1. In under 300 words, describe the two metrics, the trade-off between them, and how you would evaluate a model with hit rate *h* and detection rate *d* in monetary terms.

In insurance industry, the hit rate, also called conversion rate, refers to the percentage of cases that are flagged by a fraud detection system as potentially fraudulent which requires further investigation, out of all the cases that are reviewed. For example, if Shift’s fraud detection system flags 50 cases out of 100 reviewed cases, the hit rate would be 50%.

The detection rate refers to the percentage of actual fraud cases that are correctly identified by a fraud detection system, out of all the fraudulent cases that exist. For example, if there are 100 cases of fraud, but Shift’s fraud detection system only detects 70 of them, the detection rate would be 70%.

Both hit rate and detection rate are important metrics in evaluating the performance of a fraud detection systems in the insurance industry. High hit rates indicate that the system is flagging a high number of cases for further review, which can help identify potential fraud. However, since there are over 10,000 claims in the dataset, a high hit rates may give rise to a high estimated cost of time and money spent on manual investigation. If the detection rate is low, it may indicate that the system is flagging too many cases, including false positives, which can not only lead to additional time and costs, but also result in overlooking the real frauds due to insufficient human capacity or budget.

Conversely, a high detection rate means that the system is correctly identifying a high percentage of actual fraud cases, which can help prevent significant losses for the insurer. This will potentially reduce premiums and raise reimbursement rate which will benefit policyholders and thus help the insurance company to gain a good reputation and expand its customer basis. However, if the hit rate is low, it may indicate that the system is not flagging enough cases for review, which could result in missed opportunities to detect fraud even if the detection rate is high.

Ultimately, insurers aim to strike a balance between the hit rate *h* and detection rate *d* to effectively identify fraudulent cases while minimising false positives and missed opportunities. In this case, our objective is to evaluate a model which can help to flag a moderate but sufficient amount of suspicious cases for further review with a high accuracy which allows us to find the 100 frauds as many as possible.