

2020_머신러닝자동화 플랫폼 동향과 적용사례_ Q&A 리스트

순번	질문	답변
1	AutoML을 활용한 시를 임베디드로 타시스템과 연계해서 사용할수 있는지요?	예, 가능합니다. 현재 위세아이텍이 주관하는 일부 사업에서도 사업에 특화된 플랫폼에 WiseProphet AutoML 기능 및 저희 측의 기타 제품을 임베드하여 사용하고 있습니다.
2	AutoML이 얼마나 유용한지는 업종, 데이터 유형, 관련된 모델 클래스에 따라 다를거라고 생각되는데요, 이런 관점에서 위세아이텍에서는 AutoML 플랫폼 도입여부를 고민(소요될 재정적, 시간적 비용과 잠재적 이익 등) 하는 기업들에게 어떤 도움을 줄 수 있는지 문의드립니다.	업무에 머신러닝 도입을 고려하는 기업에서 활용 가능하며, 보유하고 있는 데이터에 대한 머신러닝 적용 가능성을 빠르게 판단할 수 있습니다.
3	머신러닝의 자동화라면 인공지능의 개념이 들어가나요. 머신러닝 자체가 인공지능을 사용한 거 아닌지요.	인공지능은 사람의 지능을 모방하여 사람이 하는 것과 같이 복잡한 일을 할 수 있는 기계를 뜻합니다. 이 인공지능을 구현하는 방법 중 중요한 방법이 머신러닝입니다. 따라서 머신러닝 자동화에 인공지능의 개념이 들어간다고 할 수 있습니다. 머신러닝의 자동화란 머신러닝 전문 지식이 부족한 사용자도 비즈니스에 인공지능 기술을 쉽게 적용, 지원하기 위해 최적의 인공지능을 기술을 활용하는 방법입니다.
4	AutoML은 준비과정에서 인간의 개입이 많이 요구되는 복잡한 작업으로 알고있습니다. ML을 최대한 효과적으로 활용하려면 고도로 훈련된 데이터 과학자들로 구성된 팀이 준비되고 동원되어 각 단계마다 개입하면서 모델을 구축, 적용, 최적화를 해야하는데 이런 이슈들이 기업에게는 큰 부담으로 작용합니다. 이런 관점에서 위세아이텍에서는 어떤 도움을 줄 수 있는지 문의드립니다.	최소한의 개입을 줄이기 위해 클릭 기반으로 사용 가능한 AutoML 솔루션을 개발하였으며, 이슈에 대한 해결 및 심층 분석을 위해 컨설팅 및 교육을 제공하고 있습니다.
5	AutoML플랫폼에서 지원하는 알고리즘은 어떤 것들이 있으며 RNN, LSTM, GRU 알고리즘도 적용할 수 있는 것인지 설명 부탁드립니다.	현재 안정적으로 구현 중인 알고리즘은 분류(Decision Tree, Random Forest, Gradient Boosting, XGBoost), 회귀(Decision Tree, Random Forest, Elastic Net, Gradient Boosting, XGBoost), 분류(KNN) 가 있으며 LSTM 알고리즘의 경우 최근에 진행한 사업에서 해당 도메인에 특화된 알고리즘으로 구현하였습니다. 앞으로도 머신러닝과 딥러닝 알고리즘을 보다 다양하게 구현할 예정에 있습니다.
6	반복적 모델 학습과 하이퍼 매개변수 튜닝의 사례, 예시를 간략히 설명해 주실 수 있을까요?	반복적 모델 학습은 일정 주기마다 전체 또는 일부의 학습 데이터를 모델에 넣고, 모델이 산출하는 결과 값에 대해 평가하여, “더 바람직한 방향으로” 모델을 업데이트하는 것입니다. 하이퍼 매개변수 튜닝은 모델을 학습시킬 때 모델의 주요 변수 입력값을 바꾸어가며 모델의 정확도를 높일 때 사용합니다.
7	전처리 준에 프로세스마이닝을 위한 모듈이 포함되어 있는지요? 아니면 기존에 보유하고 있는 프로세스 마이닝과 연계 가능한지요?	프로세스 마이닝 모듈이 어떤것인가요? 조금 더 자세히 문의 주시면 답변드리도록 하겠습니다.

