

# Software Requirement Analysis

- 딥러닝 기반 영화평 감성 분석 -  
[리뷰 해체 분석기(Film Review Dematerializer)]

## Team #4

201411273 박재범

201411275 박진호

201411283 이상민

201511244 김민우

**1. Functional/Non-Functional Requirement**

**2. Architecture Diagram**

**3. Prototype Illustration**

**4. System Test Cases**

## # Functional Requirement

### 1. 인터페이스

#### 1) 사용자 페이지

[메인 페이지]

- (1) 사용자가 Data(200자 이하의 영화평 문장 및 긍정/부정 여부)를 입력할 수 있어야 함.
- (2) '+, -' 버튼 클릭을 통해 추가 Data 입력 또는 삭제가 가능해야 함.
- (3) 작성된 입력들을 제출할 수 있어야 함.
- (4) 여러 개의 Data를 한번에 제출할 수 있도록 Excel 파일 연동 입력 기능이 제공되어야 함.

[결과 페이지]

- (5) 입력을 바탕으로 각 Data에 대한 '긍정/부정 예측', '일치 여부' 리스트와 Data Set에 대한 '정답률(일치율)', '긍정/부정 비율' 그래프, 전체적인 긍정/부정 여부에 대한 결과가 출력되어야 함.

## # Functional Requirement

### 1. 인터페이스

#### 2) 관리자 페이지

[메인 페이지]

(1) 관리자 페이지에 모델 버전 별 '정확도' 가 표 및 그래프로 출력되어야 함.

(2) 관리자가 활용 또는 Fine-Tuning할 모델 버전을 선택할 수 있어야 함.

(3) 현재 선택된 모델을 기반으로 누적 Data Set을 통해 Fine-Tuning하여 새로운 버전을 생성할 수 있어야 함.

## # Functional Requirement

### 2. 언어 모델

#### 1) Pre-Processing

- (1) Raw-Data의 오타, 띄어쓰기, 이모티콘 등을 작업에 최적화된 포맷으로 변환해야 함.
- (2) 변환한 Data를 KoELECTRA의 Vocabulary에 대응시켜 Word-Embedding해야 함.

#### 2) RNN Layer

- (1) KoELECTRA의 Input Sentence에 한국어 감성사전을 적용하여 각 Position에 맞는 가중치 벡터(senti\_w)들을 구성해야 함.
- (2) 한국어 감성사전 적용에 관한 오차 보정을 위해 [CLS] 토큰의 Hidden State를 나머지 토큰들에 Scaled Dot-Product Attention(SDA)해 주어야 함.
- (3) Position이 일치하는 'SDA를 마친 벡터'와 '가중치 벡터'들을 Contatenate하여 Bi-GRU를 통과 시키고, 문맥 흐름 정보를 담고 있는 양 끝의 출력 벡터를 FFNN Layer 입력 차원에 맞춰 Projection 해야 함.

