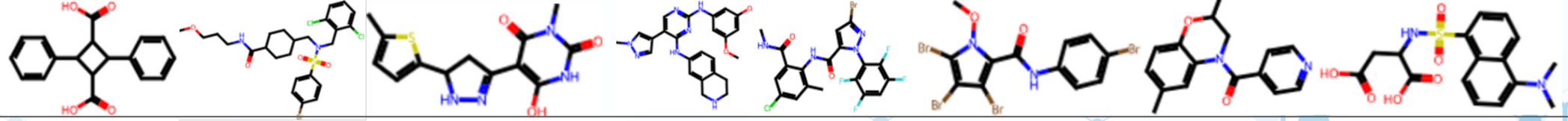
인간과 쥐의 간 세포에 대한 화합물 대사 안정성 예측

목차



01 연구 개요
- 생쥐가 왜? 실험쥐인가?
- 연구 배경 & 목적
- 신약 개발 과정 & Word Cloud
- 신약 허가 후 취소된 사례

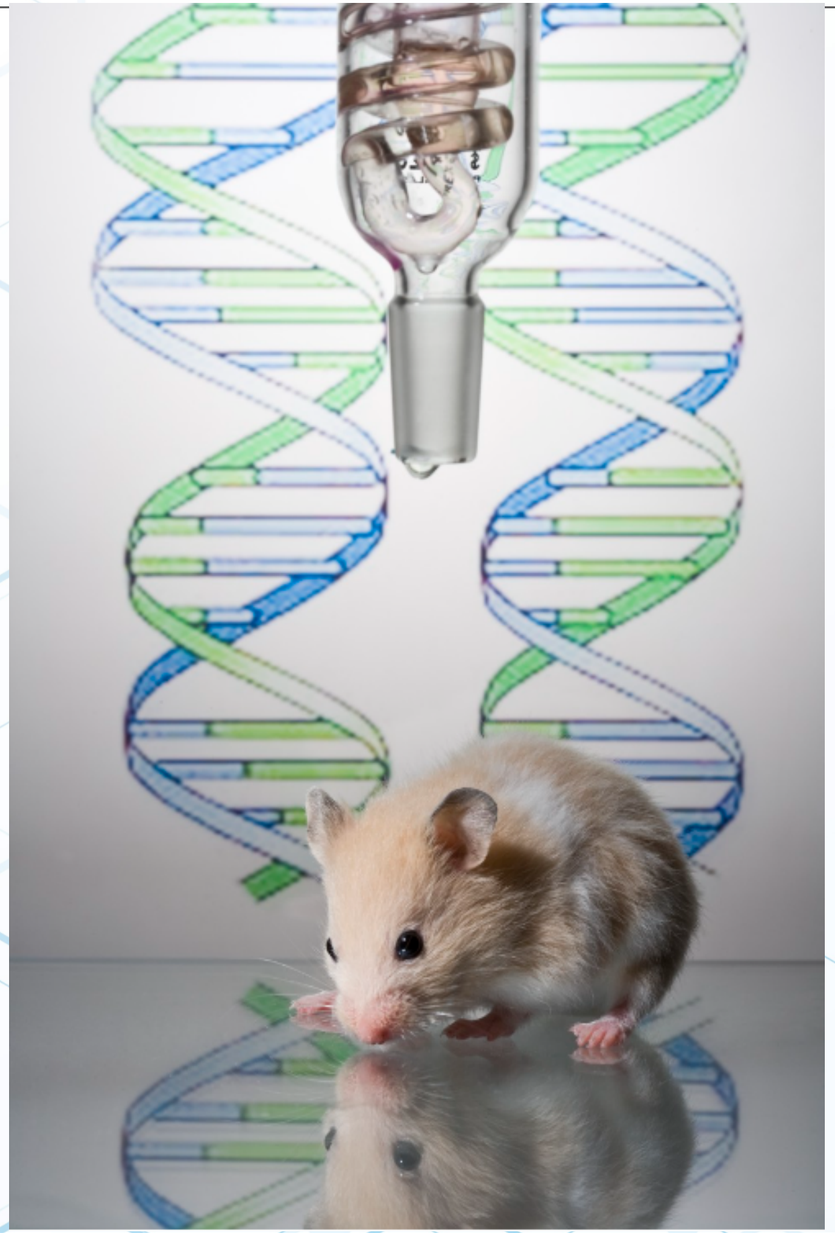
02 시장 조사
- 시장 규모
- 관련 산업 동향

03 데이터
- 전처리
- 시각화

04 모델링
- 모델 결과
- 모델 소개

05 향후 계획
- GCN 모델

01 생쥐가 왜? 실험쥐인가?



사람과 97%나 되는 유전적 상동성(相同)

쥐와 인간의 게놈은 80% 이상 동일하고 90% 이상은 비슷한 위치에 자리 잡고 있다.

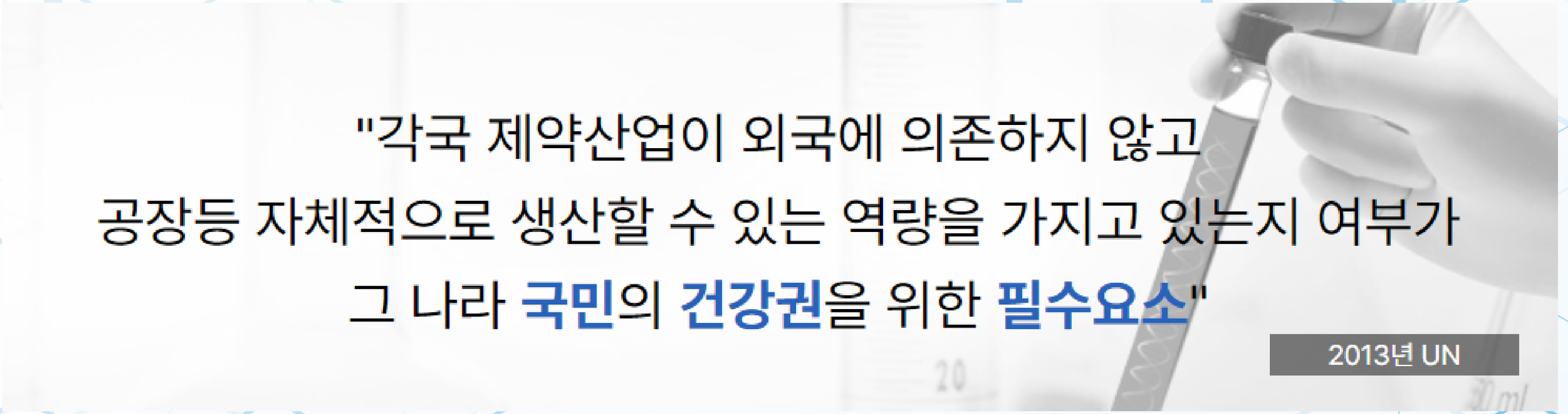
다루기 쉽고 세대가 짧아 실험 결과를 쉽게 확인

쥐의 유전자 연구 역시 인간의 유전자 기능과 역할을 파악하는 데 크게 기여

생명과학자들은 쥐 게놈을 인간 유전자의 '로제타석'으로 비유

https://blog.naver.com/open_kbsi/22179035296

01 연구 배경



"각국 제약산업이 외국에 의존하지 않고
공장등 자체적으로 생산할 수 있는 역량을 가지고 있는지 여부가
그 나라 **국민**의 **건강권**을 위한 **필수요소**"

2013년 UN

https://www.kpbma.or.kr/sub/00000000048/00000000154

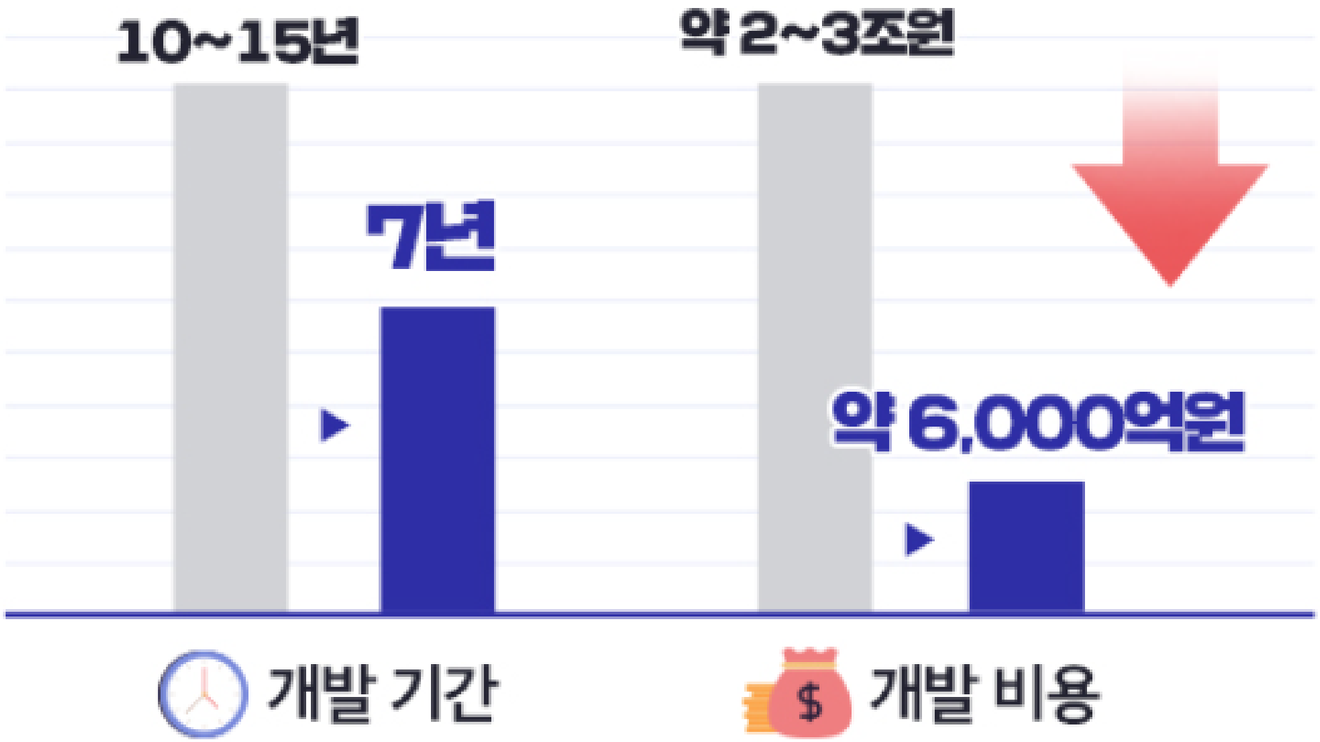
01 연구 배경

신약을 개발하는데 통상 12~15년의 기간이 걸리고 평균 **2조 6000억원**이 투입됩니다.

https://www.kpbma.or.kr/sub/00000000048/00000000154

01 연구 목적

신약 개발 기간 및 비용 단축



https://newsroom.daewoong.co.kr/archives/13497

좀비처럼 죽지않는 '항노화 약물 후보물질' AI로 확인

2023.05.08 16:45

과학자들이 인공지능(AI)을 활용해 노화를 일으키는 체내 작용과 싸우는 약물 후보물질들을 발견했다. 80만개 이상의 화합물을 분석해 추려진 이들 물질은 주변 환경의 방해에도 마치 '좀비'처럼 살아남아 항노화 효과를 보이는 것으로 나타났다. 연구팀은 "현재 연구되고 있는 항노화 화합물 중 임상시험에서 성공할 가능성이 가장 높은 물질"이라고 자신했다.

