

# HAAFOR Challenge 2020

## HAAFOR Research Korea?

HAAFOR Research Korea 는 금융시장 예측과 관련한 데이터를 생산하여 예측 모델(Prediction Model)을 연구/개발하는 R&D 회사입니다. HAAFOR Research Korea 에서는 벤더가 공급하는 금융시장 관련 데이터 외에도 자연어 처리(Natural Language Processing)와 머신 러닝(Machine Learning) 등을 활용하여 자체적으로 데이터를 생산합니다.

## Challenge 2020 개요

HAAFOR Challenge 는 Quantitative trading model 에 활용되는 응용/경제 수학, 계산과학, 기계학습 및 인공지능 분야의 우수한 인재들을 발굴하기 위하여 매년 개최하고 있습니다.

HAAFOR 에서는 금융시장 예측을 위해 다양한 데이터와 기술을 활용하고 있습니다. 그 중에서도 뉴스는 많은 정보를 함축하고 있는 데이터입니다. 시장에서 일어나는 이슈를 빠른 속도로 전달해주는 뉴스는 포트폴리오를 구성하는 데 있어서 중요한 역할을 합니다. 뉴스 Article 마다 담고 있는 정보질이 각각 다르기 때문에, 쏟아지는 수만의 Article 중 중요한 정보를 선별하는 과정은 필수적입니다.

## 대회주제

이번 대회 주제는 뉴스 Article 간의 선후관계 파악입니다. 참가자에게 총 20 만건의 짧은 뉴스 Article 쌍들이 제공되며, 제공된 Article 의 내용을 분석하여 Article 의 발행 순서를 맞추어야 합니다. 한 쌍을 이루는 두 기사 간의 선후관계를 추측할 수 있는 내용을 추출하는 것이 이번 과제의 핵심입니다.

## 1. 최초 데이터 셋 (Training.zip-training.csv)

- 제공된 최초 데이터(*training.csv*)는 **200000x5** 형태의 헤더가 포함된 CSV 테이블이며, 각 열은 다음과 같은 정보를 가지고 있습니다.
  - ① **INDEX**: 데이터 순번
  - ② **BEFORE\_HEADLINE**: "AFTER" Article 이전에 발행된 Article 의 헤드라인
  - ③ **BEFORE\_BODY**: "AFTER" Article 이전에 발행된 Article 의 내용
  - ④ **AFTER\_HEADLINE**: "BEFORE" Article 이후에 발행된 Article 의 헤드라인
  - ⑤ **AFTER\_BODY**: "BEFORE" Article 이후에 발행된 Article 의 내용

Index	BEFORE_HEADLINE	BEFORE_BODY	AFTER_HEADLINE	AFTER_BODY
1	Qualcomm's Win Over Apple in Germany Won't Force Apple to Table	Qualcomm's win over Apple in Germany, which if upheld will block the sale of some older model iPhones in Germany, could reduce Apple's quarterly sales by about \$560 million, in our estimation. It's an important win for Qualcomm and builds on its momentum from a recent injunction in China, yet we don't think a victory in this battle will put enough pressure on Apple to settle the global dispute in the near term.	Qualcomm's Germany Offensive Coming Up Short in Apple Row	Qualcomm is losing steam in Germany with every decision that falls short of entirely blocking Apple's most popular iPhone models from being sold there. So far no judge has gone that far, with a Jan. 15 Mannheim court ruling in Apple's favor the latest setback for Qualcomm. The company sued Apple in Germany because the nation frequently grants broad injunctive relief in patent cases. We think Apple is likely to escape meaningful harm in Germany.

- 작으로 제공된 Article 들은 동일한 본문에서 발췌된 일부 내용이 아니며, 서로 독립된 개별 Article 이 발행 순서에 따라 짝을 이루고 있습니다. (즉, BEFORE Article 은 AFTER Article 보다 먼저 발행된 Article 입니다.)
- 문맥만으로 Article 의 발행 순서를 알기 어려운 경우도 있을 수 있습니다.
- 하나의 동일한 Article 이 다른 Article 과 짝을 지어 여러 번 제공될 수 있습니다.

