

R

WEEK 5

The background of the image is a landscape at sunset or sunrise. The sky is a gradient of colors, transitioning from a deep blue at the top to a bright yellow and orange near the horizon. In the foreground, there are dark, silhouetted shapes that appear to be hills or mountains. The overall atmosphere is serene and dramatic.

FINALE

Linear Regression

선형회귀법

고민거리

In Machine Learning

설명력

Interpretability

유연성

Flexibility

설명력 유연성

뭐가 더 중요한지
정하는 것은
거의…



수지 vs 아이유 아이들에게 물어보았다

설명력

유연성

하나를 얻으려면

하나를 희생해야 된다.

설명력

유연성

다른 사람들을
이해시키기 쉽다.

다른 데이터에
바로 사용할 만큼
일반성이 높지 못하다.

설명력

유연성

일반성이 높아
다른 데이터에도
충분히 활용 가능하다.

다른 사람들에게
결과물을 설명하기
어려워진다.

선형회귀법은
설명력이 매우 뛰어나다.

**그래서 오늘은
다른걸 한다.**

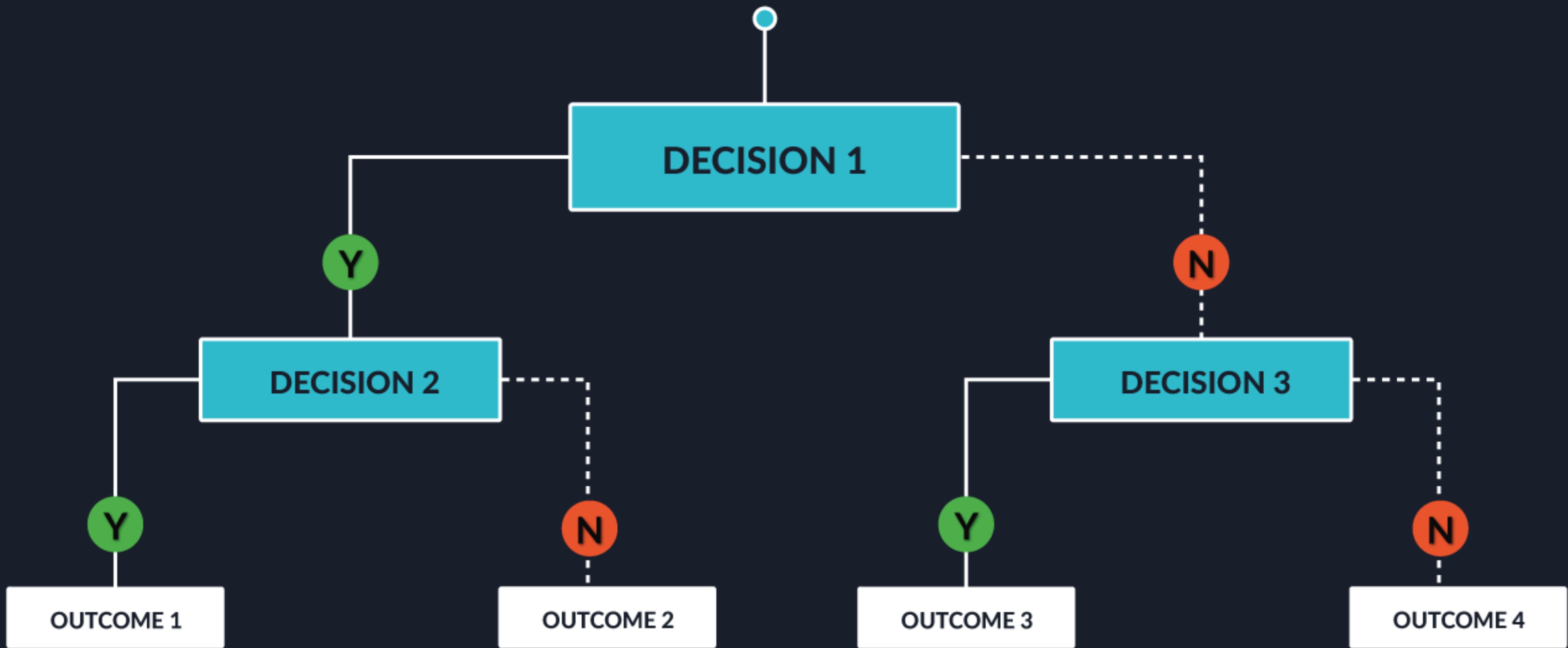
A photograph of a dense forest of tall evergreen trees, likely pines or firs. The scene is partially obscured by a thick layer of fog or mist, which is illuminated from behind by bright sunlight, creating a hazy, golden glow. The trees are dark and silhouetted against the light, with some yellow autumn leaves visible among the green.