## # Functional Requirement

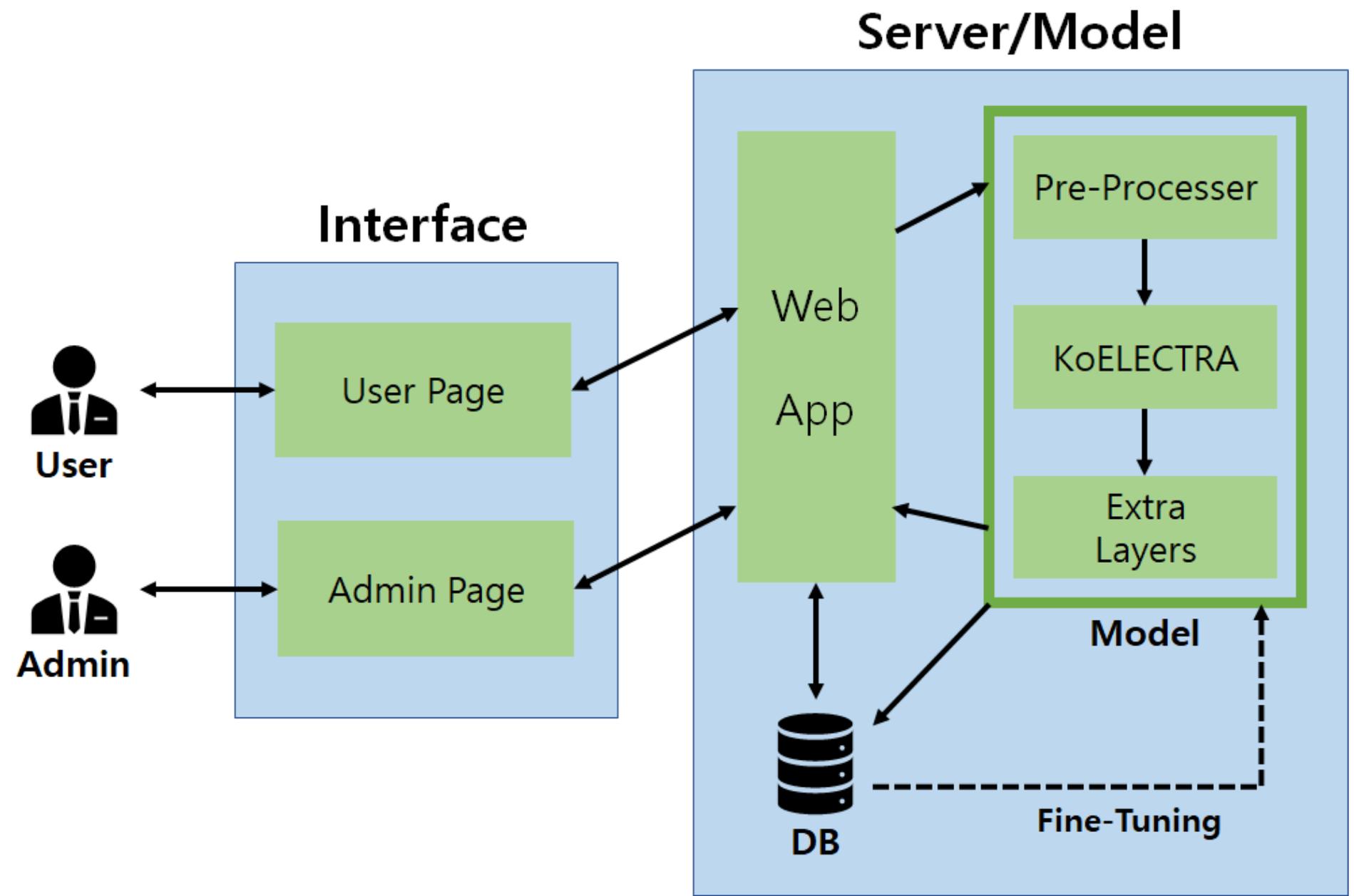
### 2. 언어 모델

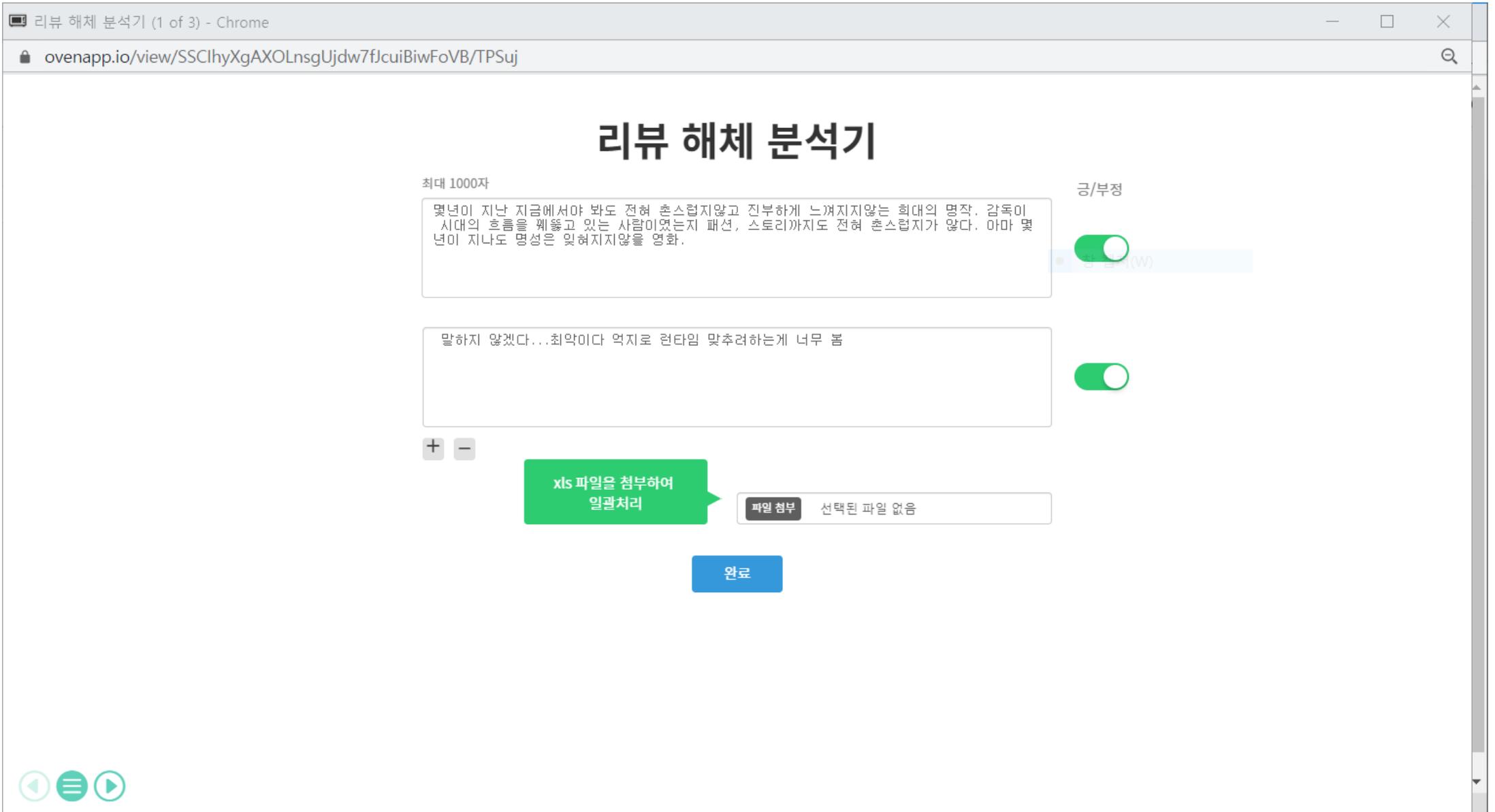
#### 3) FFNN Layer

(1) KoELECTRA의 [CLS] 토큰에 대한 Hidden State와 RNN Layer에서 나온 두 벡터를 FFNN에 통과시켜 2차원 Classification 벡터로 변환하고, 이를 바탕으로 긍정/부정 예측 결과를 출력해야 한다.

