

방송통신대학교 프라임칼리지 딥러닝 사례연구

중간평가 과제물 설명

작성자: 정지한 겸임교수

한국방송통신대학교 딥러닝사례연구 2025 기말과제

한국방송통신대학교 프라임 칼리지 "딥러닝사례연구" 기말고사 과제 결과물이 옳게 작성되었는지 테스트 해 볼 수 있는 공간입니다. 자세한 설명은 과제 설명 내용을 참고하여 주시기 바랍니다.



과제 설명

- 목표: 은행 고객 데이터 세트를 활용하여 고객의 은행 이탈을 예측하는 것이 본 과제의 목표입니다.
- 데이터: 데이터에는 은행을 떠났거나 계속 고객으로 있는 은행 고객에 대한 몇몇 정보가 포함되어 있습니다. 자세한 내용은 [한국방송통신대학교 딥러닝사례연구 2025 기말과제 | Kaggle 링크](#)를 참조하여 주시기 바랍니다.
 - 은행 고객 이탈 데이터 세트는 은행 업계에서 고객 이탈을 예측하는 데 일반적으로 사용되는 데이터 세트입니다. 여기에는 은행을 떠났거나 계속 고객으로 있는 은행 고객에 대한 정보가 포함되어 있습니다. 데이터 세트에는 다음 속성이 포함됩니다.

데이터 리스트

- Customer ID: 각 고객의 고유 식별자
- Surname: 고객의 성 혹은 이름
- Credit Score: 고객의 신용 점수를 나타내는 숫자 값
- Geography: 고객이 거주하는 국가(프랑스, 스페인 또는 독일)
- Gender: 고객의 성별(남성 또는 여성)
- Age: 고객의 나이
- Tenure: 고객이 은행에 있었던 기간(연도)

- Balance: 고객의 계좌 잔액
- NumOfProducts: 고객이 사용하는 은행 상품 수(예: 저축 계좌, 신용 카드)
- HasCrCard: 고객이 신용 카드를 가지고 있는지 여부(1 = 있음, 0 = 없음)
- IsActiveMember: 고객이 활성 회원인지 여부(1 = 있음, 0 = 없음)
- EstimatedSalary: 고객의 예상 급여
- Exited: 고객이 이탈했는지 여부(1 = 있음, 0 = 없음)

파일 리스트

- **train.csv** - 학습 데이터; Exited는 바이너리(0,1) 형태입니다.
- **test.csv** - 테스트 데이터; 목표는 Exited의 확률을 예측하는 것입니다.
- **sample_submission.csv** - 샘플 제출 파일

3)) 베이스코드 설명

- 시험자료로 제공되는 basecode.ipynb 파일에 대한 설명입니다.
- 베이스코드는 xgboost 머신러닝 알고리즘([XGBoost Documentation — xgboost 3.1.1 documentation](#))을 활용하여 고객 이탈을 예측하는 형태로 간단하게 작성되어 있습니다.

과제 제출을 위한 프로세스

1. <https://www.kaggle.com/t/690688fd13414219acde594424d6d1c5>

에 방문하여 대회 참가를 신청합니다.

The screenshot shows the Kaggle competition page for "한국방송통신대학교 딥러닝사례연구 2025 기말과제". The page includes a navigation bar with links to Overview, Data, Code, Models, Discussion, Leaderboard, and Rules. The Overview section features a timeline from "Start" (18 minutes ago) to "Close" (a month to go). It also displays a "Competition Host" (hey jihan), "Prizes & Awards" (Kudos, Does not award Points or Medals), "Participation" (0 Entrants, 0 Participants, 0 Teams, 0 Submissions), and a "Tags" section with categories like Banking, Beginner, Tabular, Binary Classification, and Roc Auc Score. The "Table of Contents" sidebar lists Overview, 평가, Citation, and Cite.

2. 코드 탭(<https://www.kaggle.com/competitions/knou-deeplearning2025-finalterm/code>)에서 "New Notebook" 버튼을 클릭합니다.
3. 노트북이 새로 커지면, 왼쪽 상단 "File" 버튼 클릭 및 Import Notebook을 클릭하여 제공된 basecode.ipynb 파일을 추가합니다.

The screenshot shows a Jupyter Notebook interface with a menu bar at the top. On the far right of the menu bar, there is a 'Save Version' button with a counter '0'. To its left is a 'Share' button. Below the menu bar, there is a toolbar with various icons for file operations like 'New Notebook', 'Import Notebook', 'Download Notebook', 'Open in Colab', 'Link to GitHub', 'Add input', 'Upload input', and 'Language'. The main area contains a code cell with some Python code related to data processing and file listing. To the right of the code cell is a sidebar titled 'Notebook' which includes sections for 'Input' (with 'Add Input' and 'Upload' buttons) and 'Output' (listing '/kaggle/working').

4. 왼쪽 상단의 "Save Version" 버튼을 클릭하고 나오는 사이드 패널에서 "save" 버튼을 클릭합니다.

