

2022-S 핀테크 최종 발표

안녕하세요 저는 레이블링팀의 최종 발표를 맡은 금융수학과 20학번 이재호입니다.

(넘김)

오늘 발표는 프로젝트 배경, 내용, 자기 평가, 산출물 페이지에 대해 말씀드리고, 데모를 하는 순서로 진행하도록 하겠습니다.

(넘김)

그러면 가장 먼저 프로젝트 배경입니다.

(넘김)

사전 발표 때 우리는 뉴스의 일부분을 여러분에게 보여주면서, 데이터 라벨링의 정의&데이터 라벨링 알바 등을 말씀드렸습니다. 데이터 라벨링이란

(넘김)

인공지능을 학습시키기 위해 데이터를 분류, 가공하는 작업을 말합니다. 주로 사진이나 동영상 등에 등장하는 사람, 동물, 문자 등에 라벨 즉, 꼬리표를 달거나 텍스트와 음성을 분류합니다.

이런 데이터 라벨링은, 지난 정부의 한국판 뉴딜 사업의 일환으로 데이터 댐이 선정되면서 그 수요와 공급이 모두 증가하였습니다. 또, 코로나 팬데믹을 거치면서 비대면 일자리 선호 현상이 나타나면서, 비대면 작업으로 시공간의 제약을 받지 않고 누구나 할 수 있는 데이터 라벨링은 주목 받고 있습니다.

(넘김)

현재는 클라우드 워크스, 레이블러, aiworks 등 데이터 라벨링을 할 수 있는 플랫폼이 있습니다.

이 플랫폼들에서는 기업, 연구소 등이 AI학습을 위한 data를 요청하면, 데이터 라벨러들에게 사진, 동영상을 찍거나 녹음하는 등의 데이터 수집, 아까 뉴스에서 보신 것처럼 데이터를 분류하는 일, 그리고 데이터 검수 등을 맡깁니다.

그렇게해서 만들어진 data set을 다시 의뢰한 기업들에게 판매하는 방식으로 data set 작업과 판매가 이루어집니다.

즉, 수요에 따라 데이터 라벨링 플랫폼들 기업들에게 받은 판매금액의 일부를 정산하는 방식으로 이루어지는 것입니다.

우리는 data set을 필요로하는 기업, 연구소, 팀 들이 조금 더 저렴한 가격에 데이터 셋을 구매하고, 라벨러들은 조금 더 높은 보상을 받아가는 방법이 있지 않을까 고민하다

(넘김)

클라우드 펀딩 요소를 도입해서

2022-S 핀테크 최종 발표

개발자들에게는 보다 저렴하게 데이터 셋을 구매할수 있게 하고, 라벨러들은 조금 더 높은 보상을 받아 새로운 일자리로 주목받고 있는 데이터 라벨링 작업에 좀 더 쉽게 접근할 수 있도록 하는 플랫폼을 만들고자 하였습니다.

(넘김)

(넘김)

그러면 클라우드 펀딩을 통해서 라벨링된 데이터를 구매할 수 있도록 하고자 하는데,

(넘김)

구매 비용, 정산 가격 등을 어떻게 할 것인지를 정해야했습니다.

(넘김)

기업이 초기에 데이터 작업 의뢰를 할 때, 데이터별 각자 해당하는 단가에 데이터 개수를 곱하고 추가 플랫폼 수익인 20%를 합한 비용을 받습니다. 임의로 텍스트 분류는 50원, 단순 음성 녹음은 70원으로 정하였습니다.

(넘김)

예를 들어 기업이 100만개의 텍스트 분류 라벨링 데이터를 의뢰한다면, 텍스트 분류 단가 50원에 100만개를 곱하고 플랫폼 수익 20%인 1000만원을 더해서 총 6천만원을 초기 작업 비용으로 받습니다.

(넘김)

이후 데이터 라벨링 작업 해서 초기 의뢰 기업에 데이터 셋을 넘긴 후, 데이터를 재판매 할 때는 원래 비용보다 20% 할인하여 저렴하게 판매할 것입니다

재판매된 비용은 초기 의뢰 기업에게 50%, 해당 데이터를 작업한 라벨러들에게 30%, 플랫폼이 20%를로 분배합니다.