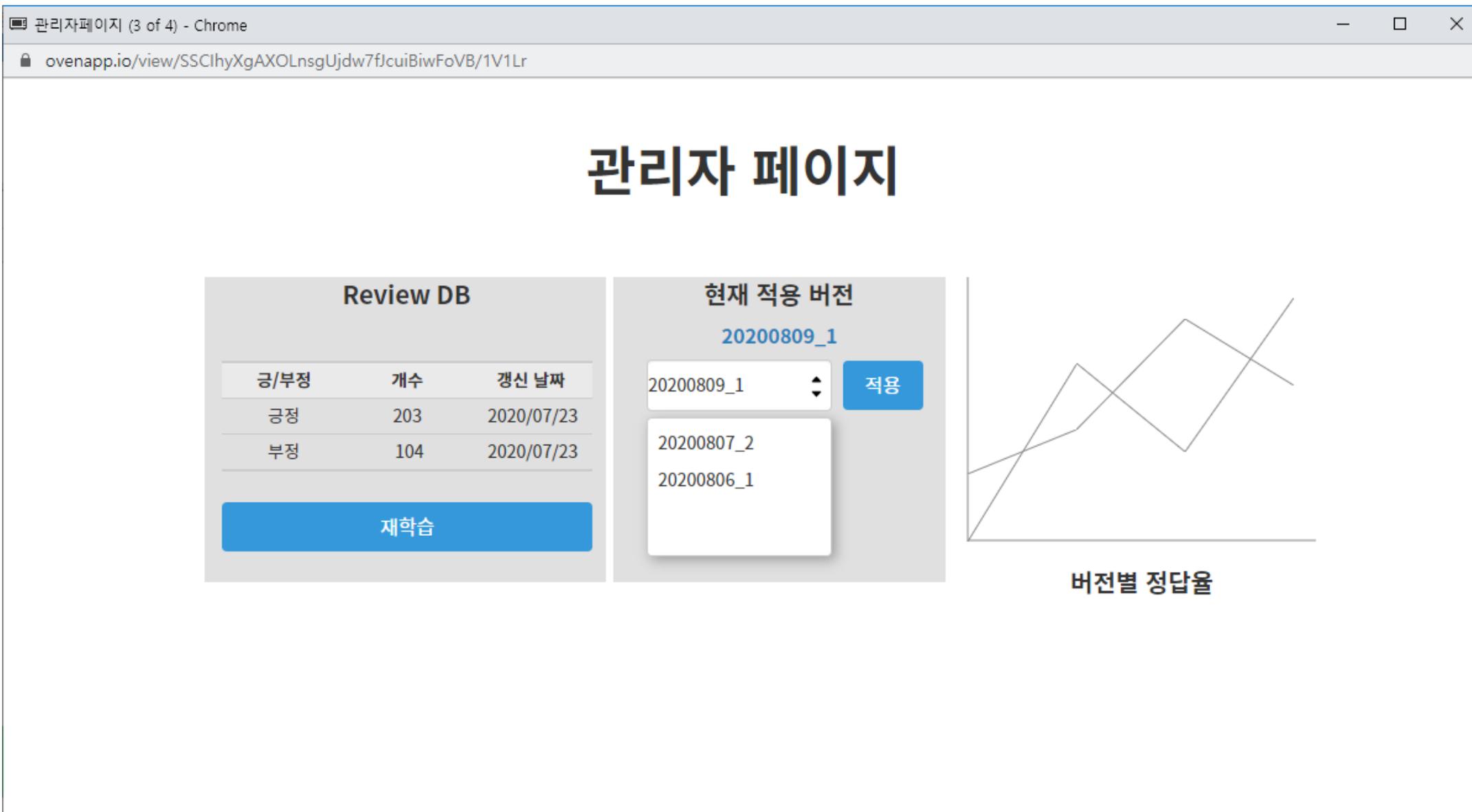
## # Non-Functional Requirement

1. 인터페이스의 명료성과 직관성에 대해 60% 이상의 사용자가 만족해야 함.
2. 입력 제출 후 최종 결과가 출력될 때까지 Input 개수 당 2초 이하의 시간이 소요되어야 함.
3. 표본이 50개 이상일 때 모델로부터 도출된 평가와 사용자의 의도의 일치도가 90% 이상이어야 함.