AI는 80만개의 후보물질 중 3개의 후보물질을 제시했다. 분석 결과 이 후보물질들은 먹어서 섭취해도 항노화 작용이 유지됐으며 적혈구를 파괴하거나 유전독성을 일으키지 않았다. 기존 항노화물질과 달리 주변 환경에 영향을 받지 않고 효과를 유지한 것이다.

http://m.dongascience.com/news.php?idx=59725

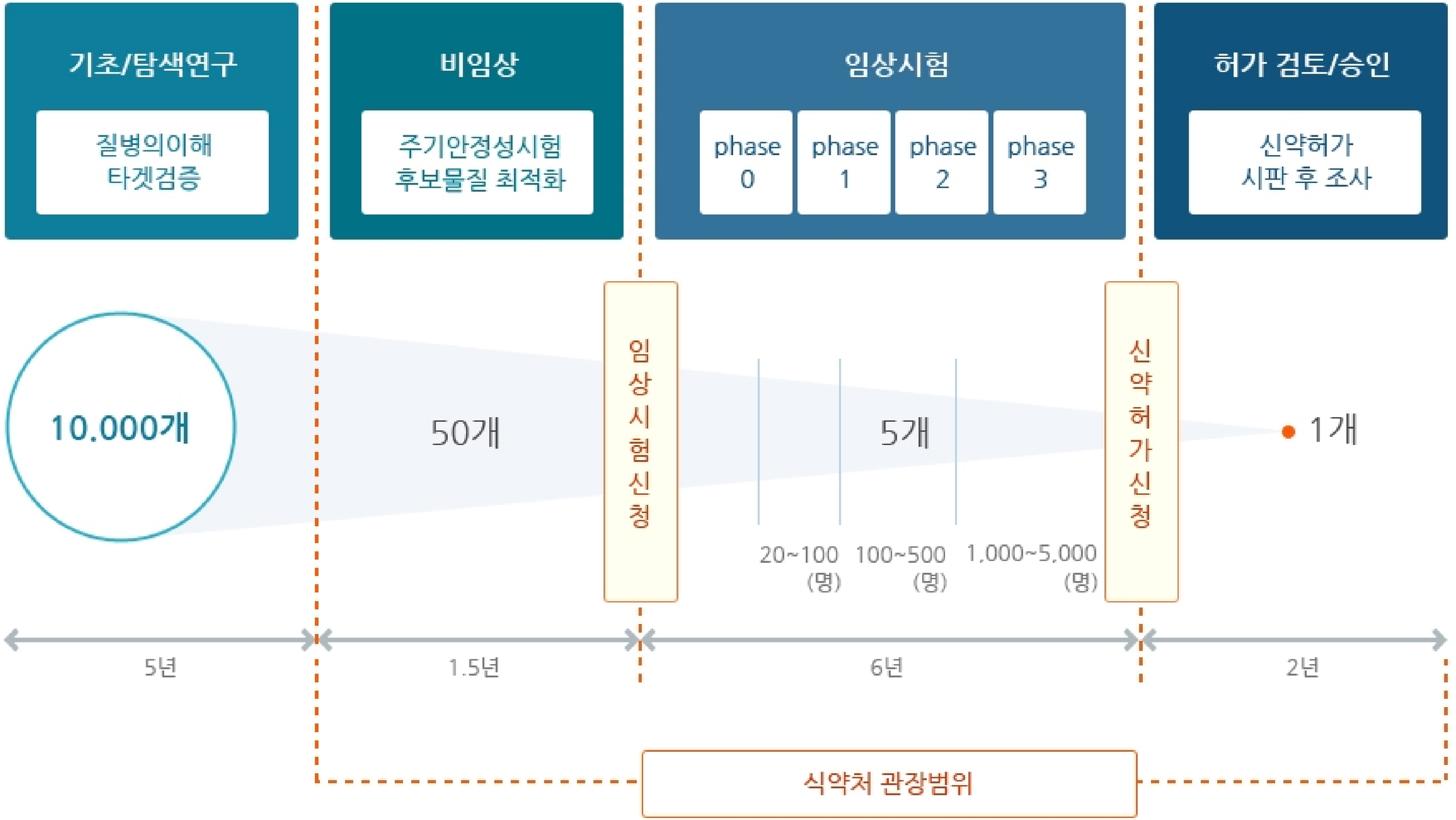
01 신약 개발 워드 클라우드

01 신약 개발 과정

https://www.iworldtoday.com/news/articleView.html?idxno=403172

[사진= 한미약품]

01 신약 개발 과정



https://www.iworldtoday.com/news/articleView.html?idxno=403172

신약개발과정 출처= 의약품통합정보시스템]

01 신약 허가 후 취소된 사례

올리타 일지

2004년	한미약품 개발 착수
2008년	한미약품 후보물질 도출
2015년	베링거인겔하임에 기술수출
2016년	식약처 조건부 허가
	베링거인겔하임 기술반환
2018년	한미약품 개발중단
2022년	허가 취소 및 급여 삭제

그래픽: 김지영 디자인기자

https://news.mt.co.kr/mtview.php?no=2022092910540472976

폐암 치료제 '올리타'(성분명 올무티닙)

02 시장 규모

의약품 1,400조원

>

반도체 500조원

출처 : IQVIA Market Prognosis 2019

https://www.kpbma.or.kr/sub/00000000048/00000000149

02 관련 산업 동향

2020.11.12

대웅제약

국외 인공지능(AI)을 활용한 신약개발 사례

얀센

Janssen

영국 베네볼런트와
인공지능으로 개발한
신약후보물질
임상 2상에 진입

리커전 파마슈티컬스

RECURSION

인공지능 기술 활용해
뇌해면상 혈관기형 치료물질
임상 1상 승인 획득

23andME

23andMe

소비자의 유전체 빅데이터 기반
인공지능, 머신러닝 통해
신약개발 가능한 항체 개발

출처: 한국보건산업진흥원, 보건산업 브리프 vol.314
「인공지능(AI)을 활용한 신약개발 국내·외 현황과 과제」(2020.09)

https://newsroom.daewoong.co.kr/archives/9082

02 관련 산업 동향

2020.11.12

대웅제약

국내 인공지능(AI)을 활용한 신약개발 사례

대웅제약

A2A 파마사와
항암 신약 공동연구 통해
AI 활용한
항암 신약 후보물질 발굴

한미약품, SK케미칼 등
AI 플랫폼 기업과 협업 통해
신약개발 진행

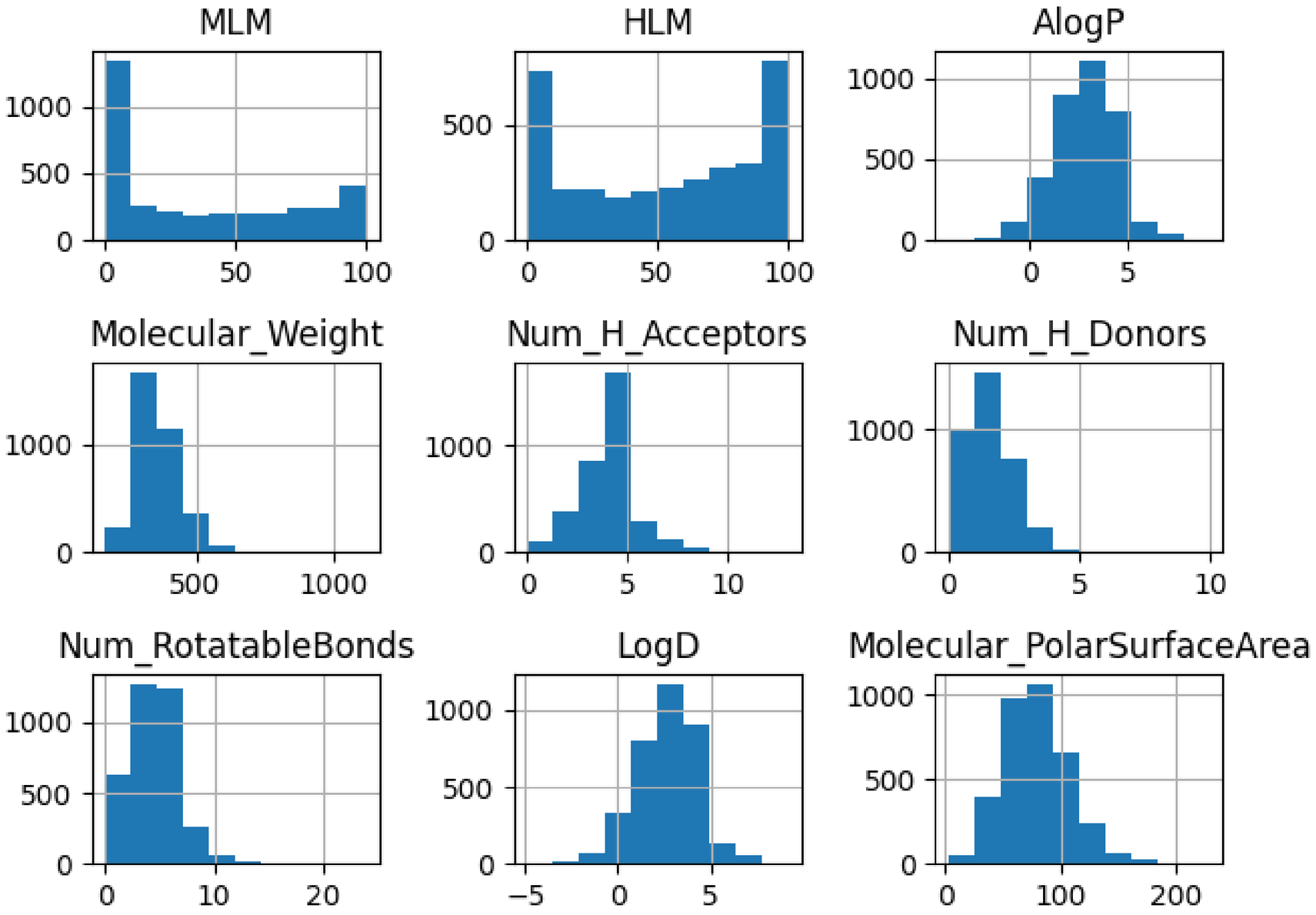
출처: 한국보건산업진흥원, 보건산업 브리프 vol.314
「인공지능(AI)을 활용한 신약개발 국내·외 현황과 과제」(2020.09)