The screenshot shows a 'Save version' dialog box overlaid on a Jupyter Notebook. The dialog has fields for 'VERSION NAME' (set to 'Version 1') and 'VERSION TYPE' (set to 'Save & Run All (Commit)'). Below these fields is a note: 'Run a fresh copy of your notebook and save the output'. At the bottom of the dialog are 'Cancel' and 'Save' buttons, with 'Save' being highlighted. The background Jupyter Notebook shows a notebook titled 'notebookf648926315' containing text about a competition and a ' 과제 설명' section with detailed instructions and a link to Kaggle.

5. 왼쪽 하단 창에서 코드가 동작하며, 코드가 정상 완료되었을 경우 캡쳐와 같이 "Successful" 문구를 확인할 수 있습니다. 이 문구를 클릭합니다.

The screenshot shows a history of notebook runs in a Jupyter Notebook. It lists three runs:

- notebookf648926315**: Version #1, Successful. A note next to it says: 'xgboost.readthedocs.io/en/stable/)을 활용하여 고객 이탈을 예측하는 형태로 간단하게 작성되어 있습니다.'
- basecode**: Version #14, Successful. A note next to it says: 'xgboost.readthedocs.io/en/stable/)을 활용하여 고객 이탈을 예측하는 형태로 간단하게 작성되어 있습니다. Kaggle.com/datasets/shubhammeshram579/bank-customer-churn-prediction)를 포함하는 오리지널 데이터를 포함하는 이유는 성능개선의 이유도 있지만, 여러분들이 참고할 코드 대부분이 본 데이터를 활용하기 때문입니다.'
- basecode**: Interactive Session, Running: 3 hours. A note next to it says: '4d77ce7d4ae8c5에 방문하여 대회 참가를 신청합니다.'

At the bottom left, there is a status bar indicating '1 Active Event'.

6. 문구를 클릭하면 동작 완료된 코드의 입력, 출력 및 로그 데이터를 확인할 수 있습니다. 여기서 오른쪽 상단 점3개 버튼(edit 버튼 오른쪽)을 클릭하고, 나온 메뉴에서 "Submit to Competition"를 클릭합니다.
7. 표시되는 사이드 패널 오른쪽 하단의 "Submit" 버튼을 클릭합니다.
8. 아래와 같이 약 20초 안에 정확도가 나오는 것을 확인할 수 있습니다. 여기까지가 제공되는 코드를 기반으로 캐글에서 정확도를 확인하는 기본 프로세스입니다.

Overview Data Code Models Discussion Leaderboard Rules Team Submissions

Submissions

Select up to 5 submissions that will count towards your final leaderboard score. If less than 5 are selected, Kaggle will automatically select from your best scoring submissions. [Learn More](#)

0/5

Auto-selection candidates [?](#)

All Successful Selected Errors Recent ▾

Submission and Description	Public Score ?	Select
 notebookf648926315 - Version 1 Complete · 18s ago	0.89478	<input type="checkbox"/>

- 3) 알고리즘 추가 및 성능 개선을 위한 코드를 작성합니다.
- 4) 코드 작성 완료 후, 모델학습 및 예측이 성공적으로 완료되었으면 노트북 기반(csv가 아닌) 제출을 통해 현재 코드에서 변경없이 0.89 이상의 정확도가 나왔다면 코드 제출까지 문제가 없음을 확인할 수 있습니다.
- 5) 완성된 코드를 pdf로 변환 후 제출하면 완료입니다.

과제 평가 기준

1) 과제 결과물(pdf 파일) 은 평가자 입장에서 전체 복사 후 캐글 내 노트북에 붙여 넣었을 때 에러 없이 동작할 수 있도록 사전 테스트를 진행하고 제출하여 주시기 바랍니다. 코드가 에러가 날 시 이유불문 0점 처리임을 알려 드립니다

2) 아래 명시된 점수는 최대 점수를 고려하여 100점이며 동작하지 않는(에러가 나는) 과제는 0점처리입니다.

3) 과제 점수는 아래와 같은 항목에 따라 0점부터 가산되는 방식이며 총점 100점이 넘는 경우 100점으로 계산됩니다.

- **정확도 향상:** 코드 내용과 관계없이 0.92 이상의 정확도가 나올 경우 40점 가산
- **데이터 전처리 코드 추가:** 데이터 항목의 'Surname','Geography','Gender'를 제거하지 않고 변환하여 사용할 경우 10점 가산
- **데이터 시각화 코드 추가:** 그래프 형태의 시각화 코드 추가될 경우 종류에 상관없이 10점 가산
- **알고리즘 추가:** xgboost 외 lightgbm, catboost 알고리즘 등 추가 시(알고리즘 종류 상관없음) 알고리즘 당 20점 가산. 추가 알고리즘 또한 단일 코드 내에 존재해야 하며(각각의 파일로 제출 불가), 각 알고리즘의 정확도는 0.85 이상의 정확도를 나타내야 합니다)