8	<p>모델이 구축, 변경될 때마다 이의 적용까지 전체 프로세스를 자동화하려고 하면 너무 많은 시간이 소요되고 있어 마케팅 팀의 불만이 적지 않습니다. 이를 해결하기 위해서 AutoML 플랫폼을 도입하고자 한다면 이런 선택이 과연 올바른 대안인지 궁금합니다.</p>	<p>모델을 만드는데 있어 걸리는 시간과 모델을 주기적으로 업데이트하는 시간을 단축시킬 수 있다면 올바른 대안이 될 것입니다. 위세아이텍의 WiseProphet을 체험할 수 있는 링크를 보내드리니 사용해보시길 권장드립니다. http://prophet.wise.co.kr/#/intro</p>
9	<p>오픈소스 모델(Python, R등에서 생성한 오픈소스 모델)과의 연계, 비교는 어떻게 되는지? 자동 재학습(어떤 기준으로 하는지) 프로세스도 궁금합니다.</p>	<p>현재 저희는 사업 및 과제에 따라 전담 연구원이 R, Python 등을 사용하여 직접 생성하는 오픈 소스모델과 AutoML 플랫폼을 사용하여 생성하는 모델을 비교하며 인사이트를 뽑아 내기도 하고, 시간 및 자원, 성능을 고루 고려하여 보다 효율적인 모델을 택하기도 하는 유연한 대처를 하고 있습니다. 재학습의 경우 WiseProphet Automl 플랫폼의 모델관리 창에서 기존에 생성한 모델을 사용하여 추가 학습, 또는 새로운 데이터에 대한 예측이 가능합니다.</p>
10	<p>3rd party 알고리즘이나 외부에서 제작된 모델들과의 연동도 가능한가요?</p>	<p>외부에서 제작된 모델들과의 연동이 가능하지만 적용하기 위해서는 별도의 솔루션 커스터마이징이 필요합니다.</p>
11	<p>AI가 최근 대세인데.. ML과 DL 등 업무 적용 기준이 있을까요?</p>	<p>사용하는 데이터에 따라 머신러닝 딥러닝 모두 활용합니다. 예를 들어 이미지, 영상의 경우에는 딥러닝을 많이 사용합니다.</p>
12	<p>이미지 분류 모델의 정확도는 어느정도까지 나오나요?</p>	<p>어떤 이미지에 대한 모델 분류에 대한지에 따라 정확도는 다릅니다. 목표에 따라 정확도에 대한 기준이 달라집니다.</p>
13	<p>비정형 데이터, 즉 사진, 동영상의 ML을 위해서는 어느 정도의 학습기간, 데이터가 필요한지요?</p>	<p>비정형 데이터의 경우 정형 데이터에 비해 데이터의 크기가 크고, 분석에 필요한 데이터 양이 많아 많은 시간이 소요되는 편입니다. 비정형 데이터는 딥러닝 모델을 많이 사용하며 일반 머신러닝에 비해 학습 소요 시간이 긴 편입니다.</p>
14	<p>AutoML은 인력을 줄여서 효율을 높일 수는 있겠지만 자칫 잘못하면 더 많은 오류를 초래할 위험이 있어 만들어진 모델 검증에 많은 시간이 투입하고 있지만 실제 비즈니스에서 어디까지 해야될지 이 균형을 잡기가 쉽지는 않습니다. 이런 관점에서 위세아이텍에서는 어떤 도움을 줄 수 있는지 문의드립니다.</p>	<p>위세아이텍은 자체 분석 인력과 컨설팅 인력을 보유하고 있으며, 교육 또한 제공합니다. 보다 더 상세한 도움이 필요하시면 contact@wise.co.kr로 요청부탁드리겠습니다.</p>
15	<p>부당청구에 대해 강화학습을 적용하는 경우와 하지 않는 경우 정확도 차이가 어느정도가 났나요?</p>	<p>F1-스코어 기준으로 10% 이상 성능향상을 보였으며, 특히 '부당'라벨에 대한 성능은 기존 모델 대비 정밀도 17%, 재현율 29% 향상하였습니다.</p>
16	<p>데이터 결측값(비어있는값) 상태로 모델 생성시 변수간 관계 왜곡으로 모델의 정확성이 떨어지게 되는데, 데이터 결측시 해결 방안이 있는지요? 어떤 방식으로 데이터 결측을 처리 가능한지? 또한 이상 데이터 처리 방법도 궁금합니다.</p>	<p>현재 데이터 결측치 처리도 가능합니다. 각 변수에 결측값 변환 설정을 하지 않을 경우 자동으로 0으로 대체됩니다.</p>