https://newsroom.daewoong.co.kr/archives/9082

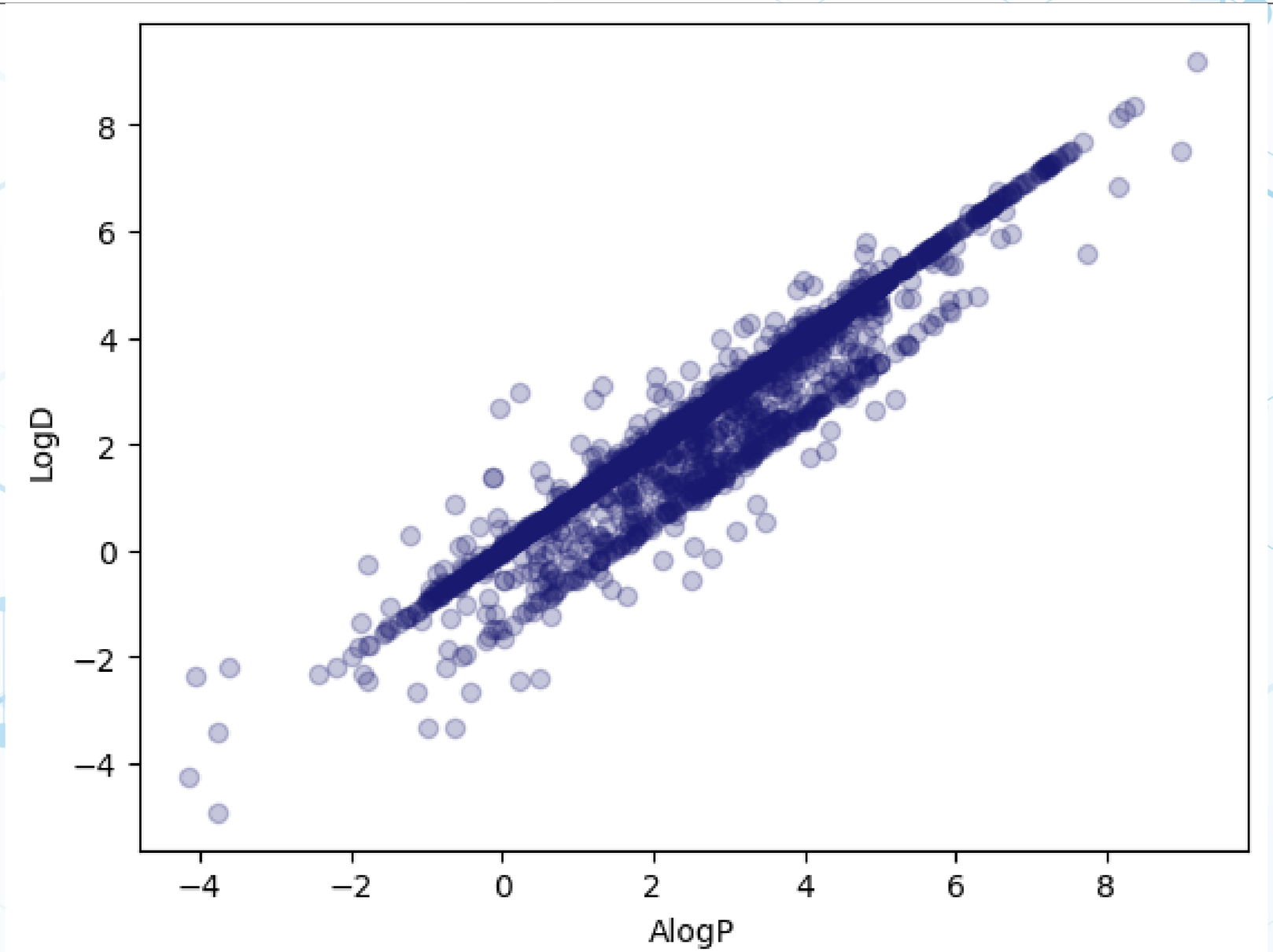
03 데이터 설명

Columns	id	SMILES	MLM	HLM	AlogP	Molecular_ weight	Num_H_ Acceptors	Num_H_Donors	Num_Rotatable Bonds	LogD	Molecular_ PolarSurfaceArea
Abbreviation		Simplified Molecular Input Line Entry System	Mouse Liver Microsome	Human Liver Microsome	logarithm partition-coefficient		number of hydrogen acceptor	number of hydrogen bond donor		logarithm Distributio n-coefficien t	
Description	화합물 id	분자 구조를 string 형태로 표현	마우스의 간 대사 효소와 화합물을 30분 동안 반응 시킨 후, 대사 되지 않고 남아있는 화합물의 양 (%)	인간의 간 대사 효소와 화합물을 30분 동안 반응 시킨 후, 대사 되지 않고 남아있는 화합물의 양 (%)	화학물질의 지방 수용성	분자량	수소 수용체	수소 공여체	회전 가능한 bond, 단일 결합인 경우 회전 가능, 안정성을 따져야한다.	물과 유기용매 사이의 분배 계수	분자에서 극성 원자가 부착 된 표면
Units	string	string	%	%	float	g/mol	개	개	번	float	Å^2
Metric					logP > 5, target에 흡수되기 어렵다.	MW <500, 350~ 400이 적절	Num_H_Accept or > 10, target에 흡수되기 어렵다.	Num_H_Donors >5, target에 흡수되기 어렵다, (HB acceptor에 bonding할 가능성이 높기 때문에)	Num_R >=10, 약효 떨어진다. 7개 이하 적정		PSA>140, 세포 투과 능력이 낮다, 약효 떨어진다. (중추신경계의 경우, 90 옹스트롬 제곱 미만의 PSA가 필요하다.)

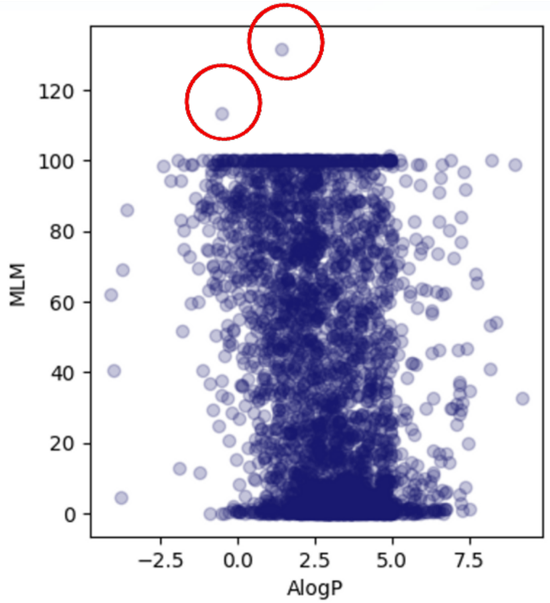
03 데이터 설명

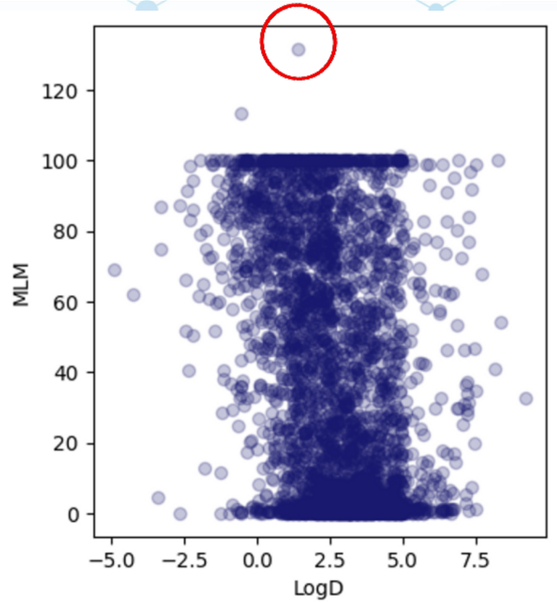

MLM
HLM
AlogP
Molecular_Weight
Num_H_Acceptors
Num_H_Donors
Num_RotatableBonds
LogD
Molecular_PolarSurfaceArea

