Your company sells all manner of raw materials to makers who use these materials to sell finished goods on Etsy. Unfortunately, you have discovered that there is a significant ramp-up of fraud in your customer portal. Here, some “customers” exploit a favorite product your company offers, which gives established customers limited-time credit, whereby they can turn over their finished goods on Etsy and pay for the raw materials with the resulting profits. A disturbingly large number of these “customers” default on their payments for your more expensive materials, after first having established a pattern of paying for cheaper materials.

Your company is considering hiring Magenta Consulting to build an AI system to help control this fraud, since you do not have your own data science team. Magenta has delivered a proposal. Knowing that you’ve taken Provost’s class, your boss asks you to come in and advise. She asks you to report back on whether Magenta’s proposal is solid. In particular:

*Are there important flaws or omissions?*

*If so, identify the* ***four most important*** *flaws/issues and offer solutions.*

Magenta’s proposal:

*We will build an AI system for identifying the defrauded accounts. We will first use machine learning to build a fraud classification model. This model will be trained from historical cases of fraud. As labeled data we will use the accounts that have defaulted on their credit. We understand that this is not exactly the same as fraud, as people may default for legitimate reasons, but we believe it will be a suitable proxy. Furthermore, we may want an alert of impending default even if it is not fraud.*

*The AI system will apply the classifier to every prospective account application at inception. If an account application is classified as “fraud”, then the application will be rejected.*

*All parties need to understand that the system will not be perfect: some good accounts will be rejected. The AI system will manage a queue of follow-up requests from these individuals, and customer service can accept them manually. There should be relatively few of these, as the system is expected to be very accurate (>90%).*