17	<p>AutoML이 알고리즘을 모르는 비 전문가 경우 사용하기 쉽다는 장점도 있으나 '불충분한 데이터 학습 또는 DB 연동 시 과적합 문제가 발생' 할 수 있다는 단점도 있는 것으로 알고 있습니다. 본 AutoML 적용하기 위한 최소 Data 수가 있는지요?</p>	<p>최소라고 말씀하셔서 100개 이상의 데이터가 있으면 적용은 되지만 결과값은 좋지 않습니다.</p>
18	<p>성능향상에 대한 기준은 무엇인가요?</p>	<p>분류모델의 경우, 평가 지표인 정확도, 재현율, 정밀도, f1-score의 값이 클수록 성능이 좋은 것이고, 회귀모델의 경우, 평가 지표인 MAPE, RMSE, MSE는 낮을수록 성능이 좋은 것이며 R2(결정계수)는 높을수록 좋은 것입니다. 부당청구 탐지 모델의 경우, F1-스코어 점수를 기준으로 평가하였습니다. F1 스코어는 데이터 라벨이 불균형 구조일 때, 모델의 성능을 평가할 때 주로 사용하는 평가 지표입니다.</p>

19	강화학습 알고리즘은 어떤 알고리즘이 있나요? 또 강화 학습 학습용 데이터 및 훈련시간은 어느정도 소요되었나요?	강화학습 알고리즘은 분석 데이터에 따라 그리고 강화학습 방향에 따라 다양하게 구현될 수 있습니다. 발표에서는 부당 청구 문제에 DQN과 DDQN 모델을 적용해 보았습니다. 훈련 시간은 데이터와 모델에 따라 달라지는 부분으로 평균 1시간 이상 소요되었습니다.
20	강화학습 알고리즘 적용 전 사전작업할 것이 있을까요? 일반적인 업무에 그대로 적용할게 아니라 선행조건을 맞춰야 할 것들이 있을텐데요...	강화학습 알고리즘을 적용하기 전에, 환경을 정의하기 위해 Feature Selection 과정을 거쳐 강화학습 알고리즘을 적용하였습니다. 실무자에게 확인 후 상세한 답변 드리겠습니다.
21	소개해주신 WiseProphet이 적용된 사례는 어떤 것이 있나요?	함께 보내드리는 소개서 참고부탁드립니다.
22	비정형 데이터는 딥러닝을 통해 진행해도 정확도가 많이 떨어지는 느낌입니다. 정형데이터에 비해 시간은 좀 더 걸려도 어느정도 정확도만 확보되면 좋을것 같은데 관련 데이터가 있을까요	딥러닝으로 비정형 데이터를 분석할 때, 분석 목적과 목표에 따라서 다양항 텍스트나 이미지 데이터로 분석을 할 수 있습니다. AI허브(https://www.aihub.or.kr/)에서 비정형인 자연어, 이미지, 영상, 지식베이스 등의 데이터를 무료로 이용하실 수 있습니다.
23	머신러닝 프로세스 공정에서 몇 퍼센트(%) 자동화가 가능한가요?	어떠한 모델을 만들지에 대한 명확한 목표가 수립되어야 있어야 하며, 이를 적용하기 위한 데이터가 준비되어 있어야 합니다.
24	AutoML을 적용하게 되면 딥러닝을 적용하는 진입장벽이 크게 낮아질 것으로 기대되는데요, PoC(인공지능 전문가의 작업과 비슷하거나 더 우수한 결과를 보이는지 확인) 지원을 요청드릴 경우 기업내에서 어떠한 준비 과정이(데이터 제공 등) 필요한지 궁금합니다.	프로핏 체험가능한 링크 보내드립니다.^^ http://prophet.wise.co.kr/#/intro
25	강화학습을 적용한 사례는 어떠한 것이 있나요?	수치형 변수의 경우는 스케일링 작업을 거쳐서 머신러닝 학습에는 문제가 없지만 원본값은 알아볼 수 없게 조치를 취합니다. 그 외의 문자로 된 기업정보들은 다른 단어로 대체하거나 일부 정보를 가리는 등 원본 값을 추론해낼 수 없도록 변경 작업을 진행합니다.
