

신입 분석가가 스타트업에서 살아남기 위한 노오오력

배병선

애자일소다

신입 분석가가
스타트업에서
살아남기 위한
노오오력



목 차

- 1) 대놓고 홍보하기 : 애자일소다
- 2) 발표자
- 3) 경험담.. 들어는 드릴께
- 4) 결론



목 차

1) 대놓고 홍보하기 : 애자일소다

- 무엇을 하는 회사인가요?
- 철학이 뭔가요?

2) 발표자

3) 경험담.. 들어는 드릴께

4) 결론



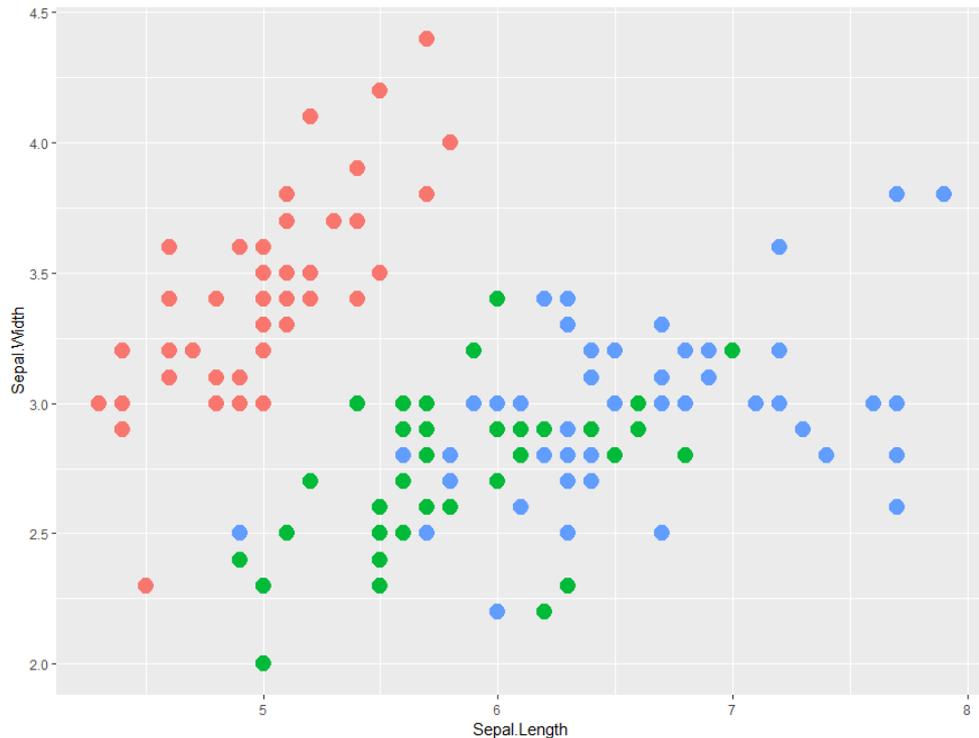
Q : 애자일소다?
뭐하는 회사인가요??



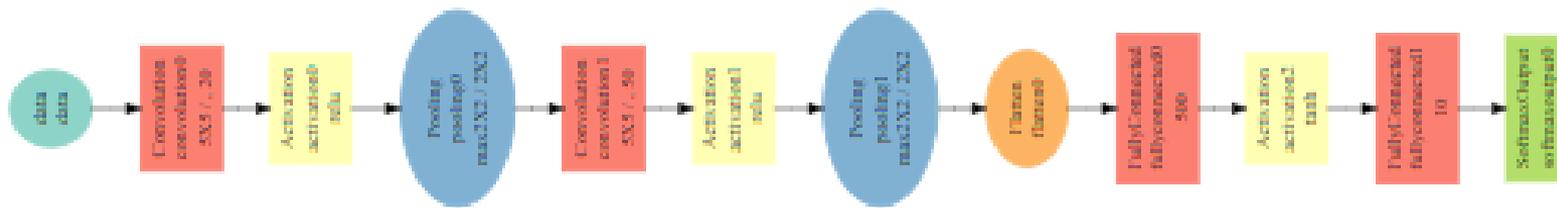
A : AI 회사입니다.



하지만 단순히 분석해주고



모델링해주는 컨설팅 회사가 아닙니다.

