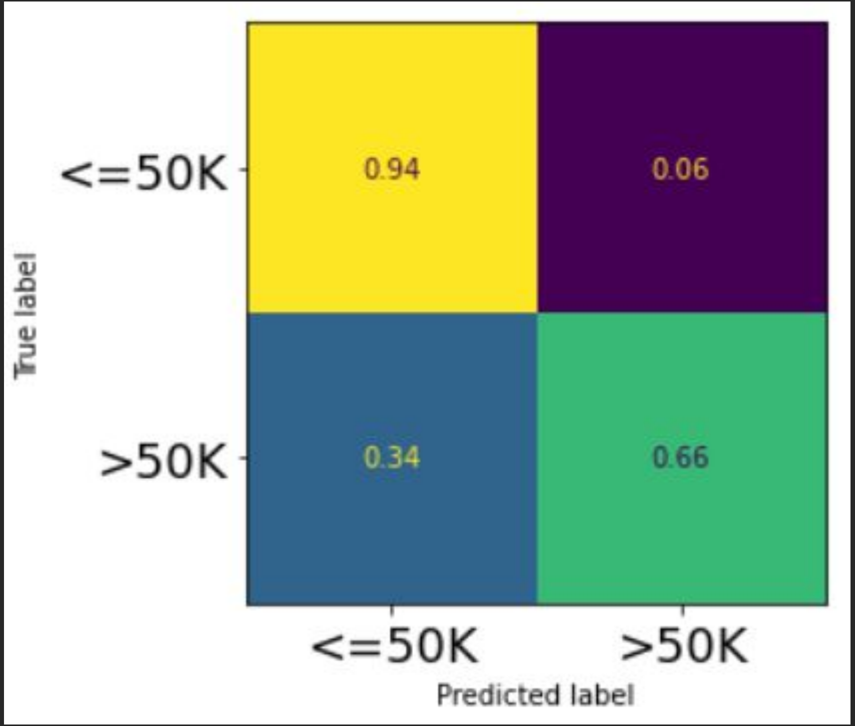
- Overall accuracy ~85%
  - Men ~75%
  - Women ~90%



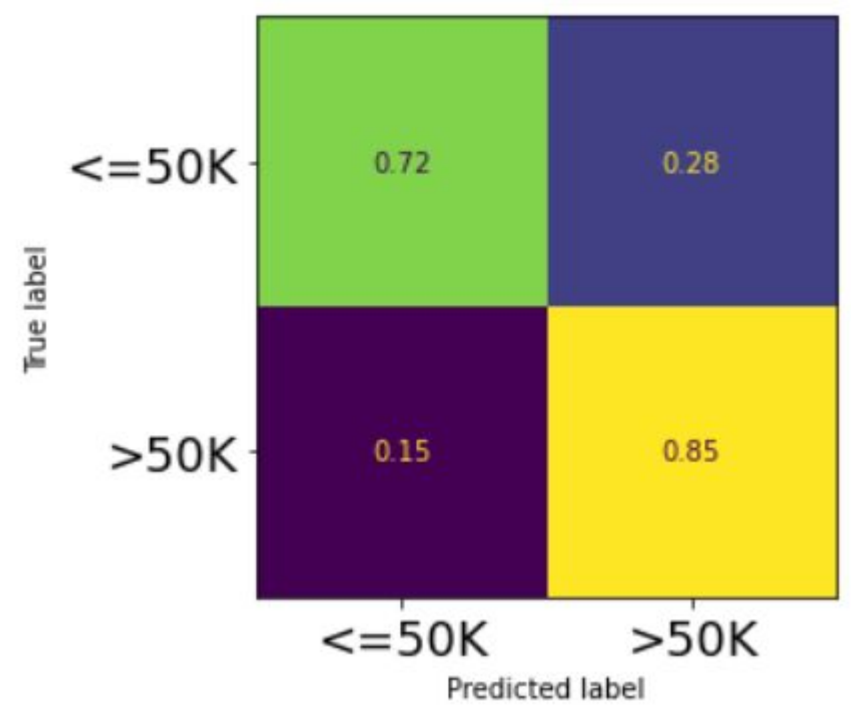
True label \ Predicted label	<=50K	>50K
<=50K	7935	1944
>50K	569	2577

Q5 - Confusion matrix by sex

Women



Men


True label
<=50K
>50K
<=50K
>50K
Predicted label