26	실무에서는 자동모델보단 반수동모델을 활용하는 것 같습니다. 전처리 전 데이터 세트를 필요로 하는데, 이런 부분에서 방금 보여주신 데모 내용은 비전문가분들도 가능할까요.	어떤 부분을 실무에 적용할 수 있을지 빠르게 검증하는 부분이 머신러닝 자동화의 장점이라고 생각합니다. 가지고 있는 데이터가 분류/예측하고자 하는 모델이 적용가능한지 빠르게 탐색이 가능합니다.
27	프로핏에서 사용자가 임의로 PYTHON이나 R 스크립트를 추가 변경해서 결과를 얻을 수 있나요?	학습효과를 높이기 위해서는 많은 데이터가 필요합니다.
28	머신러닝 자동화가 실무에서는 어떻게 적용될 수 있을까요? 앞으로의 실무와 어느정도의 연관성이 있고, 그에 따라 어떤 결과를 얻을 수 있을까요?	데이터의 양도 중요하지만 데이터의 특성과 데이터의 질, 전처리에 따라 예측 결과의 정확도는 달라질 수 있습니다.
29	사진의 경우 학습효과를 높이지 위해서는 case별도 많을수록 높은지요?	학습효과를 높이기 위해서는 많은 데이터가 필요합니다.
30	어느 정도 정확한 예측결과를 얻으려면 얼마만큼의 데이터가 필요한가요?	데이터의 양도 중요하지만 데이터의 특성과 데이터의 질, 전처리에 따라 예측 결과의 정확도는 달라질 수 있습니다.
31	생성된 예측 모델을 Code 형태로 다운로드해서 사용자가 모델을 수정하는 것이 가능한가요?	아니요. 코드형태로 다운로드하는 불가능합니다. 다만 생성하신 모델을 시스템상에 저장하여 재학습, 추가데이터에 대한 검증 등으로 활용하실 수 있습니다. 감사합니다 :)

32	머신러닝을 특화된 전문 업무에 대한 활용과 전문 업무가 아닌 일반적인 다양한 업무에 대한 활용을 하는 경우에 따른 학습 기간과 예측에 대한 오류가 차이가 있는지 궁금합니다. 또 교육을 위한 데이터 양과 질을 어떻게 준비해야 할지도 문의합니다.	일반적으로 AutoML을 통한 적용의 경우 데이터만 준비되어 있으며, 10분내에 모델 생성이 가능합니다. 그러나 이러한 수준의 문제가 아닌 더 고도화된 모델이 필요할 경우에는 데이터 분석가가 필요하며, 일반적인 프로젝트의 경우 6개월 소요 되는 것을 2개월 이내로 줄일 수 있습니다. 단순 교육을 위해서는 캐글 데이터를 사용하시는 것을 추천드립니다.
33	머신러닝자동화 플랫폼에 원본 소스 복제를 위한 CDC와 ETL 툴은 오픈소스를 통해 구성하는지요? 별도의 상용 솔루션을 구입해야 하는지 궁금합니다.	ETL의 경우 오픈소스/사용 솔루션 둘다 적용 가능합니다.
34	김건민 선임 연구원이 고객정보들 비식별한다고 하셨는데 비식별처리는 어떤식으로 처리를 하는지 궁금합니다.	수치형 변수의 경우 스케일링 처리를 하였고 문자로 구성된 정보들은 다른 단어로 치환하거나 일부 정보를 빼는 방식을 사용하였습니다.
35	일반 산업에서의 AI 학습과 데이터 분석 처리하는 종류와 양과 엔지니어링 산업에서의 AI 학습과 데이터 분석 처리하는 종류와 양이 많이 다른지요? 어떤 점을 설계할 때 주의해야 할지도 궁금합니다.	데이터 분석 종류 또는 AI 알고리즘 자체가 특별히 다른 점은 없습니다. 다만 엔지니어링 등 특화된 도메인의 경우 사용자의 도메인 사전지식에 따라 데이터 선정과 동일한 분석 결과에서 뽑아낼 수 있는 인사이트가 달라질 수 있습니다. 데이터 분석 및 AI 학습에 대한 설계 과정에서 중요시하실 것은 귀하가 분석에 사용할 데이터 파악과, 데이터의 양과 질, 그리고 목표하는 바라고 생각합니다. 감사합니다.