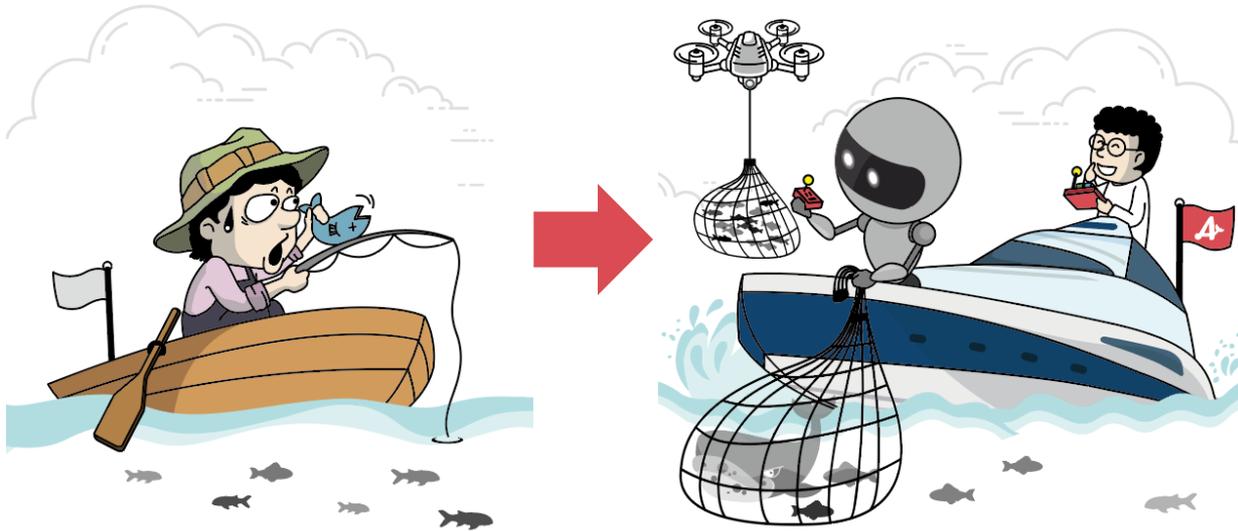


Agile Software Defined AI

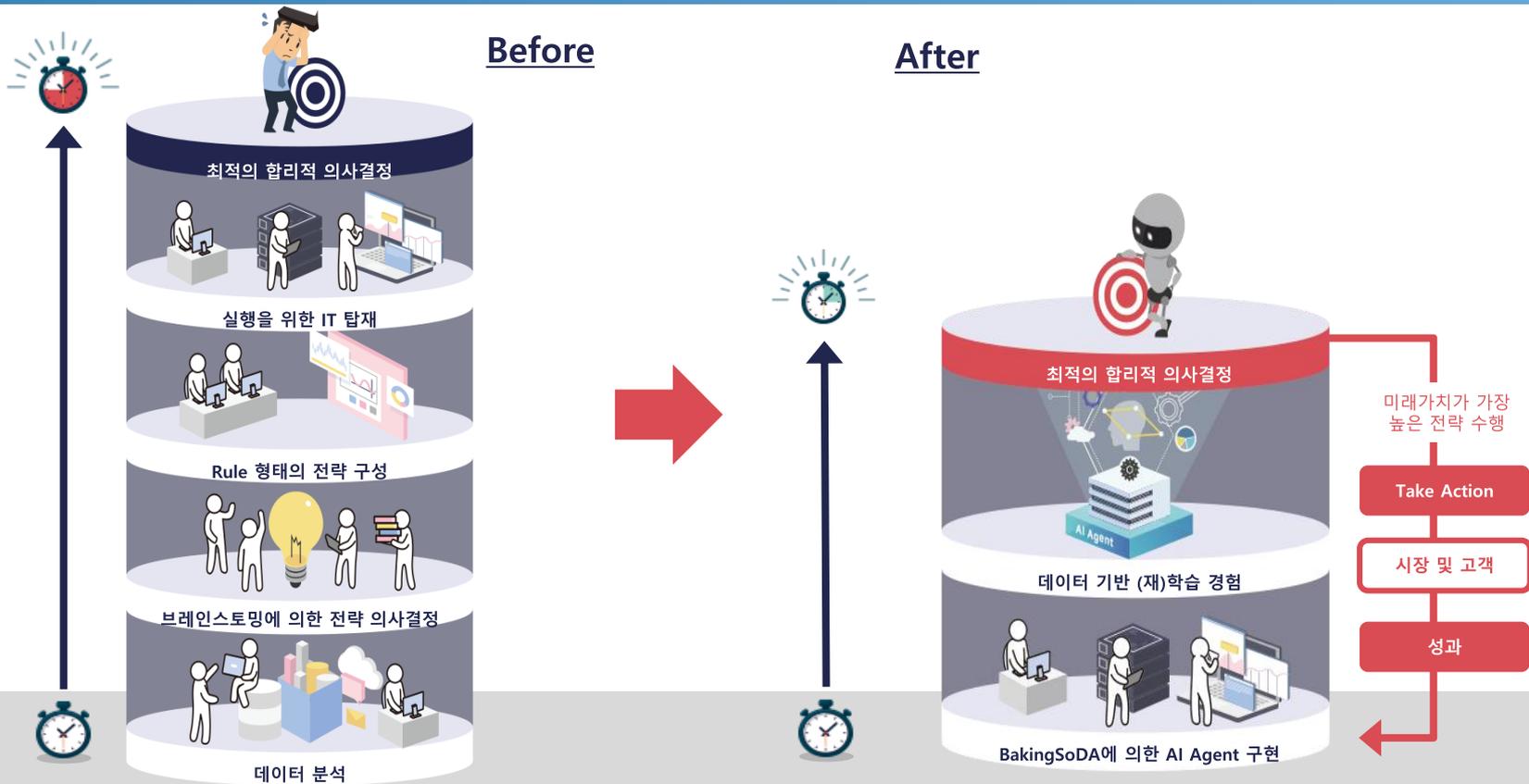
Agile Software Defined AI



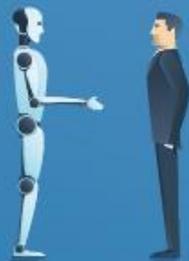
Fast Digital Transformation



Able to Optimal Making Decision



**Q : 애자일소다?
철학이 무엇인가요?**



A : Everyone's AI



모두를 위한 AI 입니다.



기업 내부에 AI 전문 인력은
턱없이 부족하고..

분석가들 간 산출물, 좋은 사례와 최신기술에 대한
공유가 어려워 기업의 자산으로 연결시키기 어렵고...

데이터도 없고 있어도 품질 등의 문제로
실제 학습에 활용할 수 없네..

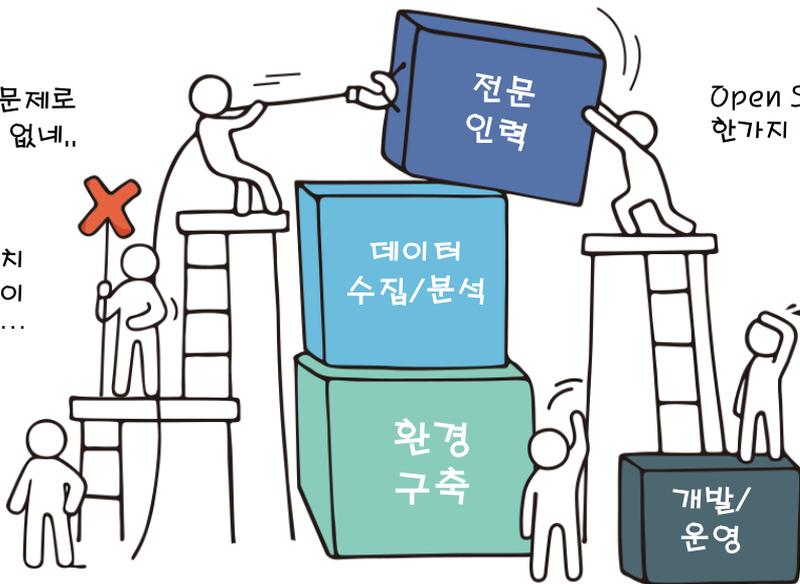
Open SW 버전관리가 어려워 데이터 분석팀은
한가지 버전만 사용할 수 있길 원합니다.

분석은 어렵고 학습 환경도 만만치
않을 뿐더러 분석과 시스템 운영이
분리되어 있고...

모델을 많이 만들면서 버전관리도 어렵고
시스템 적용까지 오래 걸리고 모델 추적도
불가능해집니다. 뭔가 솔루션이 필요..

GPU, CPU 선택적으로 사용 가능한
AI 분석환경은 없나요?

시스템에 적용한 모델이
어떤 것인지? 결과는 좋은 거야?





이러한 수요를 만족시키기 위해
애자일소다는
항상 노력하는 중입니다.



이 회사가 궁금하다면?

검색창에 **agilesoda.ai** 를 쳐보세요.





후 보 끝!

이제 저의 경험담을
이야기해볼까 합니다.



목 차

1) 대놓고 홍보하기 : 애자일소다

2) 발표자

3) 경험담.. 들어는 드릴께

4) 결론



Q : 그런데 누구시죠?



A : 저는 이런 사람입니다.





2017.01 ~ 현재	AI알고리즘연구팀	알고리즘 개발 + 데이터 분석
프로젝트 명		
대용량 데이터분석환경 개발 (스파크 엔진 기반)	모 카드사 EDC중복매출 선별 강화학습 모형개발 컨설팅 프로젝트	
머신러닝기법을 활용한 신용평가 Pilot모형 개발	бат데리 공정 이미지 품질 판정 모형 개발	
강화학습 및 딥러닝 기반 텐서플로우 엔진연구	수율 예측 알고리즘 튜닝	
증권 시세 예측 알고리즘 개발 프로젝트	애자일소다 BakingSoDA-강화학습 알고리즘 패키지 개발 프로젝트	



지금부터 들으실 이야기는
100%
(약간의 뺨을 보탠)
실화입니다.



