


# 10-월차 IPP보고서

## ( 10월)

2022년

학부	융합소프트웨어 학부	전공	데이터테크놀로지	학번	6017 1632	성명	김형진
실습기간	2022년 9월 1일 ~ 2022년 12월 31일						
실습기관	㈜코나아이	부서		AI팀			
실습직무	AI 개발	지도 담당자		엄용운 			

※ 반드시 보고서 최종본에 멘토 확인 후 서명 하여 업로드.

## 2-차월 현장실습 일지

기간	주요 실습내용
1주차 (10/3~10/7)	<p>&lt;직무&gt;</p> <ul style="list-style-type: none"> <li>- Crypttrade 불건전 발언 관련 예문 서칭</li> <li>- 욕설/성적 발언 약 2400개 수집</li> <li>- uuid4를 사용하여 각 발언에 고유의 id를 부여하고 json형식으로 저장</li> </ul> <p>&lt;공통&gt;</p> <ul style="list-style-type: none"> <li>- 와이어 프레임 수정사항 종합 및 토의</li> <li>- 와이어 프레임 수정사항 반영</li> </ul>
2주차 (10/10~10/14)	<p>&lt;직무&gt;</p> <ul style="list-style-type: none"> <li>- 데이터 정제</li> <li>- CF 알고리즘 개념 학습</li> <li>- 기본 CF 알고리즘 구현</li> </ul> <p>&lt;공통&gt;</p> <ul style="list-style-type: none"> <li>- 소비자, 가맹점주 요구사항 분석</li> </ul>
3주차 (10/17~10/21)	<p>&lt;직무&gt;</p> <ul style="list-style-type: none"> <li>- Python을 이용한 개인화 추천 시스템 강의 수강 <ul style="list-style-type: none"> <li>● 인기제품 추천 방식</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>● 추천시스템의 정확도 측정</li> <li>● CF 알고리즘</li> <li>● 이웃기반 CF 알고리즘</li> </ul> <p>&lt;공통&gt;</p> <ul style="list-style-type: none"> <li>- Figma Prototype 연결</li> </ul>
<p>4주차</p> <p>(10/24~10/31)</p>	<p>&lt;직무&gt;</p> <ul style="list-style-type: none"> <li>- Python을 이용한 개인화 추천 시스템 강의 학습</li> <li>● 추천 시스템의 평가 지표(Precision, Recall..)</li> <li>● 아이템 기반 CF 알고리즘</li> <li>● 추천 시스템의 정확도를 높이기 위한 파라미터 튜닝</li> </ul> <p>&lt;공통&gt;</p> <ul style="list-style-type: none"> <li>- 최종 발표 준비 및 시연</li> </ul>

## 현장실습 내용 주요 목차

### I. Cryptrade 불건전 발언 관련 업무 보조

- 1, 욕설/성적 발언 데이터 약 2,400개 수집
- 2, .txt파일을 json형식으로 변환

### II. 추천 알고리즘 학습

1. CF 알고리즘에 관한 내용 학습
2. 인프런 – Python을 이용한 개인화 추천 시스템 강의 수강

## 현장실습 주요내용

### I. Cryptrade 불건전 발언 관련 업무 보조

1. 욕설/성적 발언 데이터 약 2,400개 수집
  - 1) 구글링을 통해 약 2,000개, 약 700개의 욕설/성적 발언이 포함된 txt파일 다운로드
2. .txt파일을 json형식으로 변환
  - 1) 두 txt파일을 하나로 합치고 중복되는 내용제거, 특수문자 및 공백을 제거
  - 2) uuid4로 랜덤Id를 생성한 후 Dialogflow의 json형식에 맞추어 json파일 생성 후 저장

### II. 추천 알고리즘 학습

## 1. CF알고리즘에 관한 내용 학습

### 1) CF의 정의

- (1) CF는 Collaborative Filtering의 줄임말로 협업 필터링이라고도 한다.
- (2) 나와 비슷한 취향을 가진 유저들은 어떠한 아이템에 대해 비슷한 선호도를 가질 것이라는점을 가정한다.

### 2) 모델 기반의 접근 방식

- (1) Latent Factor 방식(Matrix Factorization): 아이템 Latent Vector (잠재 벡터)와 유저 Latent Vector 간 Inner Product로 아이템에 대한 유저의 선호도를 모델링하는 방식
- (2) Classification/Regression(분류/회귀) 방식: 콘텐츠 기반 추천과 쉽게 융합이 가능

### 3) 메모리 기반의 접근 방식

- (1) User based CF: 사용자 A와 비슷한 사용자 B를 찾고 B의 item 평가를 기반으로 A가 평가하지 않은 item의 평가를 예측
- (2) Item based CF: 사용자 A가 평가하지 않은 item A와 비슷한 Item B를 찾고 item B의 평가를 기반으로 item A에 대한 평가를 예측

## 2. 인프런 – Python을 이용한 개인화 추천 시스템 강의 수강

### 1) 인기제품 추천 방식

- (1) 개별 사용자 정보가 없을 때 사용할 수 있는 간단한 추천 방식
- (2) Best-Seller 제품을 추천해주는 방식

### 2) 추천 시스템의 성과측정지표

- (1) 각 아이템의 예상 평점과 실제 평점의 차이를 성과측정지표로 사용하고 보통 RMSE를 가장 많이 사용한다.

(2) 추천한 아이템과 사용자 실제 선택과 비교

		예측 결과	
		TRUE	FALSE
실제 정답	TRUE	TP ( True Positive )	FN ( False Negative )
	FALSE	FP ( False Positive )	TN ( True Negative )

3) 이웃기반 CF 알고리즘

- (1) KNN + CF 알고리즘: 사용자와 유사도가 높은 K명의 데이터를 통해 사용자에게 추천하는 방식
- (2) 최적의 이웃 크기 결정: 너무 클 경우에는 전체 집단으로 추천하는 것과 크게 다르지 않고, 반대로 너무 작을 경우에는 편향된 결과로 신뢰성이 떨어지게 됨

4) 사용자의 평가경향을 고려한 CF 알고리즘

- (1) 각 사용자의 평점을 구한 후 평점 편차를 계산
- (2) 평점 편차로부터 예측값을 계산
- (3) 평점 편차 예측값 + 평균값을 통해 실제 예측값을 계산

## 느낀점

어느덧 코나아이에서 인턴 생활을 시작한지도 4개월차다. 팀이 주로 어떤 업무를 맡고 있는지 직장 생활을 어떻게 돌아가는지 알 것 같다. 4개월간 진행한 인턴 공통프로젝트도 2등이라는 만족스럽기도하고 아쉽기도 한 결과로 끝났다. 인턴 기간도 끝나가고 한창 공채 시즌이라서 미래에 대한 고민이 많다.