36	인공지능AI학습적용또한미래직업선택알선지원가능한지	인공지능 AI 학습 적용은 미래 직업선택의 문제는 아닌것 같습니다. 업무 담당자 분들도 충분히 적용가능한 영역이라 별도의 직업은 아닐 것 같습니다.
37	프로핏 체험을 잠시해봤는데요, 데이터 로딩 시 소요되는 시간에 대한 성능 보정은 솔루션보다 인프라에 크게 좌우될까요?	네 맞습니다. 인프라 성능에 따라 속도가 많이 달라집니다.
38	AI기반의 서비스를 하기 위해 준비할 때 가장 비용 및 시간이 많이 소요되는 과정은 어떤 것이 있을까요?	어떤 AI 서비스를 할 것인지에 대한 목적을 선택하고 이를 분석하기 위한 데이터의 수집 및 전처리에 시간이 많이 걸릴 수 있습니다. 데이터가 없다면 수집 또한 많은 비용 및 시간이 걸릴 수 있습니다.
39	엔지니어링 산업에서 활용했던 빅데이터 분석 시스템 데모를 비전문가가 실행해 볼 수 있는 데모페이지가 있을까요?	데모 페이지는 제공하고 있지 않습니다. 도입을 원하실 경우 contact@sie.co.kr 로 데모 요청을 하시면 개별적인 데모를 진행하도록 하겠습니다. 혹은 프로핏 체험가능한 링크 보내드립니다.^^ http://prophet.wise.co.kr/#/intro
40	데이터 업로드 및 전처리 과정의 오류 발생 가능성 최소화를 위한 방법은 어떻게 있을까요?	학습이 가능한 형태로 데이터를 수집하고 전처리하는 것이 최소화 할 수 있을 것 같습니다.
41	ML모델을 구축할때 분석에 적합한 데이터를 수집하는 것이 쉽지 않은것으로 알고 있는데 데이터 수집은 어떤 과정으로 이루어지나요?	우선적으로 분석 목표를 수립하고, 이를 학습하기 위한 데이터에 대해 선행 조사를 진행해야 합니다. 내부 데이터 외에 외부데이터를 통한 수집 방법도 확인을 해야 합니다.
42	드래그 앤 드롭을 통한 손쉬운 인터페이스를 통해 데이터 품질 및 준비 과정을 자동화 가능한지요? 또한 변수 분포도 및 기타 척도를 토대로 시각화한 결과값도 볼 수 있는지? 사용자 중심의 편의성에 어떤 강점이 있는지 궁금합니다.	현재 WiseProphet은 인공지능 자동 학습에 집중하는 제품입니다. 추가로 데이터 품질 관리의 경우 Wise DQ라는 자사 제품에서 전담하고 있으며, 시각화는 Wise Intelligence(WI) 제품에서 전담하고 있습니다. 위세아이텍은 과제 또는 사업에서 의뢰 측의 니즈에 따라 단일 또는 다수의 솔루션을 적용하여 목표하는 결과를 도출하고자 노력하고 있습니다. 추가적인 문의는 contact@wise.co.kr로 연락 주시면 자세하고 성실한 답변 드리겠습니다.

43	산업군마다 정형된 데이터가 상이 할 수 있는데 사용하셨던 알고리즘 말고 다른 알고리즘을 이용해 정확도를 높일 수 있나요?	데이터 또는 도메인에 따라 현재 WiseProphet에서 구현되어 있는 알고리즘 외 R, Python 등을 사용하여 직접 구현하시는 알고리즘이 보다 정확할 수도 있습니다. 그러나 AutoML 플랫폼의 강점은 시간 및 인력 자원 등을 고려할 때 IT 비전문가도 충분히 훌륭한 결과를 도출할 수 있다는 점입니다. 또한 사업/과제의 경우 연구원이 직접 생성하는 다양한 모델과 WiseProphet을 통해 생성하는 모델을 비교하며 최적의 결과를 도출하는데 최선을 다하고 있습니다. 감사합니다.
44	머신러닝을 위해서는 데이터가 많이 축적되어야 하지 않는지요? 유의미한 데이터의 양에 따라 결과가 달라 나올 수 있지 않지요?	네 그렇습니다. 데이터 양이 풍부할수록, 데이터의 질이 좋을수록 머신러닝의 효과가 더 좋습니다. 유의미한 데이터의 양이 많을수록 예측 결과의 신뢰성도 높고 결과도 더 좋은 경향이 있습니다. 특히 학습용 데이터의 질이 가장 중요하며, 주요한 피처를 얼마나 잘 추출하는지가 중요합니다.