## 4. System Test Cases

No.	Description	Expected Result
F.1.1.1	사용자 메인 페이지에서 입력 박스에 200자를 초과하지 않는 텍스트 입력이 가능한지, 그리고 스위치 버튼을 통한 긍정/부정 여부 선택이 원활하게 작동하는지 확인.	200자 이상 입력 시 입력 불가, 긍정/부정 버튼 토크 정상 작동.
F.1.1.2	사용자 메인 페이지에서 + 버튼을 눌렀을 때 새로운 입력 박스가 생성되는지, - 버튼을 눌렀을 때 생성된 입력 박스가 삭제되는지 확인.	+ 버튼 클릭 시 새 박스 생성, - 버튼 클릭 시 최하단에 위치하던 박스 삭제.
F.1.1.3	사용자 메인 페이지에서 완료 버튼을 눌러 입력들을 제출할 수 있는지 확인.	결과 페이지가 출력됨.
F.1.1.4	사용자 메인 페이지에서 .xlsx 확장자의 파일 등록 및 제출이 정상적으로 이루어지는지 확인.	포맷에 맞지 않는 파일 등록 시 오류 메시지, 포맷에 맞는 파일 등록 시 정상 등록 및 제출 처리.
F.1.1.5	사용자 결과 페이지에서 제출된 입력들에 대한 분석 결과와 통계들이 정상적으로 출력되는지 확인.	입력 시 선택한 긍정/부정이 일치하며 정상적으로 통계 출력.
F.1.2.1	관리자 메인 페이지에서 모델 버전 별 정확도가 표 및 그래프로 출력되는지 확인.	DB 데이터와 일치하는 정확도 표 및 그래프 출력.
F.1.2.2	관리자 메인 페이지에서 현재 분석에 적용하거나 Fine-Tuning 기반이 될 모델 버전을 선택할 수 있는지, 그리고 선택한 모델이 정상적으로 반영되고 통계가 갱신되는지 확인.	선택 및 적용된 모델 버전으로 분석이 진행, 통계 갱신됨.
F.1.2.3	관리자 메인 페이지에서 선택된 모델을 기반으로 DB의 누적 Data Set을 통해 Fine-Tuning이 가능한지, 결과에 대한 새로운 버전의 모델이 정상적으로 저장되는지 확인.	선택 및 적용된 모델 버전으로 Fine-Tuning이 진행됨, DB 및 리스트에 새 모델 추가.

## 4. System Test Cases

No.	Description	Expected Result
F.2.1.1	모델의 Pre-Processing 단계에서 입력 문장의 오타, 띄어쓰기, 이모티콘 등에 대한 포매팅 작업이 정상적으로 이루어지는지 확인.	설계한 알고리즘에 맞추어 문자열 포매팅이 이루어짐.
F.2.1.2	모델의 Pre-Processing 단계에서 포매팅 작업이 끝난 입력 문장에 대해 KoELECTRA의 Vocabulary에 대응되는 WordPiece 토큰으로 임베딩 되는지 확인.	문자와 대응되는 KoELECTRA의 Vocabulary Index가 정상적임.
F.2.2.1	모델의 RNN Layer 단계에서 입력 문장에 대한 한국어 감성사전 가중치 벡터가 정상적으로 구성되는지 확인.	토큰 개수와 일치하는 가중치 벡터가 생성됨.
F.2.2.2	모델의 RNN Layer 단계에서 각 Hidden State들에 [CLS] 토큰의 Hidden State가 정상적으로 Scaled Dot-Product Attention 되는지 확인.	결과 벡터에 문맥 영향 정보가 반영됨.
F.2.2.3	모델의 RNN Layer 단계에서 Position이 일치하는 'SDA를 마친 벡터'와 '가중치 벡터'들이 Contatenate 되는지, Bi-GRU 출력 벡터가 FFNN Layer에 맞는 차원의 벡터로 Projection 되는지 확인.	두 벡터가 Concat됨, 출력 벡터의 차원이 Projection됨.
F.2.3.1	모델의 FFNN Layer 단계에서 최종 긍정/부정 예측 결과가 출력되는지 확인.	긍정/부정의 의미가 출력됨.
N.1	시연 시 60% 이상의 테스터가 인터페이스의 명료성과 직관성에 대해 만족하는지 리뷰를 통해 확인.	-
N.2	입력 문장 당 2초의 프로세싱 시간을 초과하지 않는지 확인.	Runtime <= 2sec/input
N.3	새로운 표본이 50개 이상일 때 전체 모델 중 최고 정확도가 90% 이상인지 확인.	모델 통계 상 최고 정확도가 90% 이상임.