

# 2022 Artificial Intelligence: Final Exam Essay

ID and name:

You are a manager at “Awesome Homes,” a real estate company specializing in vacation home rentals. Over the last five years, sales have decreased 10% annually, and customer feedback indicates that your company’s rental prices are seen as too low or too high relative to market value.

The rental prices of Awesome Homes are set through traditional means in which a team of real estate agents monitor various market metrics, and use their expertise to determine rental value. Your CEO has tasked you with managing a new team which will employ machine learning techniques in order to improve the rental price estimates.

1. Many of the real estate agents at Awesome Homes, close personal friends of yours, are worried about being replaced by computers. How can you gain their buy-in on this new initiative?

아래 같은 현상들을 대체할 수 있을 만큼 발전하지 않았으며, 너희들이 걱정하는 컴퓨터에 의해 대체되지 않을 것이다. 또한, 이 새로운 기술의 개발은 오래지 너희들의 일의 도움이 될 것이다. 가격 편지를 낼 때 결정을 도와주는 시장가를 제안하며, 너희들의 배움과 실력 상승에 큰 도움을 줄 것이다.

2. Is this initiative a supervised or unsupervised problem? Explain your reasoning.

Supervised problem 이다. 집의 다양한 데이터, 경쟁 업체에 의해 제시된 해당 집의 임대 가격을 데이터를 하여 수집된 집의 데이터를 통해 임대 가격을 예측하는 문제이며, 시장가를 예측한다.

3. What types of data do you need to gather in order to proceed?

임대, 임대 가격에 관계하는 집의 데이터, 집의 임대 가격을 수집한다. 또한 해당 집의 최대한 많은 정보를 수집한다.

예를 들어, 집의 다양한 크기에 관한 값과 같은 방의 수, 방의 개수와 같은 다른 데이터, 지역, 주소와 같은 카테고리 데이터까지 최대한 많이 수집한다. Your new team presents several variations of machine learning models for estimating rental prices.

4. How do you train and test each model?

임대 가격을 예측한 집의 데이터를 위한 scaling 등의 preprocessing 과정을 거친 뒤 training과 test, validation 데이터로 나눈다.

training 데이터로 모델을 학습시키고, 학습 과정에서 validation 데이터로 성능을 평가한다. 학습이 끝난 뒤 test 데이터로 모델을 평가한다.

5. By what metrics will you determine the quality of the models’ performance?

분류가 아닌 예측 값에 대한 예측, regression 이다. 따라서, 모델의 예측값과 실제 임대 가격 데이터의 차이를 수치로 나타낼 수 있는

MSE (Mean-squared error) 같은 지표를 사용하여 모델의 성능을 측정한다.

6. How will you determine whether your models are coming up with better rental price estimates than were obtained from the traditional reliance on real estate agents’ expertise?

임대 가격에 관계하는 데이터에 대해, 모델의 예측값과 공인중개사의 예측값을 만든다. 모델의 예측값에 대해 실제 시장가와 error 값,

공인중개사의 예측값과 시장가와 error 값을 각각 구하며, 모델의 error 값이 더 작으면 모델의 예측이 더 정확했다는 사실을 정량적으로 입증할 수 있다.

You applied one of your models to the training data, the model’s predictions of rental prices is sound and consistent. However, the model’s success is seriously compromised when new unseen data are entered.

7. What is going on with this model?

Overfitting 이 일어난 것을 가능성이 높다. 주어진 데이터에 대해서만 모델의 예측 성능이 높아지도록 학습이 되었다는 것이다.

이런 경우 새로운 데이터에 대해서는 예측 성능이 일정하지 않거나 좋지 않다.

8. What steps can be taken to improve this model’s prediction capabilities?

Bootstrap Aggregating 을 통해 다양한 모델을 생성하며, ensemble 기법을 도입한다. 또한 다양한 모델을 학습시키는

stacking 기법을 사용한다. 이러한 기법을 사용하면 overfitting에 대해 robust한 모델을 구축할 수 있다.

Your team has assembled three models whose success rates are more or less identical. Here are some observations:

Model	Time for train	Speed of prediction	Average prediction error
A	24 hours	One second or less	5%
B	Six hours	Six hours	1%
C	Five minutes	Two minutes	10%

9. Which model will you recommend to your CEO? Explain your decision.

A와 B 모델 두 가지를 추천할 것이다. A 모델은 가격 산정을 비로 받고자 하는 고객에게 즉각적으로 예측된 값을 보여줄 수 있는 장점으로 활용하고, 몇몇 시간 뒤 정확한 임대 가격 편지를 발송할 수 있다고 알린다. 이후 B 모델을 통해 더 낮은 오차를 예측된 임대 가격을 고객에게 안내한다. 보통의 가계로 한 번의 금액에 변화가 많으므로 A와 B 모델의 긴 훈련 시간은 충분히 정당화 가능하다. 그러나 5%의 오차는 미미한 수준의 영향이 있을 수 있으므로, A 모델은 대표적인 가격만을 안내하고, B 모델의 예측값과 전문가의 계산 및 판단 결과에 따라 제 안내한다.