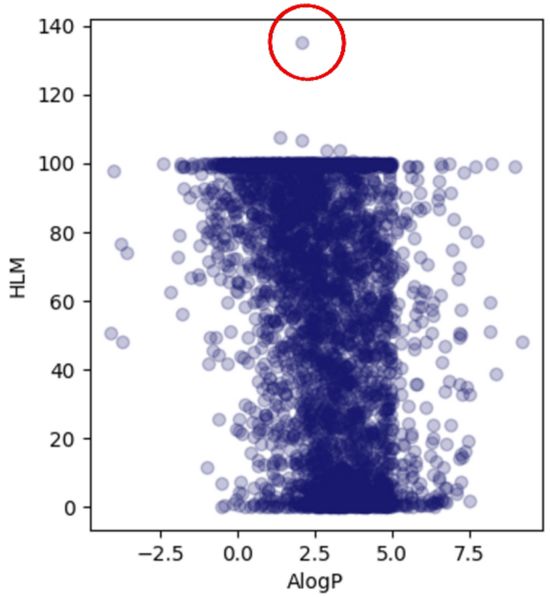
03 데이터 설명

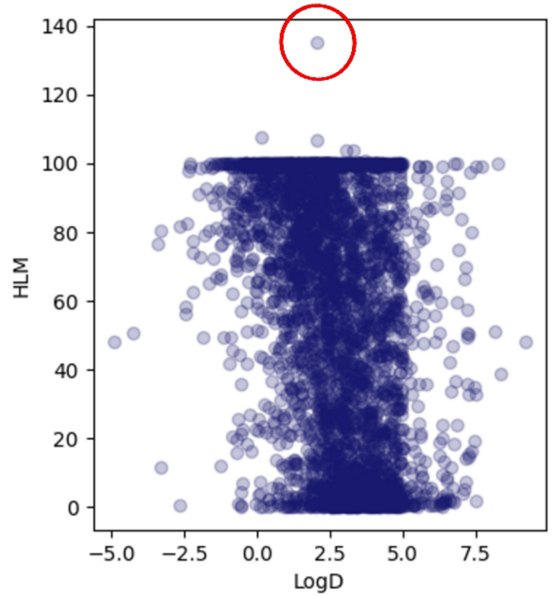


03 데이터 설명


MLM
AlogP
0
20
40
60
80
100
120
−2.5
0.0
2.5
5.0
7.5


MLM
LogD
0
20
40
60
80
100
120
−5.0
−2.5
0.0
2.5
5.0
7.5


HLM
AlogP
0
20
40
60
80
100
120
140
−2.5
0.0
2.5
5.0
7.5


HLM
LogD
0
20
40
60
80
100
120
140
−5.0
−2.5
0.0
2.5
5.0
7.5

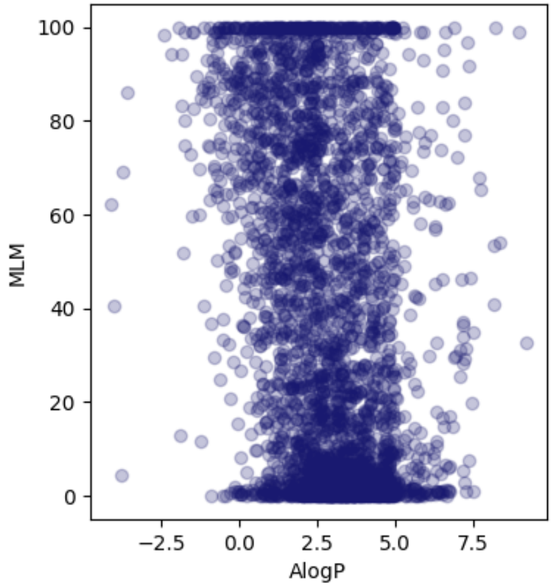
03 데이터 설명

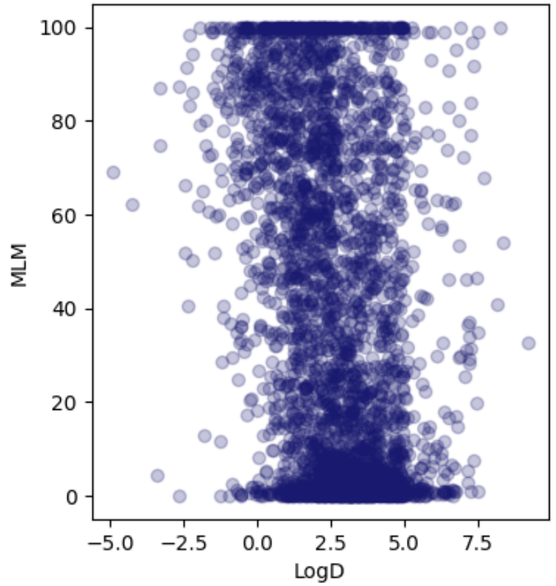
```
1 print(len(df[(df['MLM'] > 100.0) | (df['HLM'] > 100.0)]))
2 display(df[(df['MLM'] > 100.0) | (df['HLM'] > 100.0)])
```

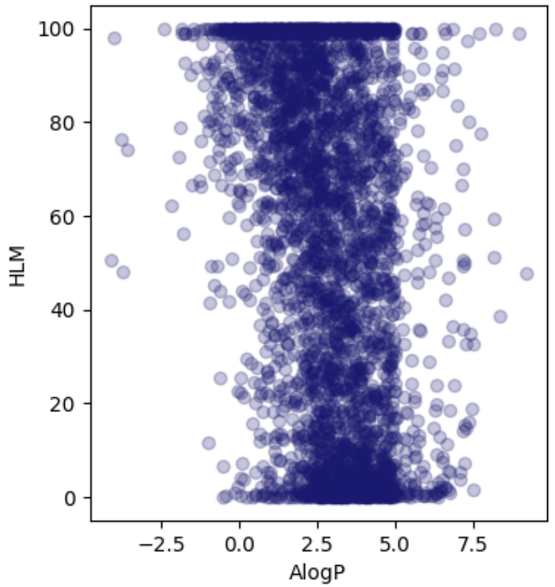
9

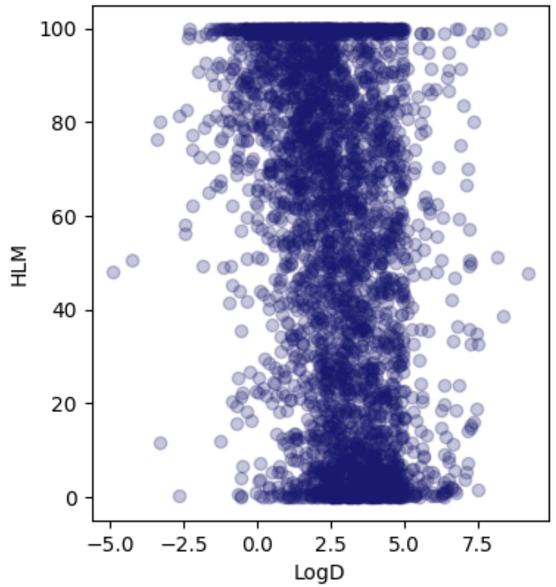
	id	SMILES	MLM	HLM	AlogP	Molecular_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotatableBonds	LogD	Molecular_PolarSurfaceArea
662	TRAIN_0662	O=C(Nc1ccccc1)C1CCCN1C1=NS(=O)(=O)c2ccccc21	131.720	99.570	1.412	355.411	4	1	3	1.412	87.22
1092	TRAIN_1092	COc1c(NC(=O)c2ccc[nH]2)cc(Cl)cc1C(=O)N(C)C	3.820	106.510	2.061	321.759	3	2	4	2.061	74.43
1584	TRAIN_1584	CN(C)S(=O)(=O)CCNCc1ccc(-c2ccccc2)cc1	1.745	135.336	2.075	318.434	3	1	7	2.052	57.79
2159	TRAIN_2159	N#Cc1nccnc1OC1CCN(C(=O)N2CCNC2=O)C1	113.210	99.000	-0.533	302.289	6	1	2	-0.533	111.44
2410	TRAIN_2410	CC(C)CCC1CCN(C(=O)C2CC(O)CN2)CC1	86.878	107.323	1.345	268.395	3	2	4	0.139	52.57
2586	TRAIN_2586	Cc1nc(-c2c[nH]c(C(=O)N3CCOc4cc(F)ccc43)c2)cs1	98.550	103.720	2.876	343.375	3	1	2	3.032	86.46
2948	TRAIN_2948	CCCC(=O)Nc1cc(C(=O)NC2CCCCC2)ccc1S(=O)(=O)c1cc...	101.380	52.330	4.906	462.989	4	2	7	4.906	100.72
3157	TRAIN_3157	CN1C(=O)c2cccc3c2C1=Cc1ccc2ccccc2c1O3	52.847	103.907	3.319	299.323	2	0	0	3.319	29.54
3403	TRAIN_3403	c1cnc2c(C3NCCc4c3[nH]c3ccccc43)cccc2c1	8.890	100.830	3.719	299.369	2	2	1	3.719	40.71