## 2. 평가문제 (evaluation.csv)

- 평가문제(*evaluation.csv*)는 10 만 개의 뉴스를 최초 데이터셋(*training.csv*)과 같은 형태로 **100000x5** 형태의 헤더가 포함된 csv 파일입니다. (**INDEX, NEWS1\_HEADLINE, NEWS\_1\_BODY, NEWS2\_HEADLINE, NEWS2\_BODY**)
- 단, 최초 데이터셋(*training.csv*)과는 달리, 짝을 이루는 NEWS1 과 NEWS2 가 발행순서에 맞지 않게 나열된 경우도 있습니다. (즉, NEWS1 이 NEWS2 보다 나중에 발행된 경우도 있습니다.)
- 평가문제에 대한 답안(*answer.csv*)은 **100000x3** 형태이며 헤더가 없는 csv 파일로 제출해주시면 됩니다. (**INDEX, NEWS1, NEWS2**)
  - ✓ 먼저 발행된 Article 를 "0", 이후에 발행된 Article 를 "1" 로 기재
  - ✓ 제공된 평가문제와 같은 순서로 행을 나열.
  - ✓ 평가는 순서를 올바르게 예측한 경우를 1 점으로 하여 총 100000 점이 만점이며, 부분점수는 없습니다.

<answer.csv (예시)>

```
1, 0, 1
2, 1, 0
...
100000, 1, 0
```

## 3. 지원 신청 절차

- 참가신청서 접수 기간: 2020 년 7 월 6 일 (월) ~ 2020 년 8 월 19 일 (수)
- 참가신청서 접수 방법: [hfchallenge@haafor.com](mailto:hfchallenge@haafor.com) 으로 참가신청서를 제출
- 참가신청서 및 개인정보수집동의서는 <http://www.haafor.com/challenge> 에서 다운로드 가능합니다.

참가신청서를 제출하신 후 2-3 일 이내로, 최초 데이터셋(*training.csv*)을 받아 보실 수 있습니다.

#### 4. 제출 안내 2020 년 8 월 22 일 (토)

참가자들은 평가문제(*evaluation.csv*)가 공개된 후, 3 시간 이내에 답안을 [hfchallenge@haafor.com](mailto:hfchallenge@haafor.com)으로 제출해야 합니다.

- **평가문제 공개 시각:** 2020 년 8 월 22 일 오전 9 시 (싱가포르 시각)/오전 10 시 (대한민국 시각)  
정각에 평가문제가 이메일로 발송됩니다.
- **답안제출 마감 시각:** 2020 년 8 월 22 일 오후 12 시(싱가포르 시각)/오후 1 시 (대한민국 시각)
- **제출해야 할 답안 (4 가지):**
  - ✓ 평가문제에 대한 답안(*answer.csv*)
  - ✓ 모델에 대한 간략한 서술(자유형태, A4 1 장 이내)
  - ✓ 기계 학습이 완료된 최종 모델을 담고 있는 소스 코드 파일
  - ✓ 기계 학습 과정을 담고 있는 프로젝트 예) Jupyter notebook, 프로젝트 압축파일 등
- 마감시각 이후 또는 미완료된 형식으로 제출된 답안도 평가 대상이나, 감점됩니다.
- 제출 파일들은 공개된 Git Repository 링크로 제출 가능합니다.
- 프로젝트 코드가 같이 제출된 학습모델을 설명할 수 없거나, 정상적이지 않은 방법으로 답안(*answer.csv*)을 생성한 경우 시상에서 제외됩니다.

#### 5. 시상

1~2등: Apple iPad Pro (256GB)  
3~5등: 기계식 키보드  
5명(Random): 고급형 헤드셋

또한, 우수 참가자가 입사 지원할 시, 1 차 전형이 면제됩니다.

#### 6. 기타

- 추가 변동사항은 홈페이지(<http://www.haafor.com/challenge>)에 업데이트 될 예정이니, 수시로 확인 바랍니다.
- 기타 문의가 있을 경우, [hfchallenge@haafor.com](mailto:hfchallenge@haafor.com) 또는 담당자(이현아 02-6415-5703)로 문의주시기 바랍니다.