경험담.. 들어는 드릴게



목 차

1) 대놓고 홍보하기 : 애자일소다

2) 발표자

3) 경험담.. 들어는 드릴께

- 알고리즘 개발하면서
- 데이터 분석 프로젝트를 하면서
- 기술 영업을 하면서
- 공부해서 남들에게 퍼주기

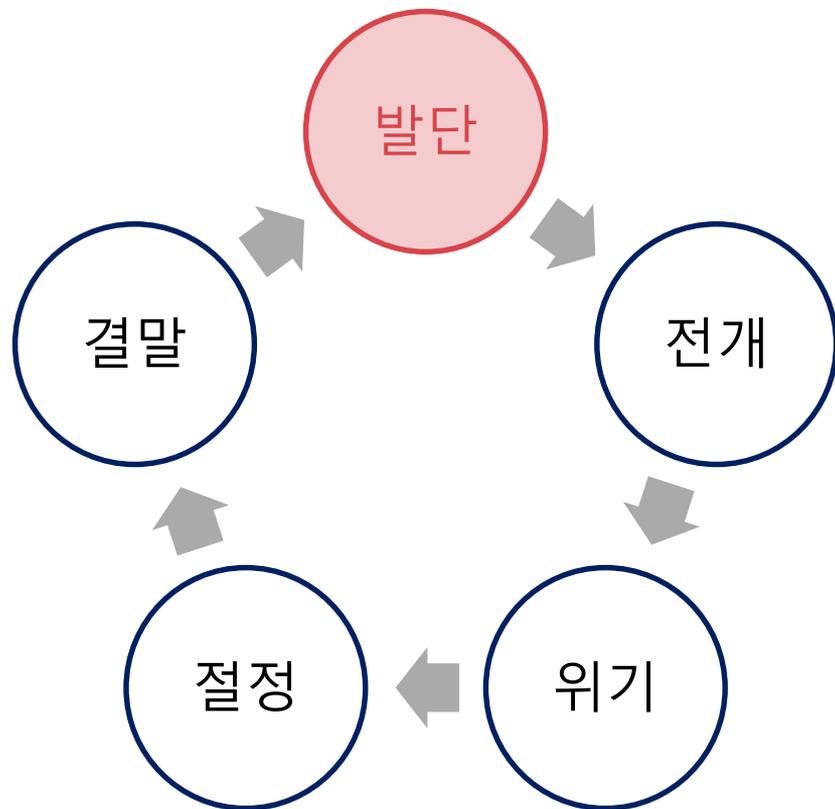
4) 결론



알고리즘 개발을 하면서



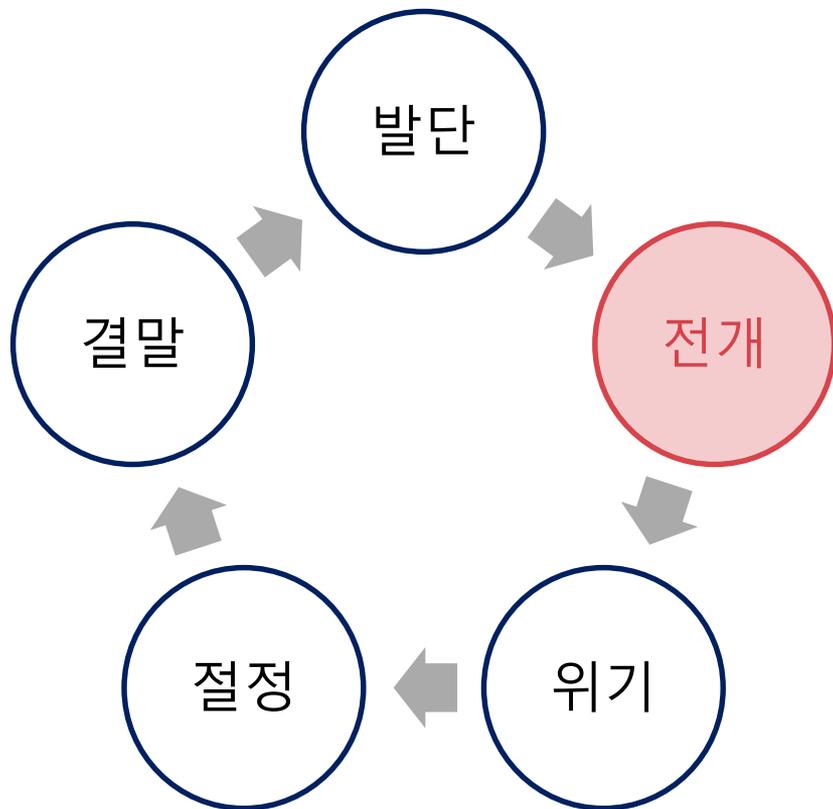
알고리즘 개발을 하면서



알고리즘 개발을 하면서

나란사람.연구팀.제안작업

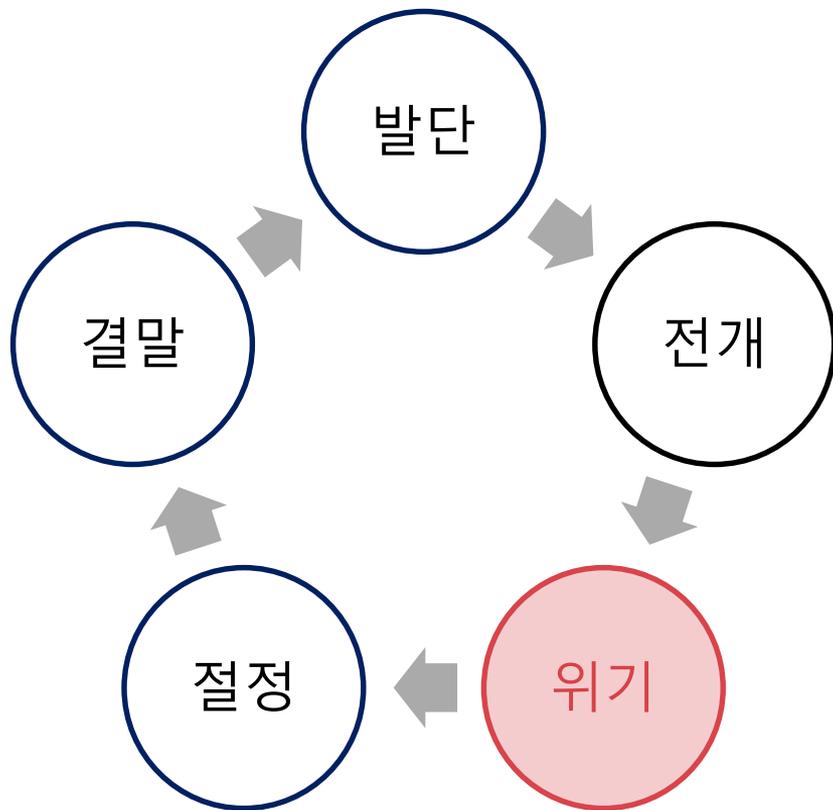
알고리즘 개발을 하면서



알고리즘 개발을 하면서

제한사항.팀원.질투폭발

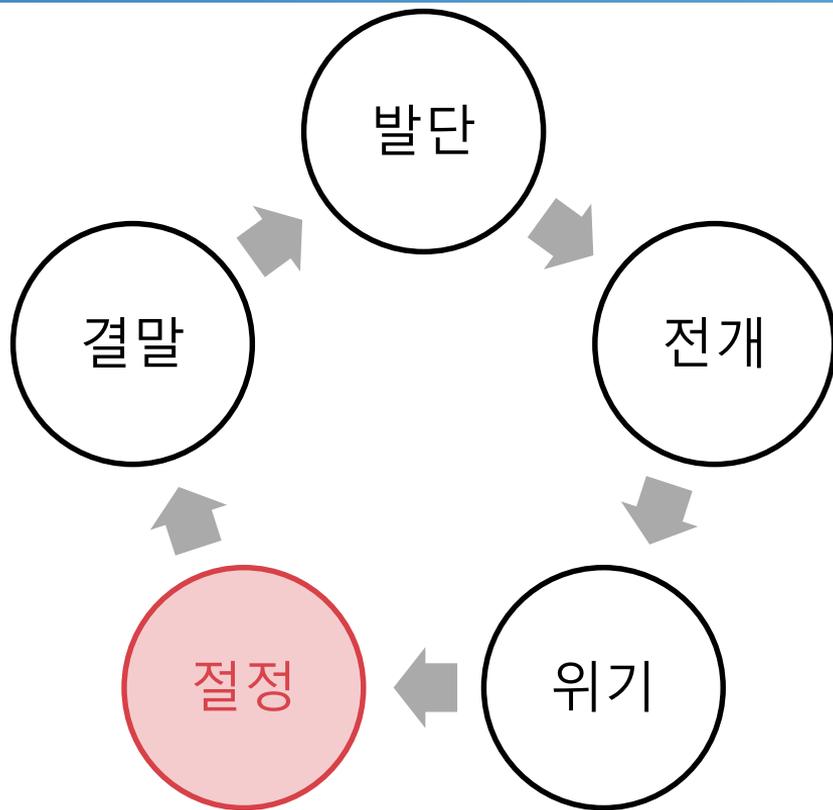
알고리즘 개발을 하면서



알고리즘 개발을 하면서

자존심.의견충돌.다툼

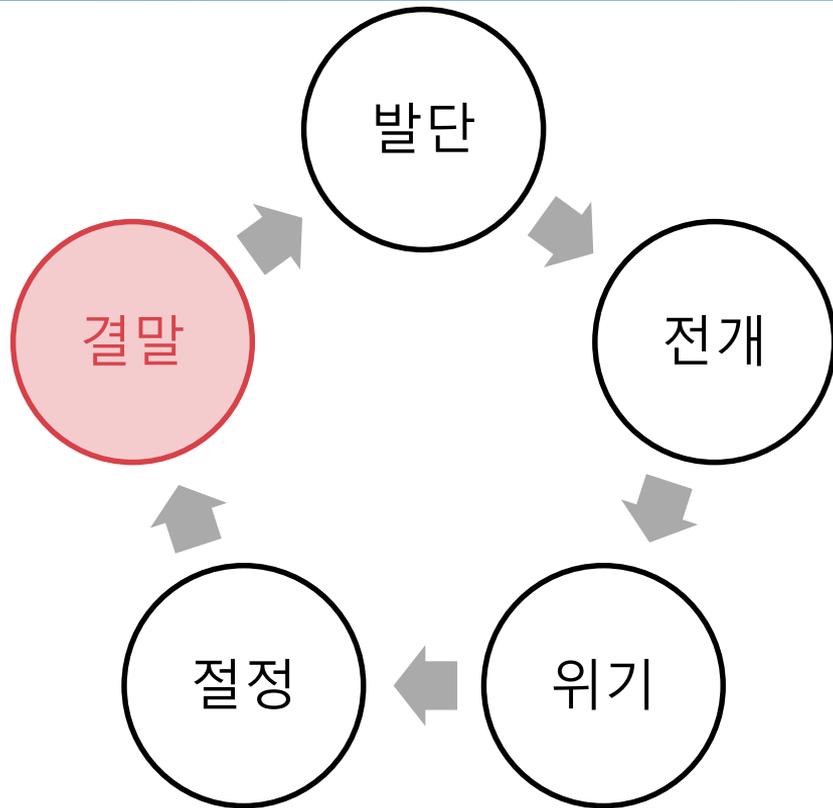
알고리즘 개발을 하면서



알고리즘 개발을 하면서

관찰.반성.인정

알고리즘 개발을 하면서



알고리즘 개발을 하면서

현재상태
집중
새로운걸 찾아서

Q : 화자의 이야기를 통해
얻을 수 있는 교훈은?