45	인스타그램의 사진과 같은 비정형데이터의 경우 별도의 피처를 사용자가 직접 입력해야하나요? 인플루언서의 인스타그램 이미지를 분석하여 최신 트렌드 등 예측가능한 분류값들이 무엇이 있을지 궁금합니다.	비정형 데이터라고 하여도 라벨링이 선행되어야 합니다. 인플루언서의 인스타그램 이미지에서 어떠한 최신 트렌드가 나오는지에 대한 라벨링 작업이 선행되어야 합니다. 그게 아니면 이미지와 해시태그를 결합하여 분석하는 방법도 있을 것 같습니다.
46	[질문] 금형에서 Si를 통한 비정형 데이터 분석에서는 재질과 형상, 구조 해석의 자료를 통한 학습을 어떻게 준비하고 진행이 되는지요? 비정형 데이터를 수집하는 빅데이터 비정형 데이터의 질과 양은 사전 준비 작업도 문의합니다.	우선, 발표 내용 대로 3D도면을 이미지화해 학습하고 있습니다. 즉, 학습에는 이미지의 형상만을 학습하고 형상만이 유사한지 판별합니다. 그러나, 실제 업무에서는 형상 뿐만 아니라, 재질, 크기도 중요합니다. 따라서 이러한 메타정보는 유사도면 검색 후 필터링 조건으로 사용합니다. 추가적으로 저희가 수집한 3D도면은 대략 2500건이고, stp, igs, catia 파일만을 대상으로 했습니다.
47	발표 내용 중 예측과 판단에 있어서 정확도가 100%가 되지 않는 상황에서 사람의 예측 오류율과 비교해 현재 인공지능의 예측 정확도가 어느 정도 선인가요?	산업 도메인별로 다를 것 같습니다. 단순 반복적인 작업의 경우 인공지능을 통한 예측 정확도가 더 높습니다. 사람과 인공지능 적용에 대해서는 지속적으로 보완을 해줘야 하는 부분입니다. 인공지능이 예측한 결과를 사람이 검증을 통해 피드백을 주면서 지속적으로 모델을 진화시켜야 합니다.
48	이미지분석을 머신비전이 아닌 딥러닝 기술을 사용한 특별한 이유가 있나요?	일반적으로 이미지 분석은 머신비전이 아닌 딥러닝 기술을 사용합니다. 특히 기존 이미지에 대한 딥러닝 기반의 학습 모델들이 많아 이러한 모델들을 활용하기 위해 딥러닝을 적용합니다.
49	위세 아이텍 머신러닝 솔루션이 타 경쟁사 대비 어떤 장 단점이 있나요?	클릭 기반으로 솔루션을 쉽게 사용할 수 있다는 것이 가장 큰 장점이며, 예측 결과에 대한 시각화 분석 제공 및 모델 운영 환경을 제공한다는 점입니다.
50	딥러닝 모델 구축시 훈련 데이터와 테스트 데이터만으로도 훈련 척도 판단이 가능하지만, 훈련 데이터에 대한 학습만을 바탕으로 모델의 설정을 튜닝하게 되면 과대적합(Overfitting)이 일어날 가능성이 크고, 테스트 데이터는 학습에서 모델에 간접적으로라도 영향을 미치면 안되므로, 테스트 데이터로 검증을 해서는 안되는데, 검증 데이터 셋 평가 및 튜닝 방안이 있는지요?	모델 생성 과정에서 입력한 데이터셋 중 일부 비율을 정하여 (ex. 80% : 20%) 검증 데이터로 사용할 수 있습니다. 해당 데이터는 학습 과정이 아닌 모델 성능을 검증하는 부분에서만 사용되고 있습니다.
51	유사 도면 검색에 대해서 유사도가 9 이상만 비슷한 것으로 인식되어진다고 하셨는데, 결과적으로 9 이상으로 유사도를 높이신건가요?	유사도가 9미만인 데이터를 검증했을 때 유사해 보이지 않고 판단되었기 때문에 유사도가 9이상인 이미지만 유사한 이미지라고 판단하여 사용자에게 표출될 수 있도록 하였습니다. 감사합니다.

