

# 뉴욕시 택시요금 예측 프로그램

## 1. 모델 구현

### a. 데이터 전처리

utils.py에 있는 함수를 import하여 결측값을 제거하고, 데이터의 이상치를 제거한다.

### b. 신경망 설계

- 1) 입력층 - 256개의 뉴런, ReLU 활성화 함수 사용
- 2) 은닉층 -  $128 \rightarrow 64 \rightarrow 32 \rightarrow 16 \rightarrow 8$ 개 뉴런으로 점진적으로 감소, ReLU 활성화 함수 사용
- 3) 출력층 - 1개의 뉴런, 활성화 함수 없음(회귀 문제)
- 4) 손실함수 - Mean Squared Error(MSE)
- 5) 최적화 기법 - Adam Optimizer

### c. 모델 학습

데이터셋 분할 8:2 = 훈련/테스트 셋

epochs = 1  $\rightarrow$  한번의 반복 학습

## 2. 예측 결과

### 1) 출력 분석

- **RMSE (Root Mean Squared Error):**

- **Train RMSE:** 3.41

- **Test RMSE:** 3.42

테스트 데이터의 RMSE가 학습 데이터와 유사하므로, 모델이 과적합되지 않았음을 확인할 수 있다.

### 2) 랜덤 샘플 예측 결과

#### a. Friday, 17:00hrs:

- 실제 요금: \$12.90

- 예측 요금: \$9.76

- **RMSE:** \$3.14

주말 교통 혼잡 등 추가 요인이 반영되지 않아 약간의 오차 발생

#### b. Sunday, 0:00hrs:

- 실제 요금: \$6.00

- 예측 요금: \$5.33

- RMSE: \$0.67

심야 시간대에 정확히 예측

**c. Wednesday, 13:00hrs:**

- 실제 요금: \$7.50

- 예측 요금: \$7.72

- RMSE: \$0.22

점심시간대에도 높은 정확도를 보여줌

**d. Sunday, 11:00hrs:**

- 실제 요금: \$10.50

- 예측 요금: \$11.40

- RMSE: \$0.90

약간의 오차가 발생

**e. Thursday, 12:00hrs:**

- 실제 요금: \$18.00

- 예측 요금: \$19.01

- RMSE: \$1.01

장거리 이동에 대한 예측도 상당히 정확

**3. 발전 방향**

교통 혼잡시의 요금 데이터를 반영하여 예측 성능을 더욱 극대화 시키고, 심야 할증 등의 데이터를 반영하여, 국내에 적용되는 택시요금 예측 프로그램을 제작할 수 있다.