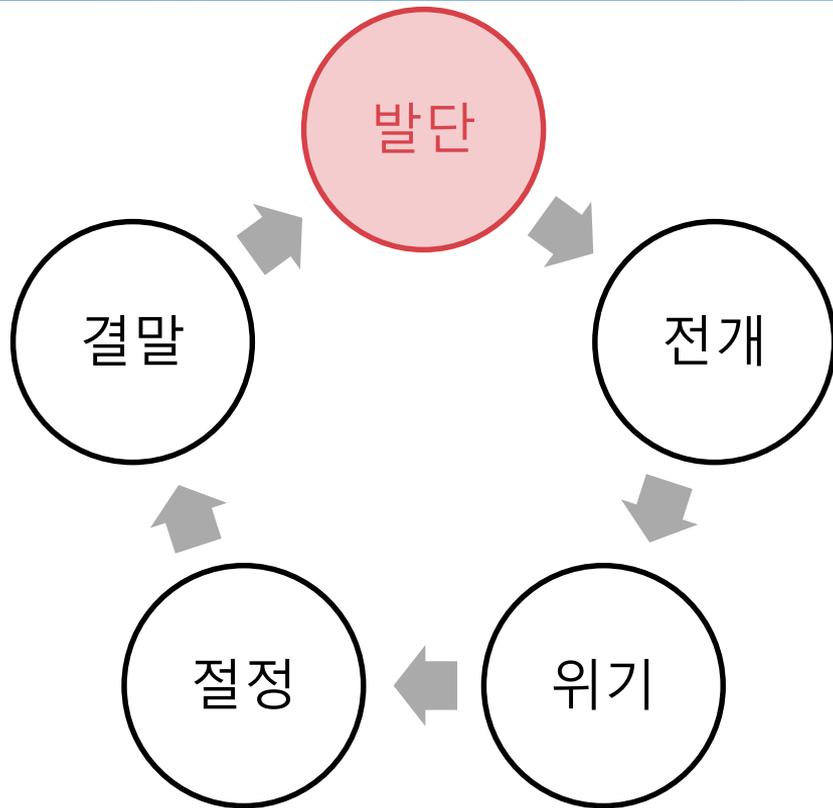
알고리즘 개발을 하면서

1. 타인을 인정하는 자세
2. 내가 잘할 수 있는 일 찾아보기
3. 감정 컨트롤도 개인의 능력

분석프로젝트를 하면서



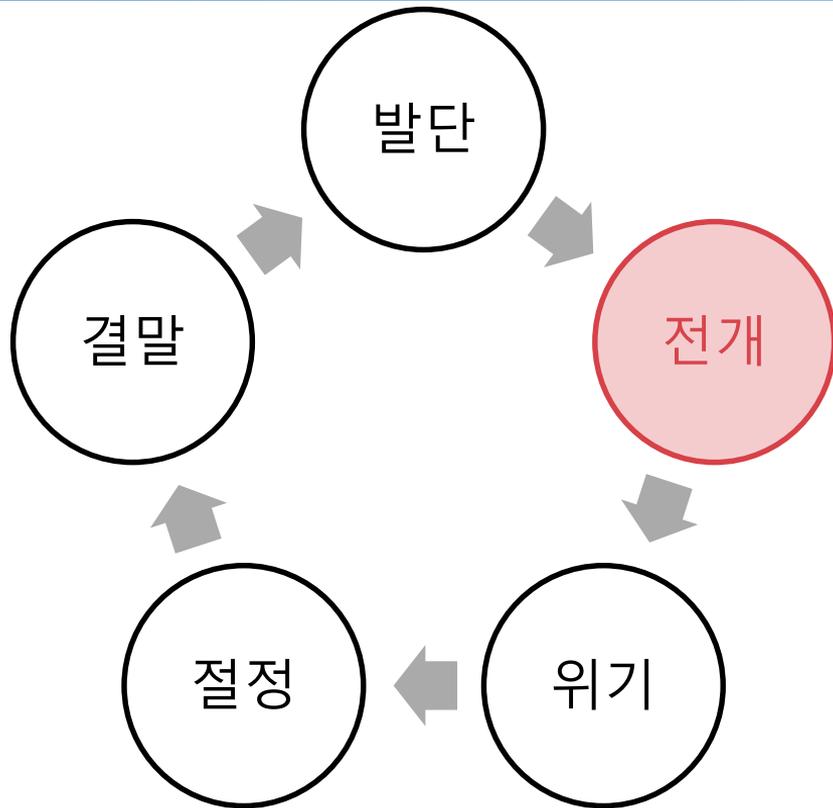
분석프로젝트를 하면서



분석프로젝트를 하면서

나혼자.PoC.분석및모델링

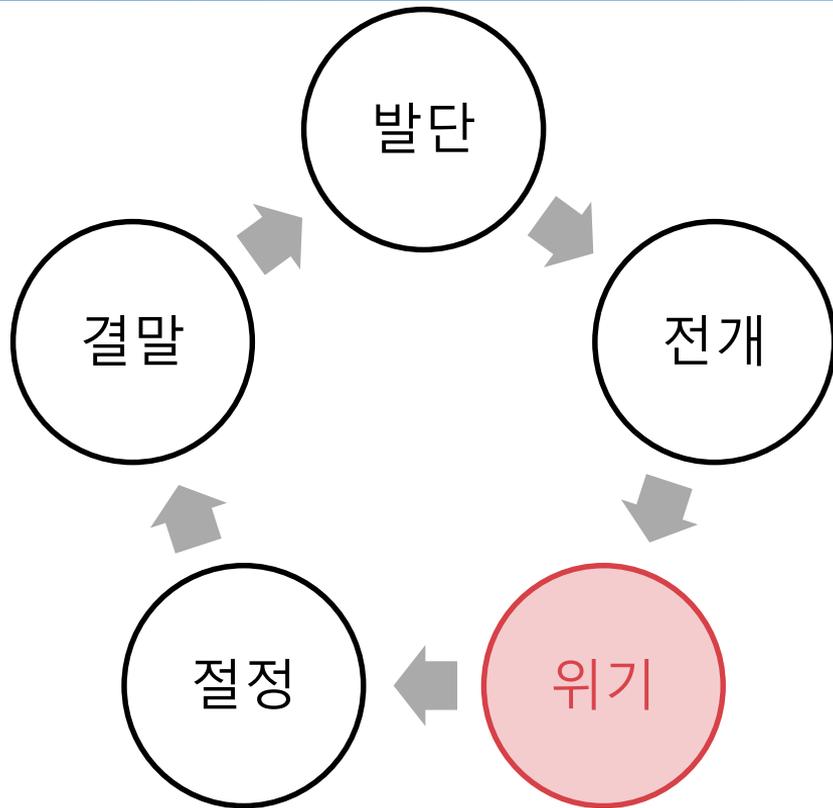
분석프로젝트를 하면서



분석프로젝트를 하면서

순조로움.결과.최선의노력

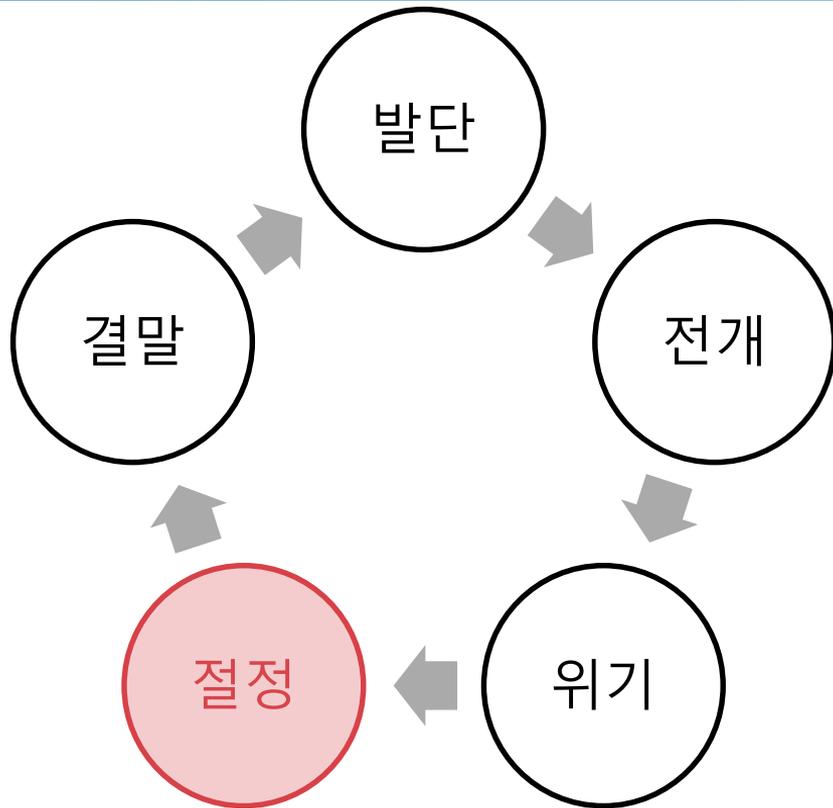
분석프로젝트를 하면서



분석프로젝트를 하면서

실망.망각.지나침

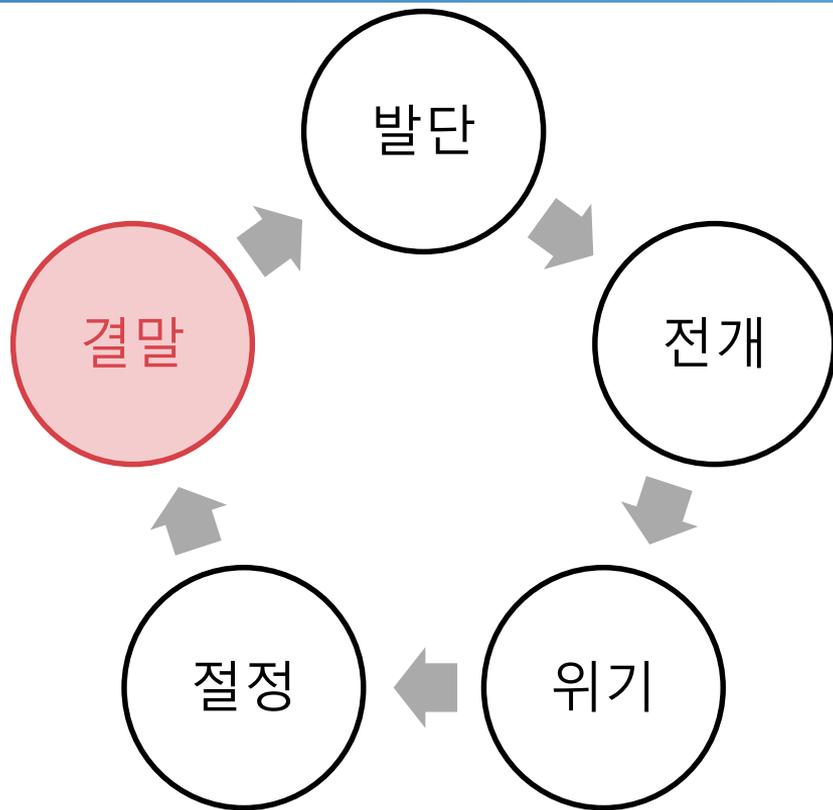
분석프로젝트를 하면서



분석프로젝트를 하면서

질책.반성.깨달음

분석프로젝트를 하면서



분석프로젝트를 하면서

돌파구.노력.공유생활화

Q : 화자의 이야기를 통해
얻을 수 있는 교훈은?



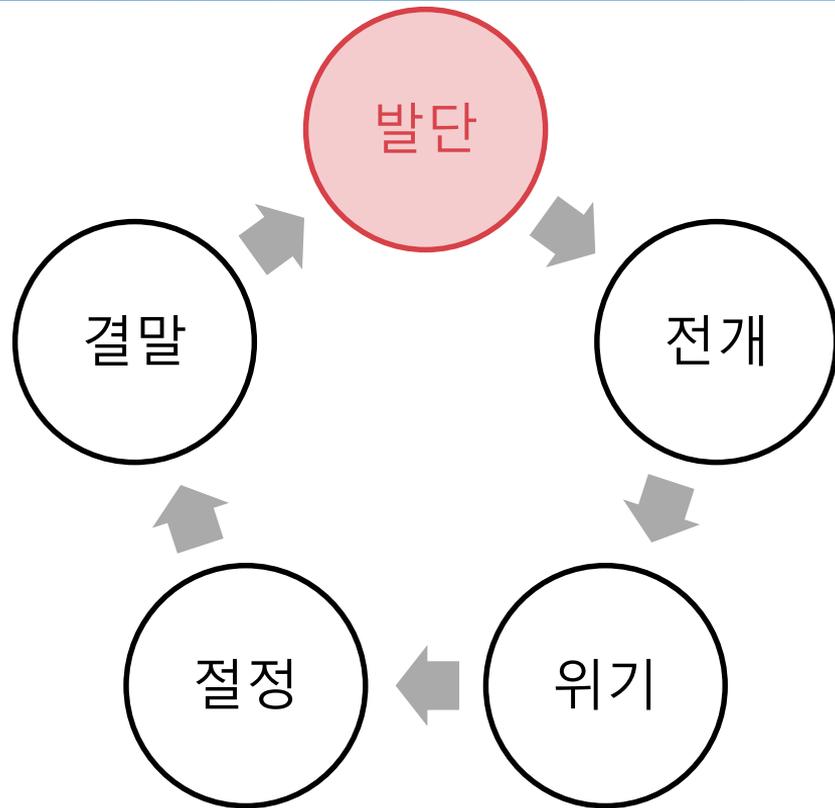
분석프로젝트를 하면서

1. 요구사항은 정확하게 파악할 것
2. 확인 그리고 또 확인!
3. 솔직한 태도

기술영업을 하면서



기술영업을 하면서



기술영업을 하면서

대표님 : 요새 뭐하니?

나 : 저 이것저것 하고 있습니다.

대표님 : 그래? 그럼 나 하나만 부탁하자.

기술영업을 하면서

대표님 : H2O사에서 만든

Driverless AI의 내용을

설명해야하는데 너가 적격인거 같아.

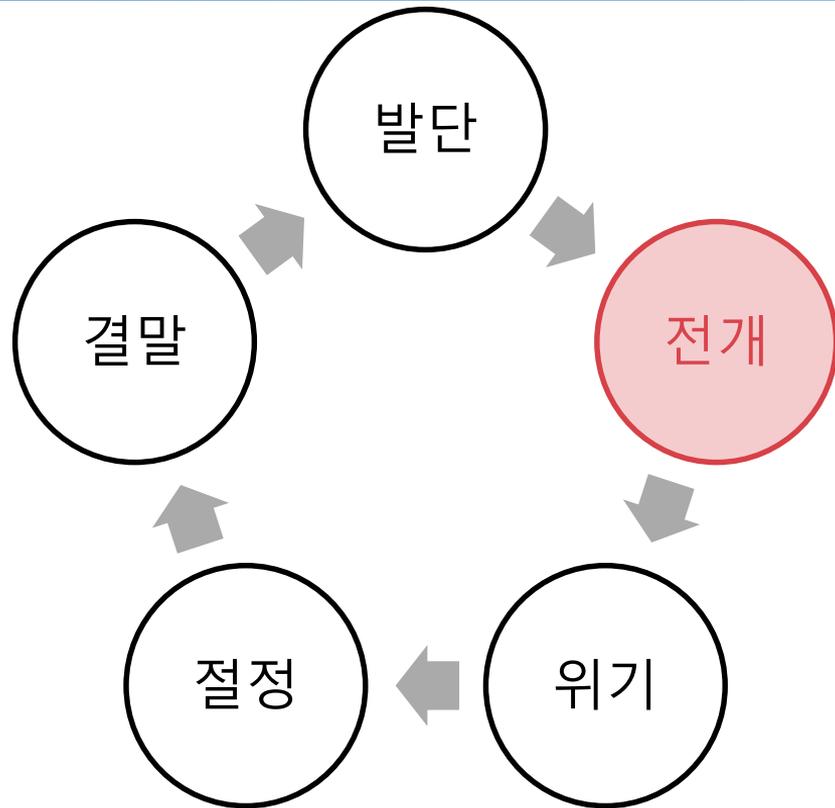
기술영업을 하면서

나 : 한번 해보겠습니다!

기술영업을 하면서

그때부터였어요...
영업에도 발을 들여놓게 된게...

기술영업을 하면서



기술영업을 하면서

WHY.문제점.자료구성

기술영업을 하면서

What are Auto ML & XAI?

The Current State of Automated Machine Learning - KDnuggets

<https://www.kdnuggets.com/2017/01/current-state-automated-machine-learning.html>
Automated Machine Learning (AutoML) has become a topic of considerable interest over the past year. A recent KDnuggets blog competition focused on this topic, resulting in a handful of interesting ideas and projects. Several AutoML tools have been generating notable interest and gaining respect and notoriety in this...

AutoML - Machine Learning for Automated Algorithm Design

www.mila.ubc.ca/ml/
News: We will hold an AutoML workshop at the Federated AI Meeting (FCM, UCAI, AASIS and ICCBR) in Stockholm. Check the workshop page for further information. What is AutoML? Automated Machine Learning provides methods and processes to make Machine Learning available for non-Machine Learning experts.
