Exercise 1
Amir Hossein Eyvazkhani
1747696

Task 1)

1. If we have training data that for each review, the positivity (label) is given, then predicting the next texts based on a trained model is going to be a classification (supervised learning) problem. However, if the labels are not given (which is rare in this domain), it will be a clustering (unsupervised learning) problem.
2. Detecting an anomaly (outlier detection) is an unsupervised learning. Because there are no labels for the data, and just by clustering the data from what it is called normal to what is not (outliers) does the job.
3. Because it is a price prediction, we must have training data with previous prices available for the question to make sense. So, it is a supervised learning and because we are predicting a continuous amount, it is called a regression in practice.
4. Because a live bitrate is needed, we going to need a model to be flexible enough to learn from the feedback of the environment, and that's when we use reinforcement learning.
5. In This model, you are trying to classify whether a user will click on an ad (binary outcome: click or no click) based on various features or attributes of the user, ad, and webpage. So, it is a classification and supervised learning.
6. Overall, generative AI is about finding patterns in the given data and producing similar data records. It can be classified as unsupervised (like association analysis).
7. Spam detection is a basic example of classification and is supervised. That is because trains on the given labeled data and predicts based on it.
8. Learning to play a game is reinforcement learning. Because in time the model learns from its actions and uses the feedback as its input.

Task 2)

1. The review can break down to a word occurrence representation vector. Each element in the vector will be treated as a word with a given value of 0 or 1 for occurred/not occurred, respectively.
2. The channel's number of packets per minute, the outbound internet throughput that is being used, the source IP of the inbound requests, and the resources being asked from the incoming requests can be the factors to be considered as the attributes of a network.
3. A house's features can be its location, area of the house (squared meterage), number of rooms, age, availability of parking and basement, number of apartments in the building, availability of a yard, and so on.
4. The state of the environment can be the current connection quality, the state of the video buffer, the location of the user, the download speed of the user, and whether his mouse is moving on the video controls.

5. The features can be the user properties like his tendencies, previous search, previously clicked ads, location, age, gender, and also the webpage and its category, the webpage's popularity, and the ad itself. Like its category and target group properties.
6. The data here can be photos, which can be represented as a matrix of pixels.
7. The email can be represented as a word occurrence matrix and the email sender's location, domain name and also IP can be useful.
8. The state of the game and whether it is going to win or not is a good representer of the environment here.

Task 3)

1. Advertising campaigns can use more sophisticated methods and tricks to send emails. The context of the emails themselves is also variable. Hence, the distribution of data can deviate a lot.
2. The local accent of people who saw the ad in that area is used as the data. Hence, when testing on different, very unusual accents from different parts of the world, the model clearly wouldn't perform well, as the data is different.
3. While this data is good, it is not enough. Predicting stock prices may be affected by numerous environmental, political, financial, and even health-related situations. Hence, the data can deviate a lot, because it comes in a totally different environment. Imagine the impact of Crona virus on stuck markets as an example.
4. First of all, just a front facing camera may not be enough. Also, the data from just two states, does not represent the whole world, and testing in different states can lead to harmful outcomes! Also, even if we test it in the same states, the three months that were used for train data may differ from the time of testing. For example, if you train on rainy days and test in summer, the patterns and hence the distribution of data differs.