(넘김)

보통 기계학습이나 데이터들이 필요할때, 기본 만장씩은 필요하다고 알고 있어서 , 데이터 set을 만개로 나누어 판매한다고 가정하였습니다. 즉, 100만개의 데이터가 있으면 1만개 씩 100개의 set으로 나누는 겁니다.

(넘김)

예를 들어, 초기 작업비용이 50만원이면, 재판매는 20%할인된 40만원으로 하는 것이고 초기 의뢰 기업은 수익 50%인 20만원, 데이터 라벨러는 수익의 30%인 12만원, 마지막으로 플랫폼은 수익의 20%인 8만원을 재판매 될 때마다 갖게됩니다.

2022-S 핀테크 최종 발표

결국 라벨러들은 자신이 라벨링 한 작업물 1건당 12원씩 추가로 받게 되는 겁니다.

(넘김)

결론적으로 개발자는 자신이 필요한 Data Set을 1만개씩 원하는 만큼 구매가 가능하게되고, 데이터 라벨러들은 기본 작업으로 data 개당 비용과 resale 될때마다 해당하는 금액을 추가로 더 받게됩니다. 그러면 이러한 데이터 라벨링 플랫폼을

(넘김)

우리팀은 1차 중간 발표에서 html과 css로 제작한 웹페이지로 구현할 것이라고 말씀드렸습니다.

(넘김)

또, 웹페이지에서 4대 핵심 기능인 로그인, 라벨링, 포인트 적립, 구매창을 구현하겠다고도 말씀드렸습니다. 여기서 라벨링은, 라벨러들이 라벨링을 할 수 있는 tool을 의미하고, 포인트 적립은 라벨러들이 현금화할 수 있는 작업 대가로 받는 포인트를 적립할 수 있는 기능입니다.

또, 구매 창은 펀딩이 성사되어서 이미 라벨링되어서 초기 의뢰 기업들에게 판매된 데이터를 재 판매하는 스토어를 의미합니다.

(넘김)

그 중 라벨링에서는 데이터 수집에서는 녹음을, 분류에는 글의 의도 분류, 판별을 오픈 소스를 통해서 구현하겠다고 말씀드렸습니다.

(넘김)

그러면 최종 우리팀의 결과에 대한 자기 평가를 하도록 하겠습니다.

(넘김)

먼저 4대 핵심기능에서

로그인은 로그인과 회원가입 페이지를 제작하였을 뿐, 계정을 생성해서 로그인하는 기능을 갖추지는 못하였습니다.

(넘김)

그 다음으로 라벨링은, 앞에서 말씀드린 수집에서 녹음, 분류에서 글의 의도 분류 판별을 하는 tool을 오픈 소스를 통해 구현하겠다는 계획을 수립하였습니다. 오픈 소스로 공개되어 있는 구현하고자 하는 기능을 갖춘 tool을 찾는 것에 성공하였지만, 웹페이지에 이식하는 것에 실패하였습니다. 대신 다른 오픈 소스로 웹으로 구현되어 있는 이미지를 분류할 수 있는 툴을 하이퍼링크로 연결하는 것으로 대체하였습니다.

즉 결론적으로는 실패하였다고 평가합니다.

2022-S 핀테크 최종 발표

(넘김)

다음으로 포인트 적립입니다. 다른 기능들을 구현하고 아이디어를 구체화하느라 포인트 적립 기능은 후순위로 밀렸고, 물리적인 시간 부족으로 구현하지 못했습니다.

(넘김)

구매창에서는 실질적으로 결제 등을 할 수 있는 데이터 스토어가 아닌, 데이터 스토어의 모습을 갖춘, 웹페이지를 제작하였다는 점에서 미완이라고 평가합니다.

(넘김)

그러면 페이지를 데모하기 이전에 간략하게 구현 결과를 설명드리겠습니다.

(넘김)

먼저 전체적으로 웹페이지에서

(넘김)

탭 메뉴를 통해 핵심 페이지 및 기능에 접근할 수 있도록 하였고

(넘김)

그 중에서 User guide는 Labeler와 developer를 나누어 페이지를 제작하였습니다.

(넘김)

또, 로그인과 회원가입 페이지

(넘김)

Fact and question 페이지를

(넘김)

그리고 데이터 스토어 페이지를 만들었습니다

—— Demo ——

(넘김)

감사합니다 이상으로 발표를 마치도록 하겠습니다.