Auto-stream - RoBO - Robot Bayesian - AutoWeka

Automated machine learning - Wikipedia

https://en.wikipedia.org/wiki/Automated_machine_learning
Automated machine learning (AutoML) is the process of automating the end-to-end process of applying machine learning to real-world problems. In a typical machine learning application, practitioners must apply the appropriate data pre-processing, feature engineering, feature extraction, and feature selection methods that...
Targets of automation - Examples

Google Reveals Automatic Machine Learning: A.I. Can Create Itself ...

<https://www.inverse.com/article/ai-can-create-itself>
May 23, 2017 - Called "AutoML," the "auto-machine learning," it allows one A.I. to become the architect of another, and direct its development without the need for input from a human engineer. On the surface, that sounds like the sort of thing that could lead to the runaway evolution of the singularity, but it's actually Google's...

Cloud AutoML - Custom Machine Learning Models | Google Cloud

<https://cloud.google.com/automl/>
Cloud AutoML helps you easily train high-quality custom machine learning models with limited...

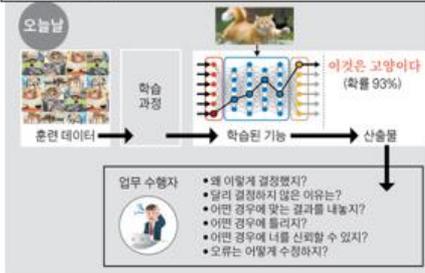
Explainable Artificial Intelligence

From Wikipedia, the free encyclopedia

An **Explainable AI (XAI)** or **Transparent AI** is an artificial intelligence (AI) whose actions can be easily understood by humans. It contrasts with "black box" AIs that employ complex opaque algorithms, where even their designers cannot explain why the AI arrived at a specific decision.^[1] XAI can be used to implement a social right to explanation.^[2] Transparency rarely comes for free; there are often

못 믿겠다 시, 설계자도 심층신경망 작동 방식 몰라

홍정민 기자 | 2018.03.25 09:58



The New York Times

Can A.I. Be Taught to Explain Itself?