03 데이터 설명

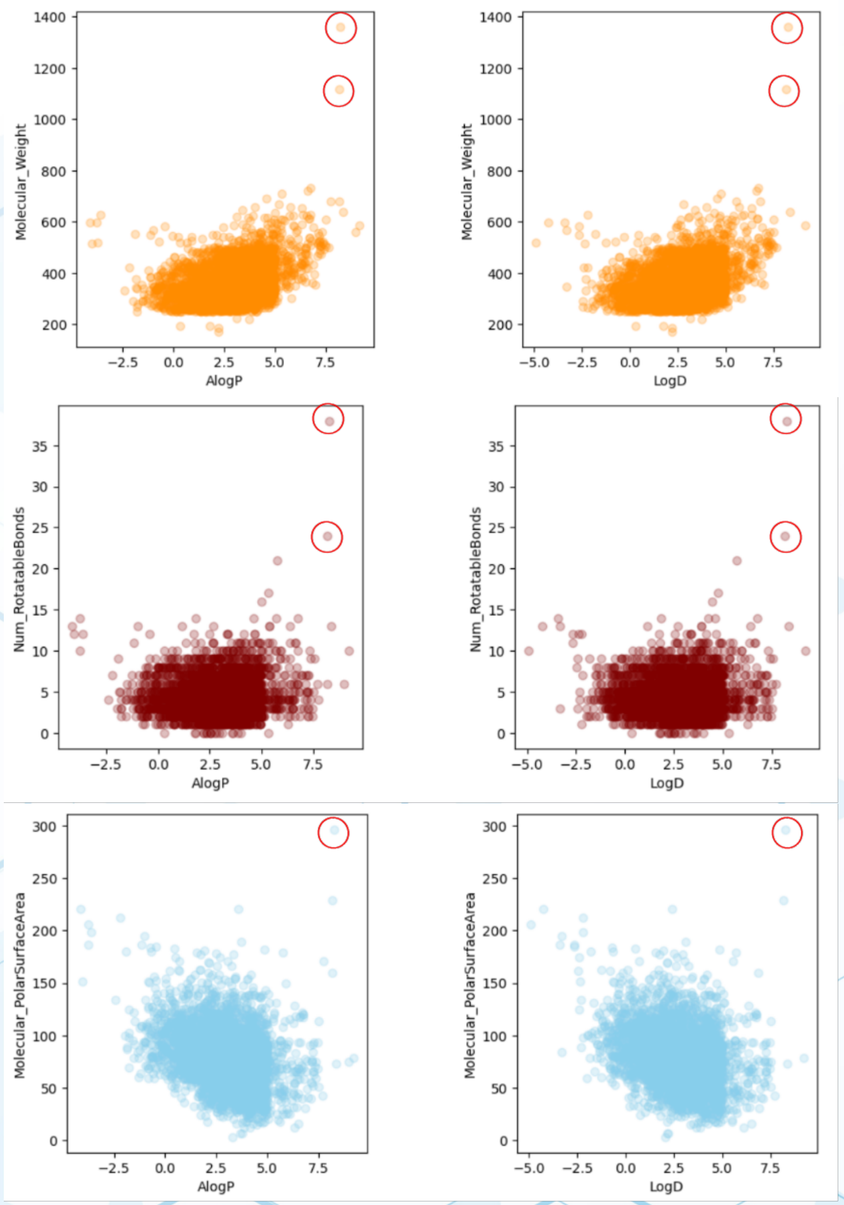

MLM
AlogP
100
80
60
40
20
0
−2.5
0.0
2.5
5.0
7.5

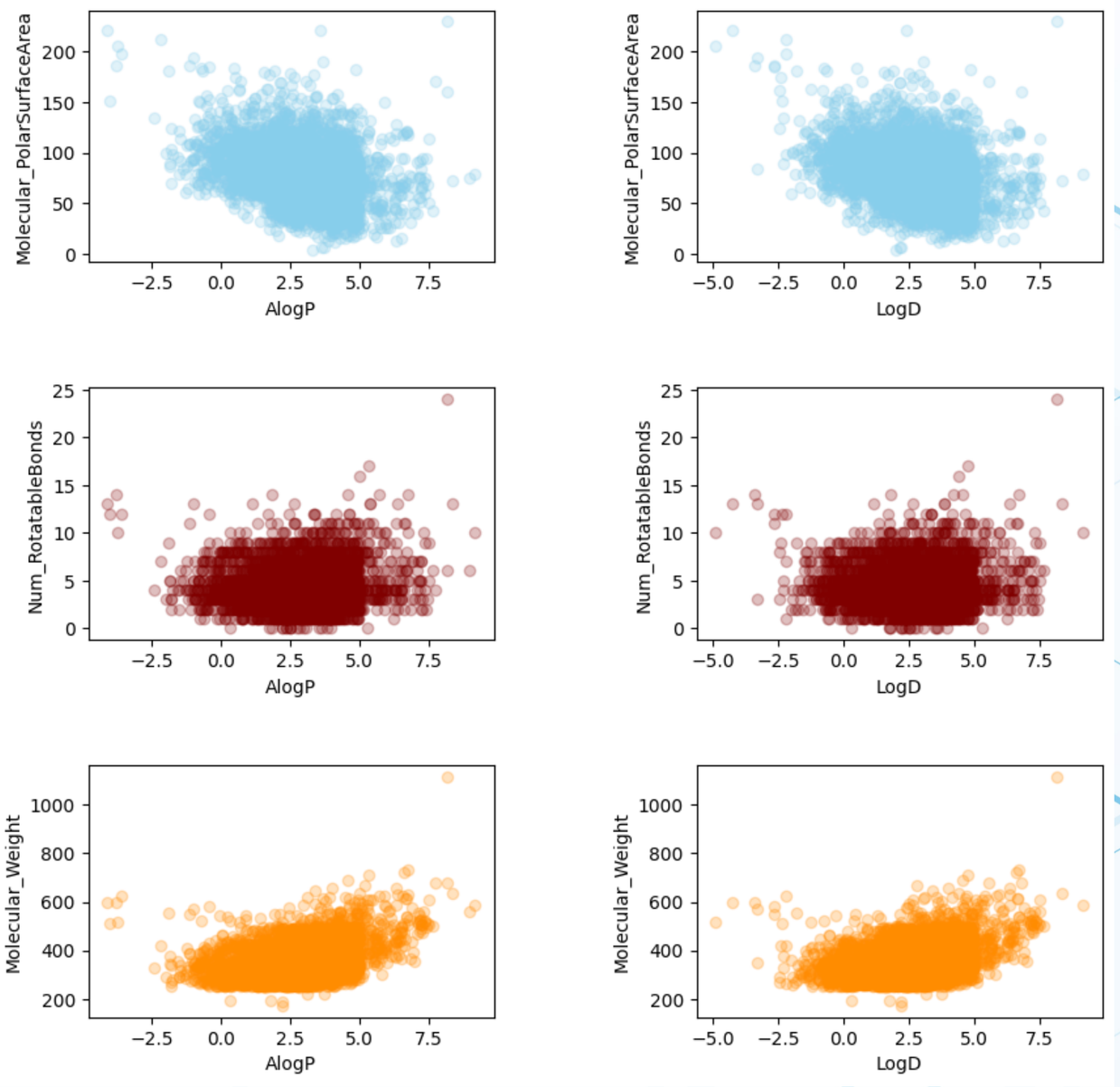

MLM
LogD
100
80
60
40
20
0
−5.0
−2.5
0.0
2.5
5.0
7.5


HLM
AlogP
100
80
60
40
20
0
−2.5
0.0
2.5
5.0
7.5


HLM
LogD
100
80
60
40
20
0
−5.0
−2.5
0.0
2.5
5.0
7.5

03 데이터 설명


Molecular_Weight
AlogP
LogD
Num_RotatableBonds
Molecular_PolarSurfaceArea


Molecular_PolarSurfaceArea
AlogP
LogD
Num_RotatableBonds
Molecular_Weight

03 데이터 설명

```
1 df[df["AlogP"].isnull()]
```

	id	SMILES	MLM	HLM	AlogP	Molecular_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotatableBonds	LogD	Molecular_PolarSurfaceArea
2796	TRAIN_2796	[H][C@]1(CC[C@@]2([H])[C@@H](C)C=CC3=C[C@H](C)...	0.549	0.2	NaN	418.566	5	1	7	4.634	72.83
3387	TRAIN_3387	COc1cc2c(cc1OC)/C(=N\#c1ccccc1)N(Cc1ccccc1F)CC2	0.240	3.2	NaN	390.450	3	0	5	4.911	34.06

03 데이터 설명

```
--------------------------
index : 2788
--------------------------
decision tree의 AlogP : [4.66665152]
random forest의 AlogP : [4.68580233]
lightgbm의  AlogP : [4.66437394]
--------------------------
index : 3377
--------------------------
decision tree의 AlogP : [4.90260656]
random forest의 AlogP : [4.79750715]
lightgbm의  AlogP : [4.895477]
```

```
([array([4.60359016]),
  array([4.68231983]),
  array([4.69181691]),
  array([4.56201974]),
  array([4.67999946]),
  array([4.8023132]),
  array([4.6012795]),
  array([4.6775005]),
  array([4.67372927]),
  array([4.57192949]),
  array([4.70309694]),
  array([4.6643199]),
  array([4.66665152]),
  array([4.68580233]),
  array([4.66437394])],
 [array([4.89463889]),
  array([4.89571932]),
  array([5.02071945]),
  array([4.88079333]),
  array([4.85331343]),
  array([5.05022853]),
  array([4.90208661]),
  array([4.81299403]),
  array([4.90457657]),
  array([4.87675177]),
  array([4.86932709]),
  array([5.07650801]),
  array([4.90260656]),
  array([4.79750715]),
  array([4.895477])])
```

03 데이터 설명

```
1 df[df["AlogP"].isnull()]
```

	id	SMILES	MLM	HLM	AlogP	Molecular_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotatableBonds	LogD	Molecular_PolarSurfaceArea
2796	TRAIN_2796	[H][C@]1(CC[C@@]2([H])[C@@H](C)C=CC3=C[C@H](C)...	0.549	0.2	NaN	418.566	5	1	7	4.634	72.83
3387	TRAIN_3387	COc1cc2c(cc1OC)/C(=N₩c1ccccc1)N(Cc1ccccc1F)CC2	0.240	3.2	NaN	390.450	3	0	5	4.911	34.06

```
1 df.loc[idx_null[0], 'AlogP'] = np.mean(alop_list1)
2 df.loc[idx_null[1], 'AlogP'] = np.mean(alop_list2)
```

```
1 df.loc[idx_null]
```

	id	SMILES	MLM	HLM	AlogP	Molecular_Weight	Num_H_Acceptors	Num_H_Donors	Num_RotatableBonds	LogD	Molecular_PolarSurfaceArea
2788	TRAIN_2796	[H][C@]1(CC[C@@]2([H])[C@@H](C)C=CC3=C[C@H](C)...	0.549	0.2	4.662050	418.566	5	1	7	4.634	72.83
3377	TRAIN_3387	COc1cc2c(cc1OC)/C(=N₩c1ccccc1)N(Cc1ccccc1F)CC2	0.240	3.2	4.908883	390.450	3	0	5	4.911	34.06