52	그렇다면 노출되지 않는 데이터에 대한 처리는 어떻게 되나요?	다른 제품을 검색할 때 유사도가 9 이상이라면 노출될 수 있도록 보관,관리됩니다.
53	소개해주신 레퍼런스에서 유사 도면 비교는 설계 변수 (형상 치수 등)들은 전혀 활용하지 않고, 캡처한 이미지에서 추출한 특성만을 활용하신건가요?	시스템의 유사 도면을 검색할 때 설계 변수를 입력하여 필터링 하고 유사 도면을 검색하는 기능이 있습니다. 감사합니다.
54	자연어 처리에 적용한 SI 모델은 무엇인지요?	lstm 알고리즘을 사용했습니다
55	한글, 중국어 등도 분석이 가능한가요?	현재는 영어로 된 텍스트 데이터만 분석 가능합니다.
56	설문조사 플랫폼, 소셜미디어 채널, 콜센터 기록 등과 같이 고객이나 직원이 생성한 상대적으로 짧은 텍스트를 분석 성능도 타사 대비 뛰어난지 궁금합니다.	현재 ITB는 영문으로 되어 있는 데이터입니다. 따라서 설문조사 플랫폼, 소셜미디어 채널, 콜센터 기록 등의 데이터가 한글 데이터라면 현재 기술문서분석과의 성능 비교가 어렵습니다. 다만 기술문서 전문가의 지식을 내재화하였기 때문에 타 기술문서 분석보다 성능이 뛰어나다고 할 수 있습니다.
57	NLP는 BERT, XLNet, OpenAI GPT3가 최신 모델로 알고 있는데 이것들을 적용할 수 있는지요?	BERT로 적용이 가능하고, 이미 BERT로 분석 모델을 만들어 보았으나, 모델 성능과 시스템화 되었을 때의 분석시간 등을 고려하여 현재는 LSTM으로 진행을 하고 있습니다.
58	금형 빅데이터 유사 도면 검색 시에, 비교 대상 이미지와 기존 DB의 많은 이미지와 유사도를 비교하려면 시간이 오래 걸릴 것 같은데 사례 동영상은 검색 결과가 빠르게 나타나는 것으로 보입니다. 이 과정은 동영상 편집인가요 아니면 빠른 결과를 출력할 수 있도록 한 방법이 별도로 있으신건가요?	네 시연영상은 시간관계상 편집이 되었으며, 실제 시스템상에서는 검색하는 도면의 개수에 따라 차이가 있지만 약 30초에서 1분정도 소요됩니다.
59	AutoML을 사용할 때에는 정확히 ML 알고리즘을 정할 필요가 없는 것으로 알고 있었기 때문에 세션3에서 다루신 내용도 AutoML을 사용한 경우인지 궁금합니다. 만약 AutoML을 사용한 경우가 아니라면 AutoML을 사용하지 않은 특별한 이유가 있는지도 설명해주시면 고맙겠습니다.	기술문서 분석은 AutoML을 사용하지 않았습니다. 기술문서 분석은 ITB를 검토하는 전문가집단의 지식을 기반으로 하고 있습니다. 그래서 개인들이 AutoML로 분석을 하는 것이 아니라 현재 기본적으로 전문가집단의 지식을 솔루션에 내재화하여 분석을 진행하였습니다.
60	위세아이텍 AutoML 도입 효과에 대한 실제 사용자의 측정지표 (AutoML을 통한 모델과 기존 모델의 성능 및 모델링 시간, 재학습 및 운영 적용에 걸리는 시간 및 비용)가 궁금합니다.	일반적으로 모델 개발 프로젝트의 경우 5~6개월이 소요됩니다. 또한 모델 운영시 별도의 운영자가 필요합니다. 그러나 위세아이텍의 WiseProphet을 활용하면 기본적인 모델의 개발 및 배포에 10분정도가 소요됩니다. 물론 기본적인 모델이 아닌 복합 모델과 기타 모델의 정확도를 더 높이기 위한 데이터 분석가의 역할을 고려 한다면 2개월 내에 모델에 대한 최적화가 가능합니다.
61	오늘 웨비나 내용을 다시볼 수 있는 방법이 있을까요?	저희 유튜브 채널에 업로드해 두었습니다. https://www.youtube.com/c/위세아이텍채널