As machine learning becomes more powerful, the field's researchers increasingly find themselves unable to account for what their algorithms know — or how they know it.

By CLIFF KUANG NOV. 21, 2017

자료만들기.과정.의외로즐거움

기술영업을 하면서

Why AutoML?

The C
https://n
Autome
yves A.
0965 ar
and 909

Autob
www.ri
News, 1
ICCBRI
Machine
Machine
Auto-94

Auton
https://n
Autome
app/100
must ap
vector
Targets

Googl
https://n
May 23,
of auto
that you
actually

Cloud
https://n
Cloud AI

Explair

From Wikipedia

An Explainabl
understood by
where even th
to implement a



Why AutoML?

- Feature Engineering을 이용해서 어떤 변수를 어떻게 조합을 해서 만들어야 성능이 좋은 모델이 나오는지 여러 실험을 거쳐야함

Appendix A: Dri

Transformations in Driverless /
engineered features. Driverless

- Filter Transformer

The Filter Transformer c

- Frequent Transformer

The Frequent Transform
either the raw count or

- Bulk Interactions Transform

The Bulk Interactions Tr
data.

- Truncated SVD Numeric Tr

Truncated SVD trains o
truncated SVD will be n

- Cross Validation Target Enc

Cross validation target i

- Cross Validation Categoric

This transformer conve
encoding is done on the

- Dates Transformer

The Dates Transformer retrieves any da

- Year
- Quarter
- Month
- Day
- Day of year
- Week
- Week day
- Hour
- Minute
- Second

- Text Transformer

The Text Transform transforms a tex
document frequency) or count (coun
reduction using truncated SVD.

- Numeric to Categorical Target Encoding

This transformer converts a numeric
target encoding is done on the binn

- Cluster Target Encoding Transformer

Selected columns in the data are clu

- Cluster Distance Transformer

Selected columns in the data are clu
calculated.

- Weight of Evidence

Creates likelihood type of features using the Weights Of Evidence (WOE) transformation method. The weight of evidence tells the predictive power of an independent variable in relation to the dependent variable, for example, the measurement of good customers in relations to bad customers.

$$WOE = \ln \left(\frac{\text{Distribution of Goods}}{\text{Distribution of Bads}} \right)$$

This only works with a binary target variable. The likelihood needs to be created within a stratified kfold if a fit_transform method is used. More information can be found here: <http://ucanalytics.com/blogs/information-value-and-weight-of-evidencebanking-case/>.

- Numeric To Categorical Weight of Evidence Transformer

This transformer converts a numeric column to categorical by binning and then creates the likelihood type of features using the WOE transformation method.

- DateTime Label Encoder

Time series frequency preserving label encoding of date times or relative times in [ns].

- DateTime Label Normalizer

Normalization of label encoded time-axis between train/valid or train/test.

- Lags Transformer

Creation of target or feature lags.

- Lags Interaction Transformer

Creation of interactions between target/feature lags (lag2 - lag1, for instance).

- Lags Aggregates Transformer

Aggregations of target/feature lags like mean(lag7, lag14, lag21) with support for mean, min, max, median, sum, skew, kurtosis, std.

자료만들기.과정.의외로즐거움

기술영업을 하면서

Wha

The C
https://n
Autome
year. A
over, ar
and not

Autok
www.ri
New, 1
ICCBRI
Machine
Auto-94

Auton
https://n
Autome
applic
must ap
selector
Targets

Googl
https://n
May 23,
of adult
that you
actually

Cloud
https://n
Cloud A

Why

▪ **Featur**
나오는:

Appen

Transformat
engineered I

▪ Filter Tra

The F

▪ Frequent

The F
either

▪ Bulk Inte

The E
data.

▪ Truncate

Trunc
trunc

▪ Cross Val

Cross

▪ Cross Val

This t
encod

Why XAI? - Accuracy vs Interpretability

Focus on Accuracy

Accuracy

Real-world use case

Research on "interpretable models"

Decision Tree

Linear Classifier

$10X_1 + X_2 - 5 > 0$

Interpretability

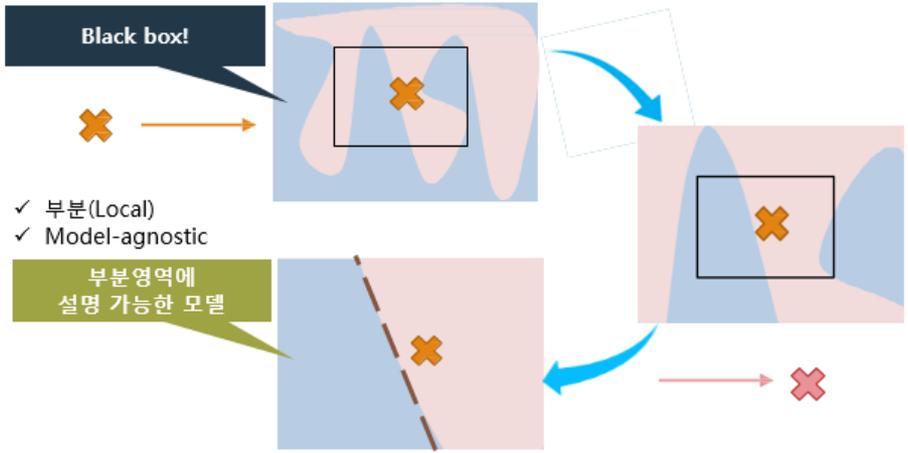
millions of weights, complex features

정확도 ↑ ↓ 해석

↑ 해석 ↓ 정확도

자료만들기.과정.의외로즐거움

기술영업을 하면서

Wha	Why	Why	LIME(Local Interpretable Model-agnostic Explanations)
<p>The C https://n Auton year A over a and not</p> <p>Auton www.ri News, 1 ICCBR) Machine Auto-94</p> <p>Auton https://n Auton applic must ap targets</p> <p>Googl https://n May 23, of auto that sou actually</p> <p>Cloud https://n Cloud A</p>	<p>▪ Featur 나오는:</p> <p>Appen</p> <p>Transformat engineered I</p> <p>▪ Filter Tra The F</p> <p>▪ Frequent The F either</p> <p>▪ Bulk Inte The E data,</p> <p>▪ Truncate Trunc trunc</p> <p>▪ Cross Val Cross</p> <p>▪ Cross Val This t encod</p>	 <p>mil con</p>	<ul style="list-style-type: none"> ▪ Black box에 대한 설명 시도 ▪ 부분영역(local)과 model-agnostic(모델 무관)  <p>✓ 부분(Local) ✓ Model-agnostic</p> <p>부분영역에 설명 가능한 모델</p>
<p>Explair From Wikipedia</p> <p>An Explainabl understood by where even th to implement a</p>			<p>출처: M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135-1144: ACM.</p>
			

자료만들기.과정.의외로즐거움

기술영업을 하면서

What

The C
https://n
Autob
year. A
over, ar
and not

Autob
www.ri
News, 1
ICCBR)
Macran
Machine
Auto-94

Autob
https://n
Autob
applic
must ap
vector
Targets

Googl
https://n
May 23,
of asset
that sou
actually

Cloud
https://n
Cloud A

Explain

From Wikipedia

An Explainabl
understood by
where even th
to implement a

Why

- Feature 나오는:
- Appen
- Transformat engineer
- Filter Tra
- The F
- Frequent
- The F either
- Bulk Inte
- The E data,
- Truncate
- Trunc trunc
- Cross Val
- Cross
- Cross Val
- This t encod

Why



mil
con

LIME(Local Interpretable Model-agnostic Explanations)