04 모델 결과

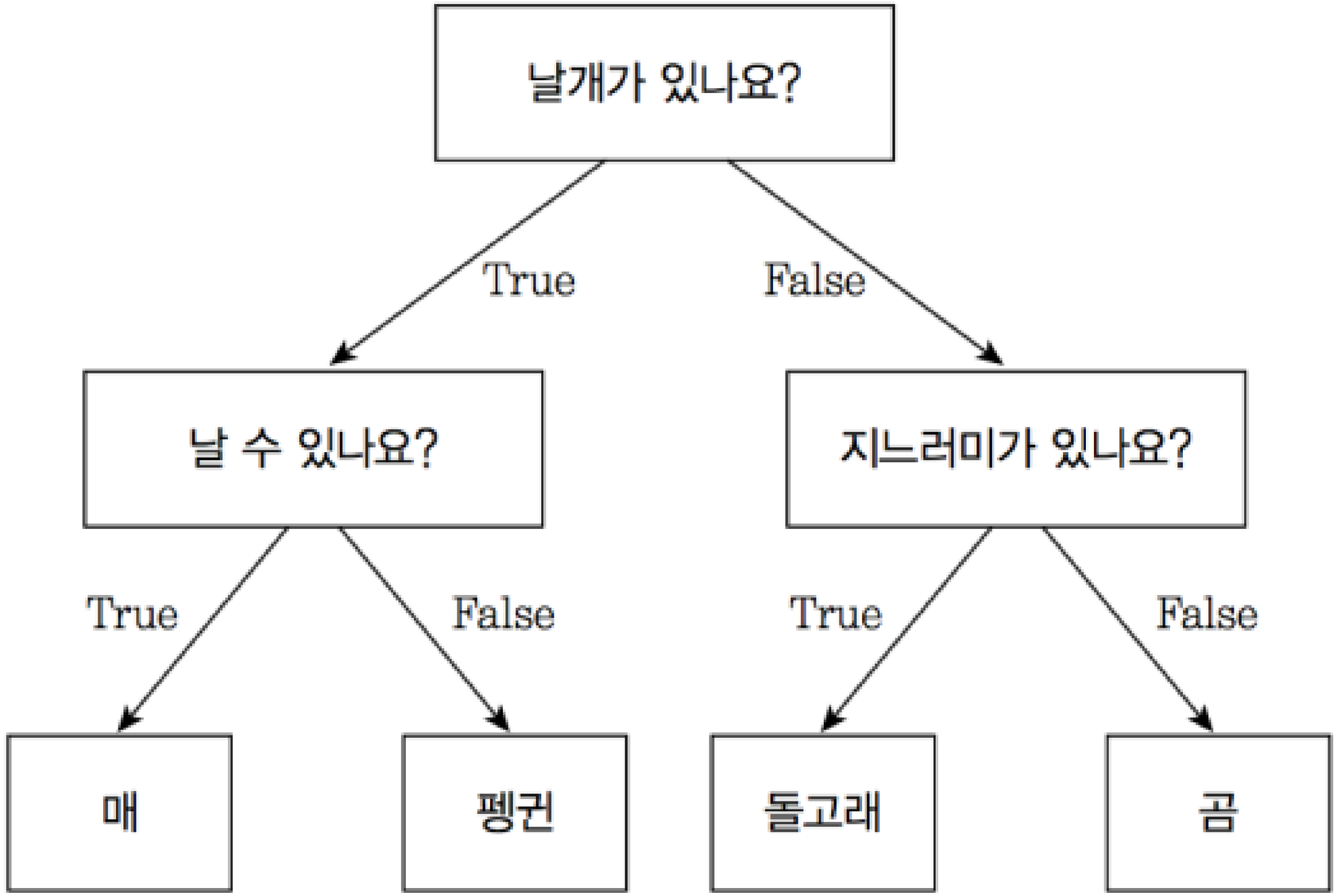
	제목	제출 일시	점수	제출선택
891161	**submission.csv** 2nd_commit(ml, RMSE: [h]33.076, [m]32.047) **edit**	2023-08-16 18:32:11	29.1169495115	☑
891085	**submission.csv** first_commit(GCN, RMSE: [h]36.194, [m]32.4348) **edit**	2023-08-16 17:18:12	38.6026354749	☐

WINNER 1% 4% 10%

전체 랭킹 >

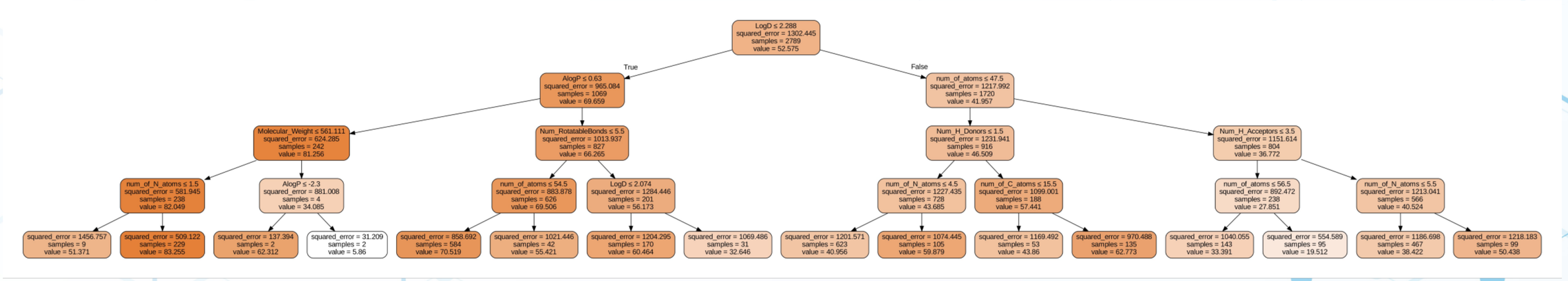
#	팀	팀 멤버	점수	제출수	등록일
39	감기조심	감기 태카 M0	29.11694	5	2분 전
1	hjm9702	hj	26.83821	21	2시간 전

04 결정 트리

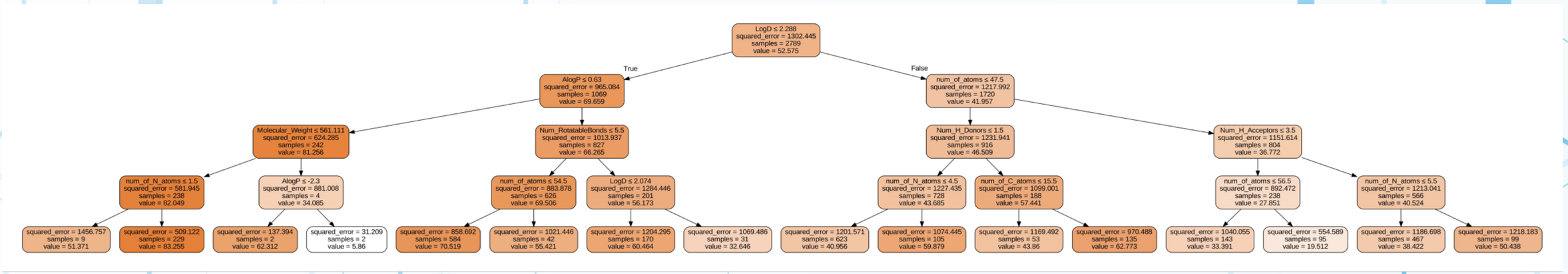


https://bkshin.tistory.com/entry/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-4-%EA%B2%B0%EC%A0%95-%ED%8A%B8%EB%A6%ACDecision-Tree

04 결정 트리



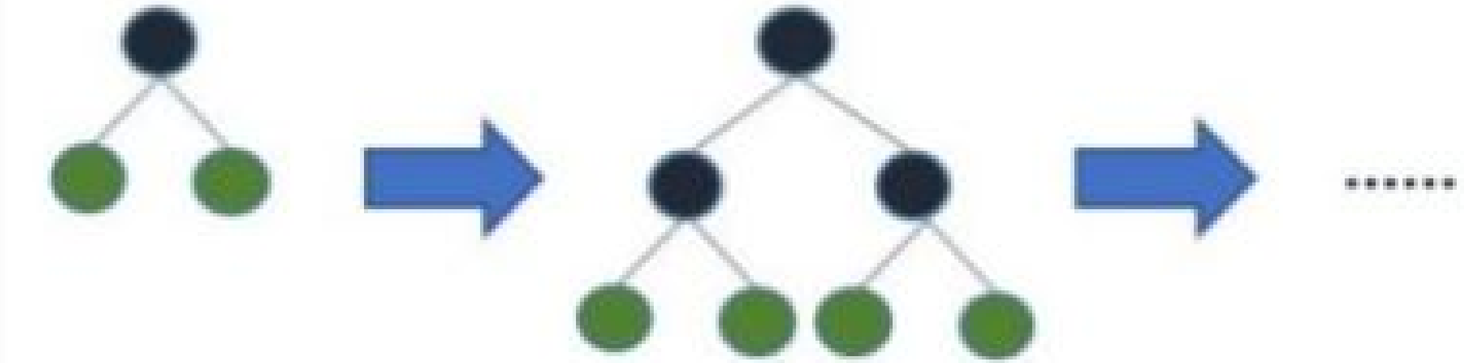
HLM decision tree



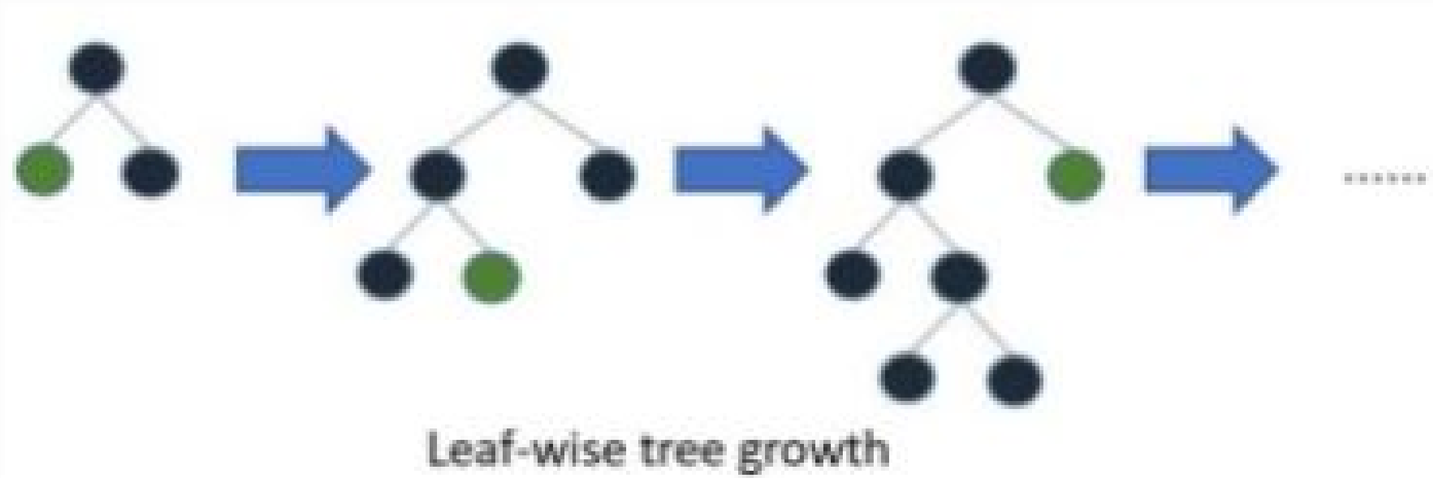
MLM decision tree

04 Light GBM vs XGBoost

XGBoost:



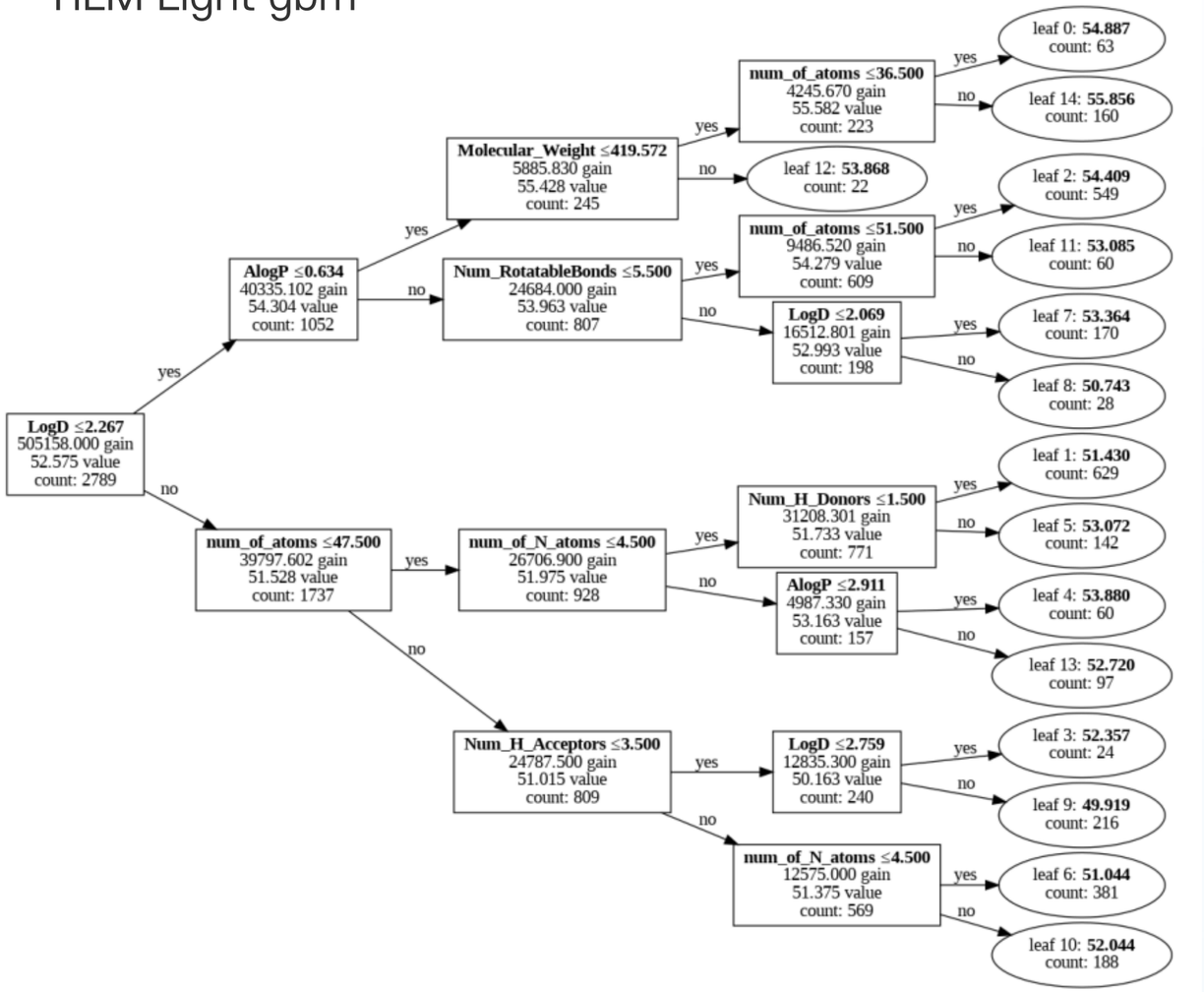
LightGBM:

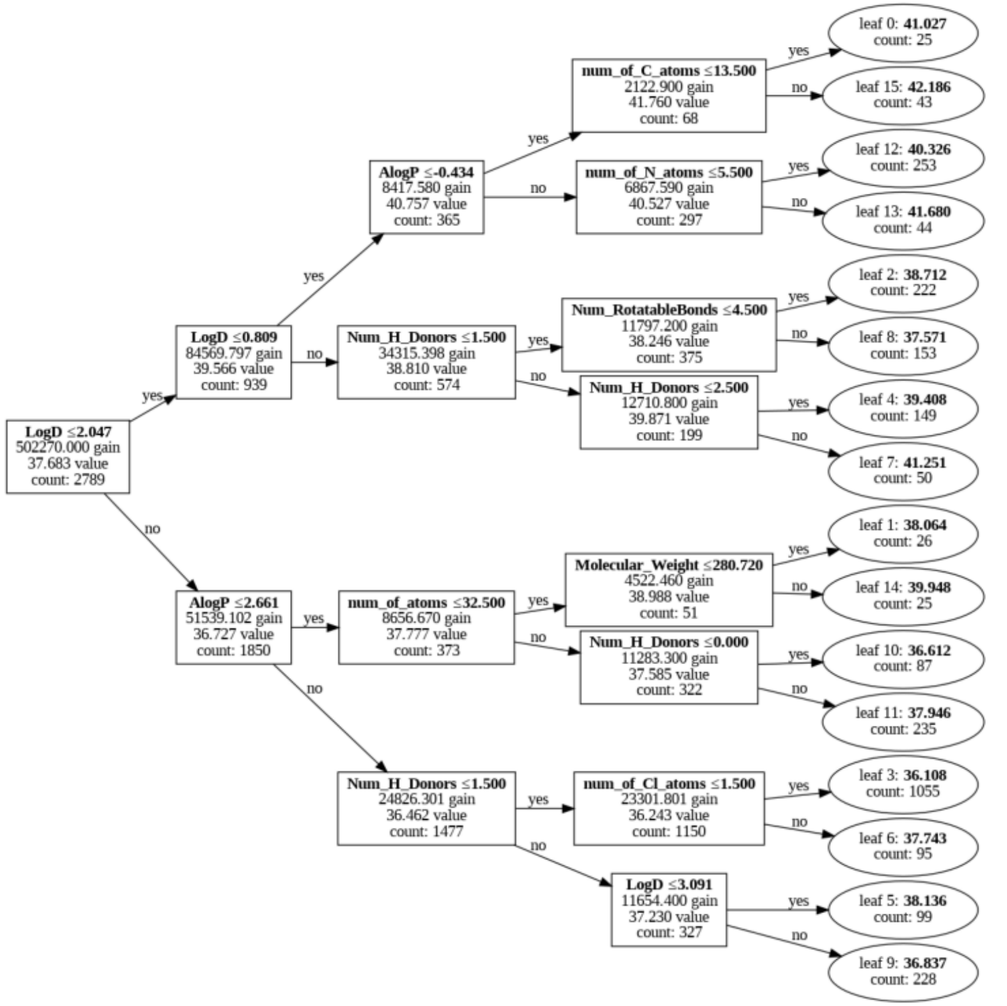


https://heeya-stupidbutstudying.tistory.com/entry/ML-GBM-%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98-%EB%B0%8F-LightGBM-%EC%86%8C%EA%B0%9C-%EA%B8%B0%EB%B3%B8%EA%B5%AC%EC%A1%B0-parameters

