**Project**

**2018.08.10**

**1. 프로젝트 목표**

“Kaggle”의 대표적인 Classification competition(“Titanic: Machine Learning from Disaster”)의 데이터에 Machine learning 알고리즘을 적용시켜보고, 각 알고리즘의 성능에 대해 비교, 분석해본다.

**2. 프로젝트 설명**

“Titanic: Machine Learning from Disaster”는 타이타닉 호에 탑승한 승객들의 정보와 생존여부를 담고 있는 데이터를 제공한다. 이러한 데이터로 학습한 알고리즘을 이용하여 탑승자의 정보만으로 생존여부를 판별하는 것이 이 competition의 목표이다.

Titanic 데이터는 “PassengerId”(승객번호), “Survived”(생존여부), “Pclass”(티켓등급), “Name”(이름), “Sex”(성별), “Age(나이)”, “Sibsp”(동승한 형재, 자매 수), “Parch”(동승한 부모, 자녀 수), “Ticket”(티켓번호), “Fare”(티켓 가격), “Cabin”(캐비넷 번호), “Embarked”(탑승위치) 총 12가지 정보로 이루어져 있다. 각 정보들은 숫자, 문자, 또는 둘의 조합으로 이루어져 있으며, 이 정보들을 machine learning 알고리즘에 적용시키기 위해서는 적절한 preprocessing을 거쳐야 한다.

다음으로 preprocessing을 거친 데이터를 통해 생존여부(0: 사망, 1: 생존)를 판별하는 classification문제를 다양한 machine learning알고리즘으로 해결한다. 본 프로젝트에서는 5가지 machine learning 알고리즘을 적용시켜보는 것을 목표로 하며, 최종적으로 가장 높은 정확성을 가진 알고리즘을 “kaggle competition”에 제출하는 것을 목표로 한다.

프로젝트 수행 후, 적용한 preprocessing, machine learning 알고리즘, 결과 비교 분석 및 kaggle competition 결과를 작성한 보고서와 작성한 소스 코드를 제출해야 한다.

**3. 보고서**

보고서는 분량 제한 없이 자유롭게 작성하되, 다음 내용을 포함하도록 한다.

* 구현한 preprocessing 설명(**project에 채워 넣은 빈칸 부분 캡처 후 첨부**)
* 6가지 machine learning 알고리즘 **설명** 및 **accuracy, ROC, AUC 비교 분석** 
  + 각 알고리즘의 accuracy, ROC, AUC 결과를 캡처하여 보고서에 첨부한 후, 비교 분석
* Kaggle 최종 스코어 (**캡처하여 제출**)
  + 가장 잘 나온 스코어만 첨부, 해당 스코어를 얻은 preprocessing, machine learning 알고리즘 설명.