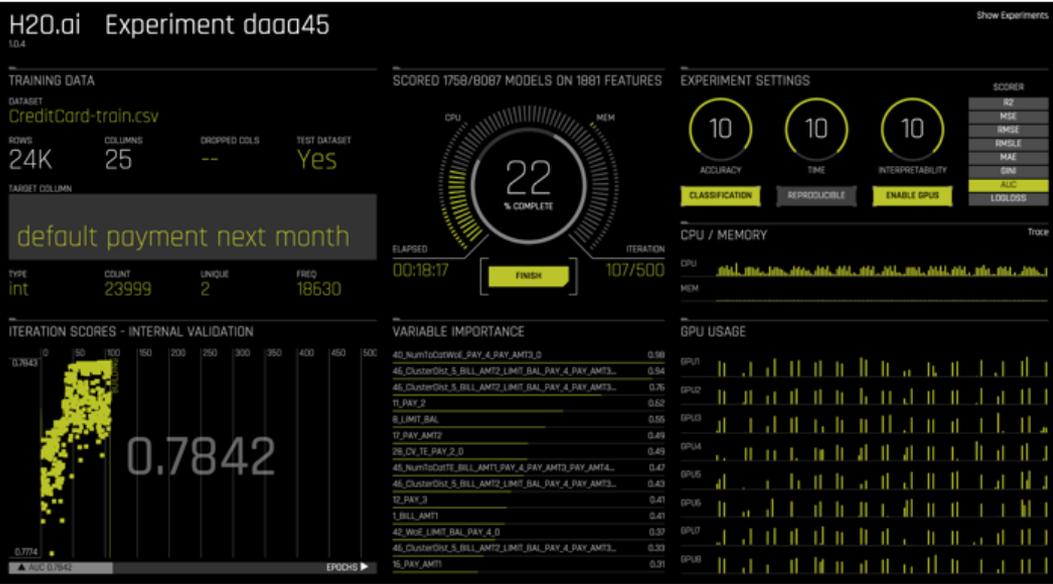
- Black box
- 부분영역(Local)

부분
Mod
설명

출처: M. T. Ribeiro, S. SIGKDD International

Driverless AI

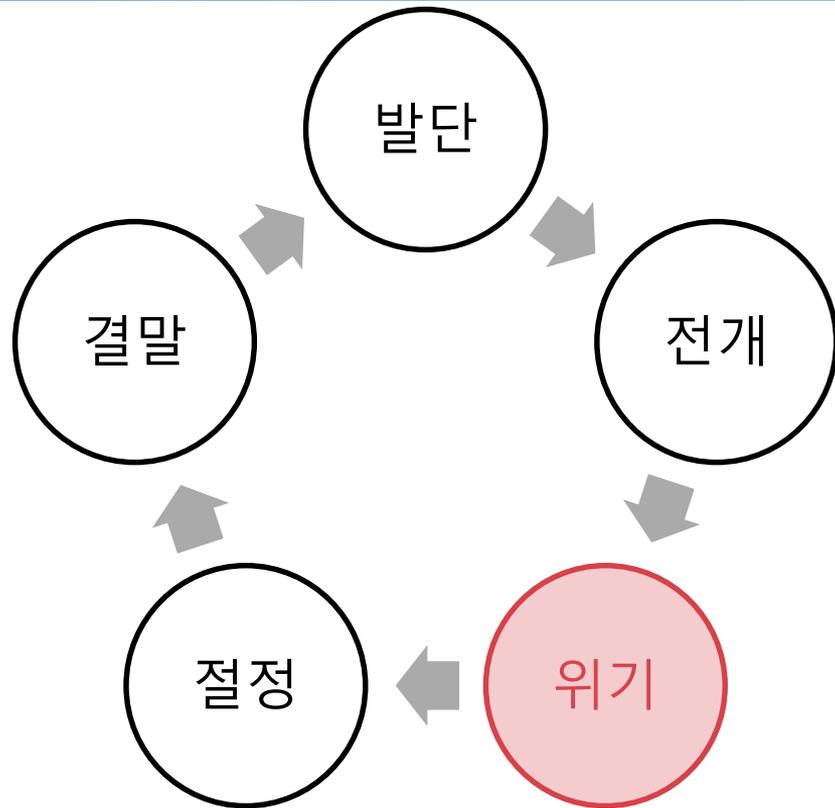
H2O.ai Experiment daaa45



The screenshot shows the H2O.ai Driverless AI interface for an experiment named 'daaa45'. It displays training data for 'CreditCard-train.csv' with 24K rows and 25 columns. A progress indicator shows 22% completion. The interface includes sections for 'EXPERIMENT SETTINGS' (Accuracy: 10, Time: 10, Interpretability: 10), 'VARIABLE IMPORTANCE' (listing features like '40_NumOfCores_Pay_4_Pay_Amt1_0' with importance 0.98), and 'ITERATION SCORES - INTERNAL VALIDATION' showing a score of 0.7842. The bottom of the interface features the Agile SoDA logo.

자료만들기.과정.의외로즐거움

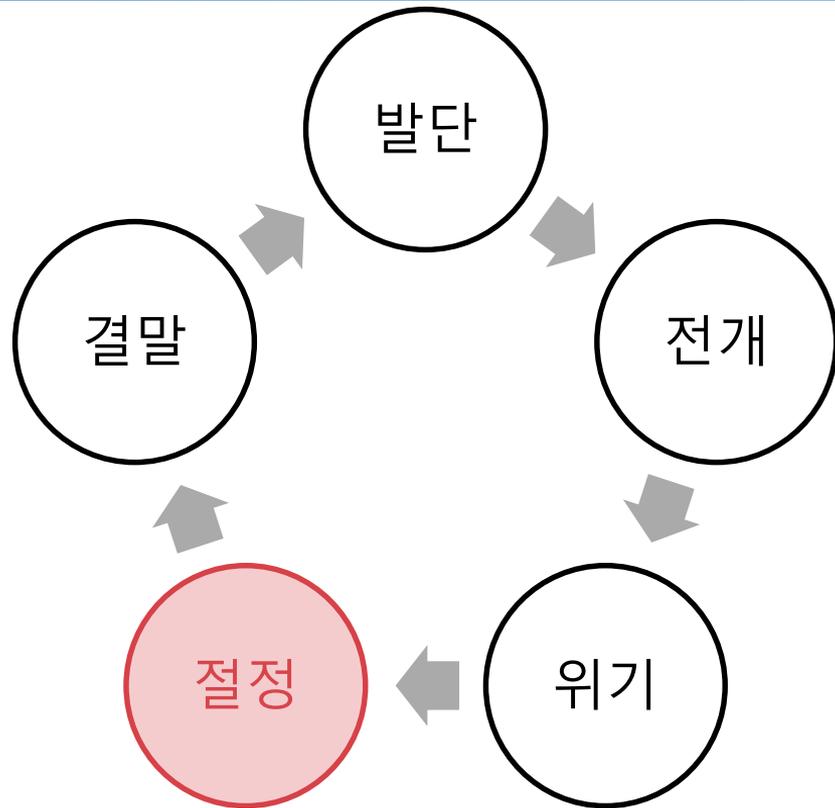
기술영업을 하면서



기술영업을 하면서

발표.긴장감.욕심

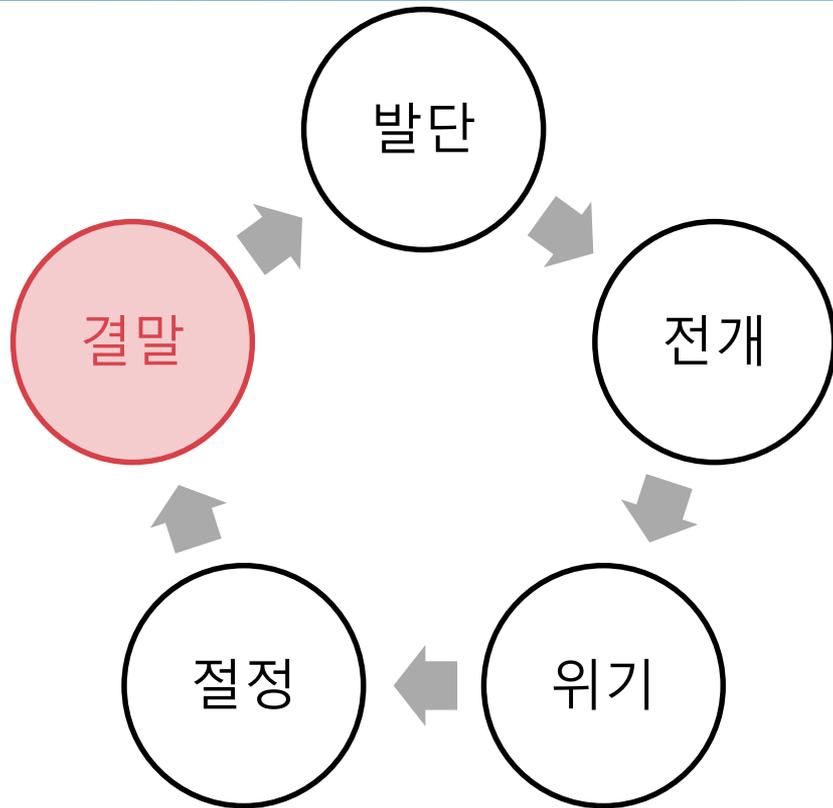
기술영업을 하면서



기술영업을 하면서

좋은느낌.질문.디펜스

기술영업을 하면서



기술영업을 하면서

아쉬움. 그리고 아쉬움.

Q : 화자의 이야기를 통해
얻을 수 있는 교훈은?



기술영업을 하면서

1. 가끔은 다른 일을 해보기
2. 어차피 해야한다면 즐겨보기

공부해서 남들에게 퍼주기



공부해서 남들에게 퍼주기

저는 어렸을 때 이런 말을 많이
들었습니다.