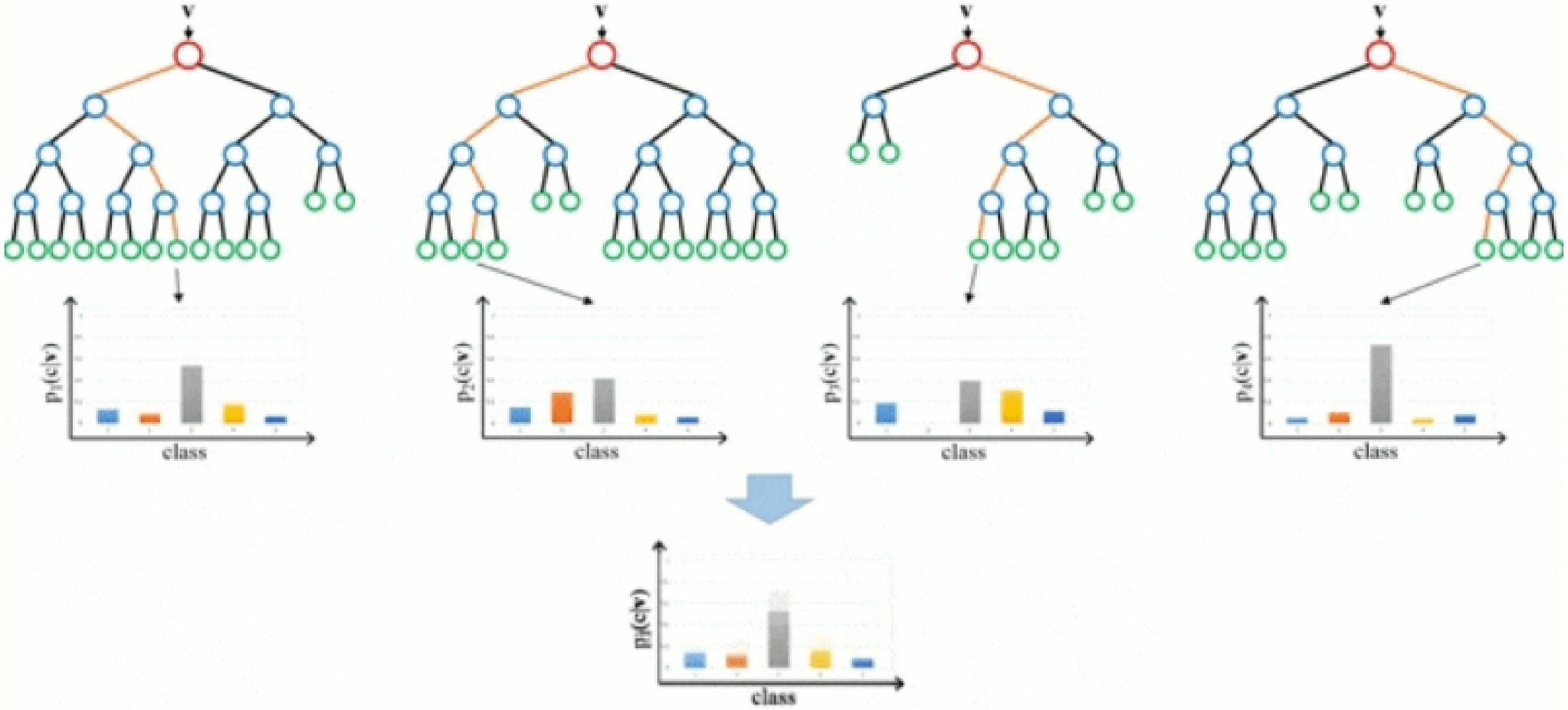
04 Light GBM

HLM Light gbm



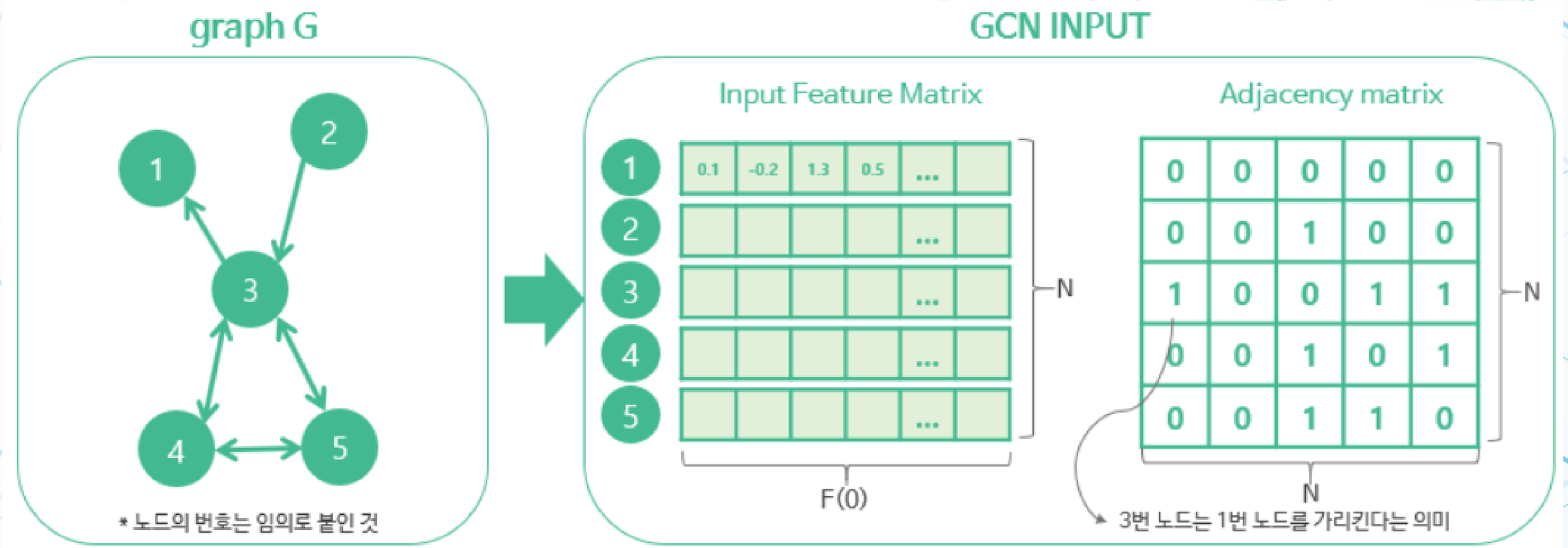


04 Random Forest



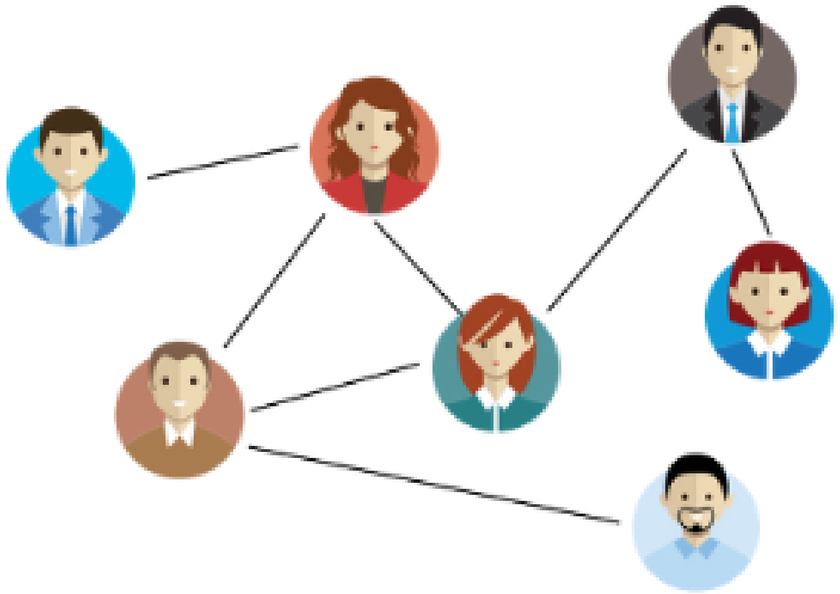
https://aytekin.tistory.com/47

05 향후 계획 Graph convolutional network (GCN)

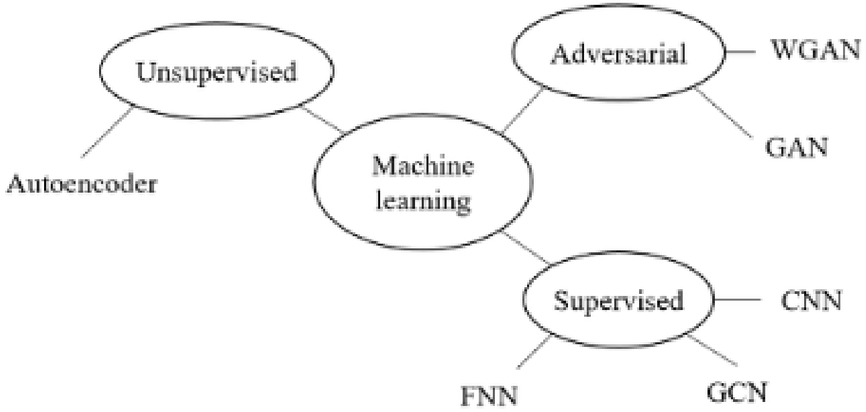


https://littlefoxdiary.tistory.com/17

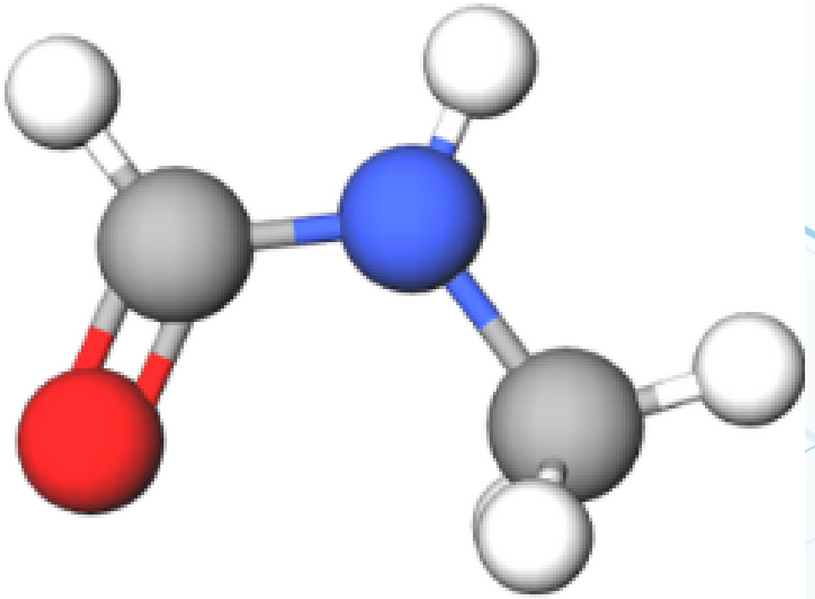
05 향후 계획 Graph convolutional network (GCN)



(a) 소셜 네트워크



(b) 관계형 데이터베이스



(c) 분자 구조

06 References

배경 https://www.freepik.com/free-vector/white-background-with-blue-tech-hexagon_4775334.htm#query=molecules&from_query=molcules&position=3&from_view=search&track=sph

아이콘 https://kr.freepik.com/free-vector/science-icons-collection_972093.htm#query=%EC%8B%A4%ED%97%98&position=3&from_view=search&track=sph

06 References

https://heeya-stupidbutstudying.tistory.com/entry/ML-GBM-%EC%95%8C%EA%B3%A0%EB%A6%AC%EC%A6%98-%EB%B0%8F-LightGBM-%EC%86%8C%EA%B0%9C-%EA%B8%B0%EB%B3%B8%EA%B5%AC%EC%A1%B0-parameters

https://bkshin.tistory.com/entry/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-4-%EA%B2%B0%EC%A0%95-%ED%8A%B8%EB%A6%ACDecision-Tree

https://untitledtblog.tistory.com/152

https://newsroom.daewoong.co.kr/archives/13497

https://news.mt.co.kr/mtview.php?no=2022092910540472976

http://m.dongascience.com/news.php?idx=59725

https://www.iworldtoday.com/news/articleView.html?idxno=403172

https://newsroom.daewoong.co.kr/archives/9082

https://www.kpbma.or.kr/sub/00000000048/00000000154

https://littlefoxdiary.tistory.com/17

https://aytekin.tistory.com/47

https://blog.naver.com/open_kbsi/221790352964

https://www.iworldtoday.com/news/articleView.html?idxno=403172

https://www.kpbma.or.kr/sub/00000000048/00000000149