공부해서 남들에게 퍼주기

**공부해서 남에게 주냐?
다 너의 것이 되니까 공부하렴!**

공부해서 남들에게 퍼주기



그런데 공부해서 남들에게 퍼주면
안되는 걸까요?

공부해서 남들에게 퍼주기

오히려 남들에게 퍼줘보아야
확실히 내 것으로 만들 수 있지
않을까요?

공부해서 남들에게 퍼주기

← → ↻ https://rpubs.com/BBSDDSD/simple_h2o_intro_usage ☆ 🔍 🌐 📄 📄

📄 Using the Graph API 📄 먹거리 만드는 열정만 📄 스텔라 확고 📄 Installing and Startin 📄 databricks/learning- 📄 Knitr with R Markdo 📄 2. Working with Dat

RPubs *brought to you by RStudio*

H2O 소개 및 간단한 사용법

h2o 사용을 위한 환경설정
h2o 사용하기
Summary
Next
References

H2O 소개 및 간단한 사용법

안녕하세요. 여러분

제가 소개할 패키지는 h2o입니다.

지난 번 페이스북에서도 말씀드렸듯이 h2o는 java 기반의 머신러닝/AI 플랫폼인데요.

제가 처음 h2o를 사용했던 건 2016년 10월부터 사용하기 시작했는데 그 때 버전이 2.x.x 대였습니다.

그런데 그 때 당시 h2o를 소개하는 documentation이 너무 산재되어있어서 사용했을 때 많이 힘들었습니다.

참 아막같은 패키지라고 생각이 들었던건.. 사용하기 힘들었는데도 불구하고 지금도 계속 사용중인데요.

- 그 이유는 제 개인적인 의견으로 다음과 같습니다.
 - 모델 배포가 편리하다. (POJO / 단, 동일 머신을 사용해야함.)
 - POJO: Plain Old Java Object
 - 확장성이 좋다. (h2o의 클러스터 모드)
 - 모델의 정보를 약간(?) 정리해서 보여준다. (h2o의 UI)

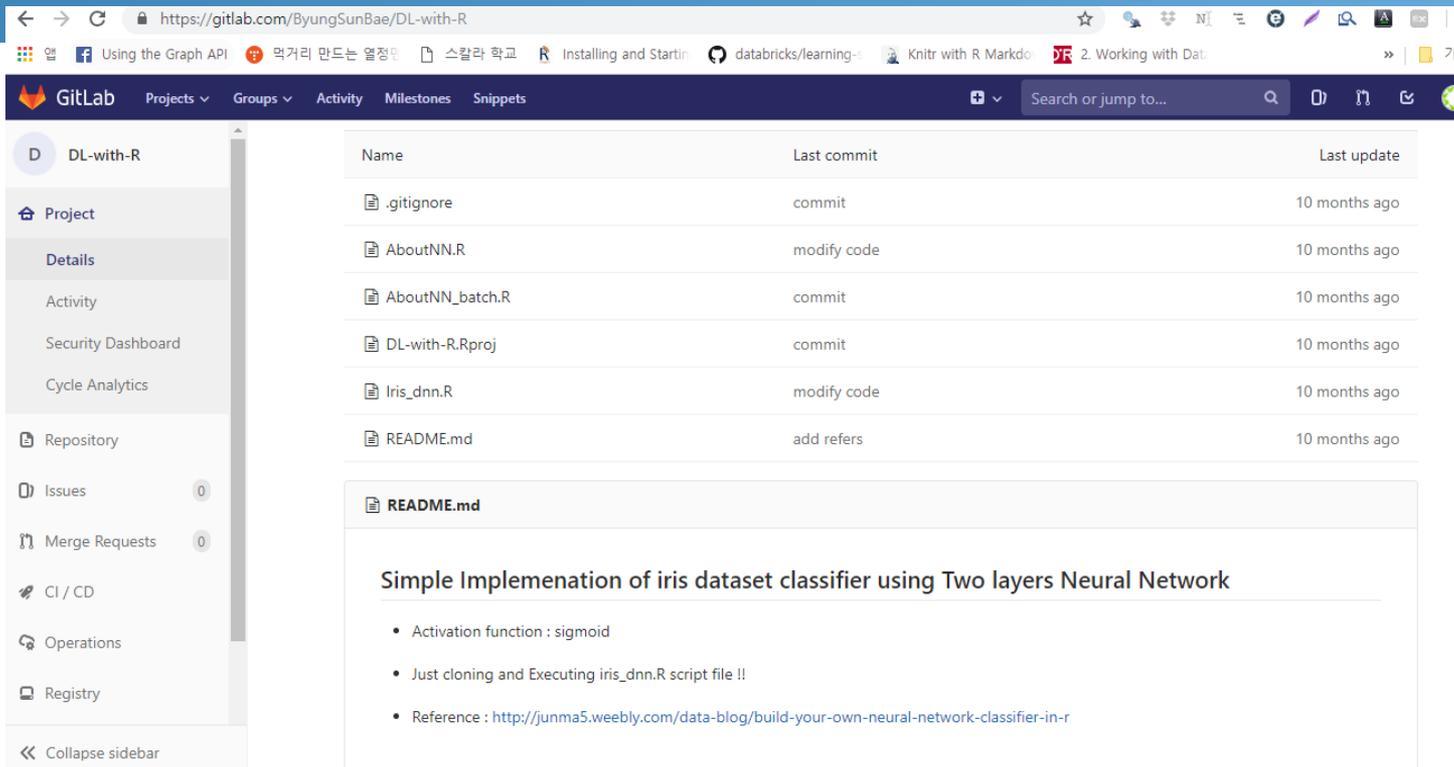
H2O 소개 및 간단한 사용법 자료 만들기

공부해서 남들에게 퍼주기



R과 H2O를 활용한 머신러닝 분석사례 강의중

공부해서 남들에게 퍼주기



The screenshot shows a GitLab repository page for a project named 'DL-with-R'. The left sidebar contains navigation options: Project, Details, Activity, Security Dashboard, Cycle Analytics, Repository, Issues (0), Merge Requests (0), CI / CD, Operations, and Registry. The main content area displays a table of files with columns for Name, Last commit, and Last update. Below the table, the README.md file is expanded, showing the title 'Simple Implementation of iris dataset classifier using Two layers Neural Network' and a bulleted list of details.

Name	Last commit	Last update
.gitignore	commit	10 months ago
AboutNN.R	modify code	10 months ago
AboutNN_batch.R	commit	10 months ago
DL-with-R,Rproj	commit	10 months ago
Iris_dnn.R	modify code	10 months ago
README.md	add refers	10 months ago

README.md

Simple Implementation of iris dataset classifier using Two layers Neural Network

- Activation function : sigmoid
- Just cloning and Executing iris_dnn.R script file !!
- Reference : <http://junma5.weebly.com/data-blog/build-your-own-neural-network-classifier-in-r>

R로 구현해보는 2개의 히든 레이어를 가지는 Neural Network

공부해서 남들에게 퍼주기

GitLab Projects Groups Activity Milestones Snippets Search or jump to...

TensorR Public Add license

Tensorflow in R
Project ID: 4471323

1 Star 0 Fork SSH git@gitlab.com:ByungSunBae

Readme Files (389 KB) Commits (47) Branch (1) Tags (0) Auto DevOps enabled Security Dashboard

Add Changelog Add Contribution guide Add Kubernetes cluster

master TensorR / History Find file Web IDE

debugging : out bounding error
ByungSunBae authored 10 months ago 07d3e0d4

Name	Last commit	Last update
1_introduction	Basic Operations	11 months ago
2_BasicModels	debugging : out bounding error	10 months ago
3_NeuralNetworks	Neural Networks using tensorflow in R	11 months ago
Add_RLinR	change gamma	10 months ago
citizens	first commit	11 months ago

R에서 Tensorflow 사용해보기

Q : 그럼 이렇게 해서
얻는게 있나요?



A : 얻는게 없을 수 있으나
적어도 손해보지는
않습니다.



공부해서 남들에게 퍼주기

페이스북 커뮤니티
사이트에 게시글 작성했다고 가정

공부해서 남들에게 퍼주기

1) 아무런 반응이 없으면?

→ 기다리면서 다른 그룹 및 사람들에게 따로 물어보기

2) 좋아요 및 공유수가 증가한다면?

→ 나와 같은 문제를 겪었던 사람이 있거나 게시한 글의 내용이 좋다는 의미

3) 댓글이 있는 경우

→ (1) 오류 등의 수정사항을 말함

→ 더 자세히 파고들거나 그 상태로 해결가능

→ (2) 또 다른 질문들이 나옴

→ 생각지도 못한 것이 발견되어 생각의 폭이 넓어질 수 있음

목 차

1) 대놓고 홍보하기 : 애자일소다

2) 발표자

3) 경험담.. 들어는 드릴께

4) 결론

- 철학이 뭔가요?





결론



Q & A



byung2070@gmail.com

제가 다니는 회사든
머신러닝이든 분석이든
궁금하신 것이 있다면
위 이메일로 연락을 주시면
일주일내로 답변 드리겠습니다.



감사합니다