Random Forest

숲은

나무가 우거진 곳을 말한다.

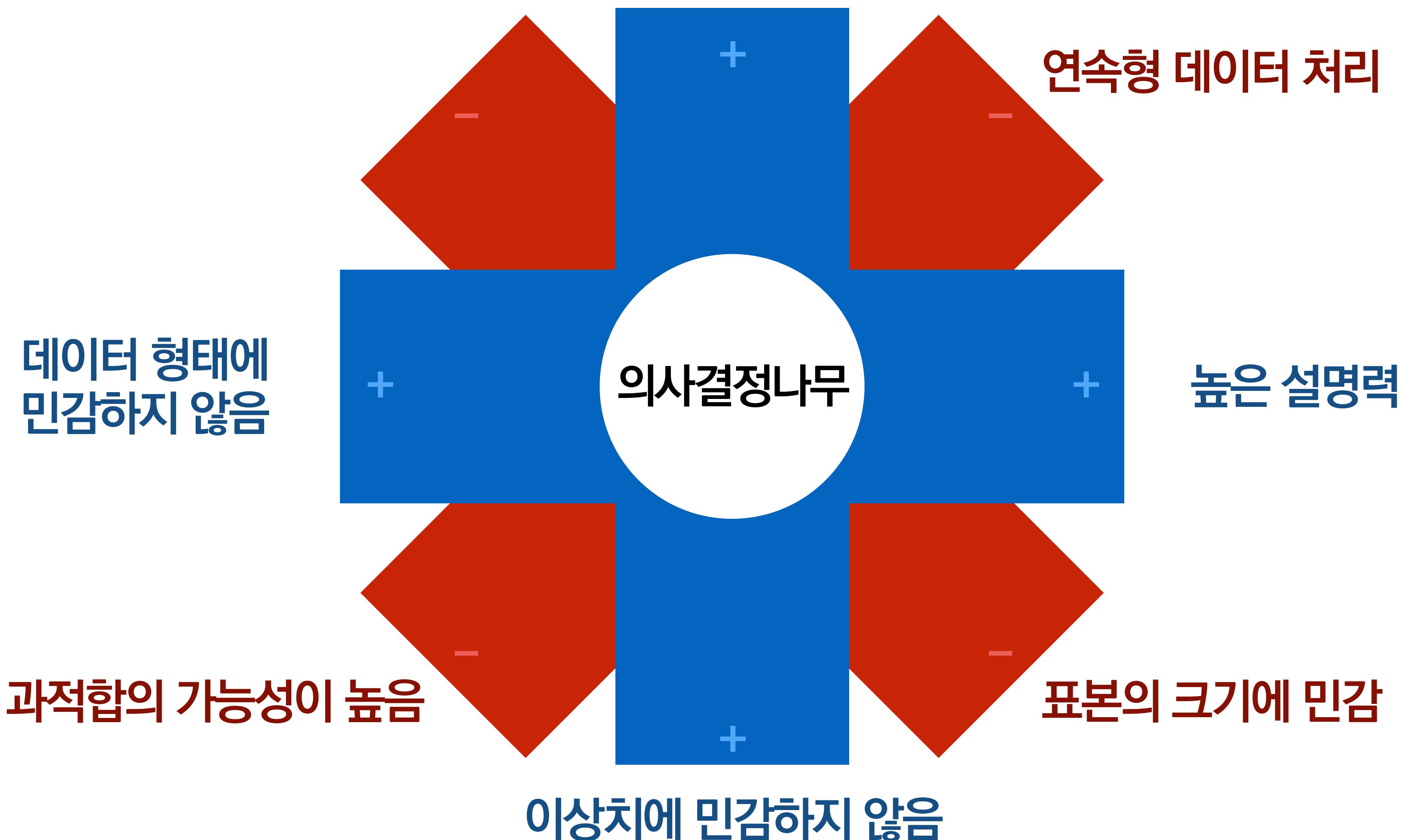
DECISION TREE



이해하기 쉽고
단순하다.

**설명력이 높다…?
그럼 유연성은…?**

모형 구축 시간이 짧음



과적합의 가능성이 높음

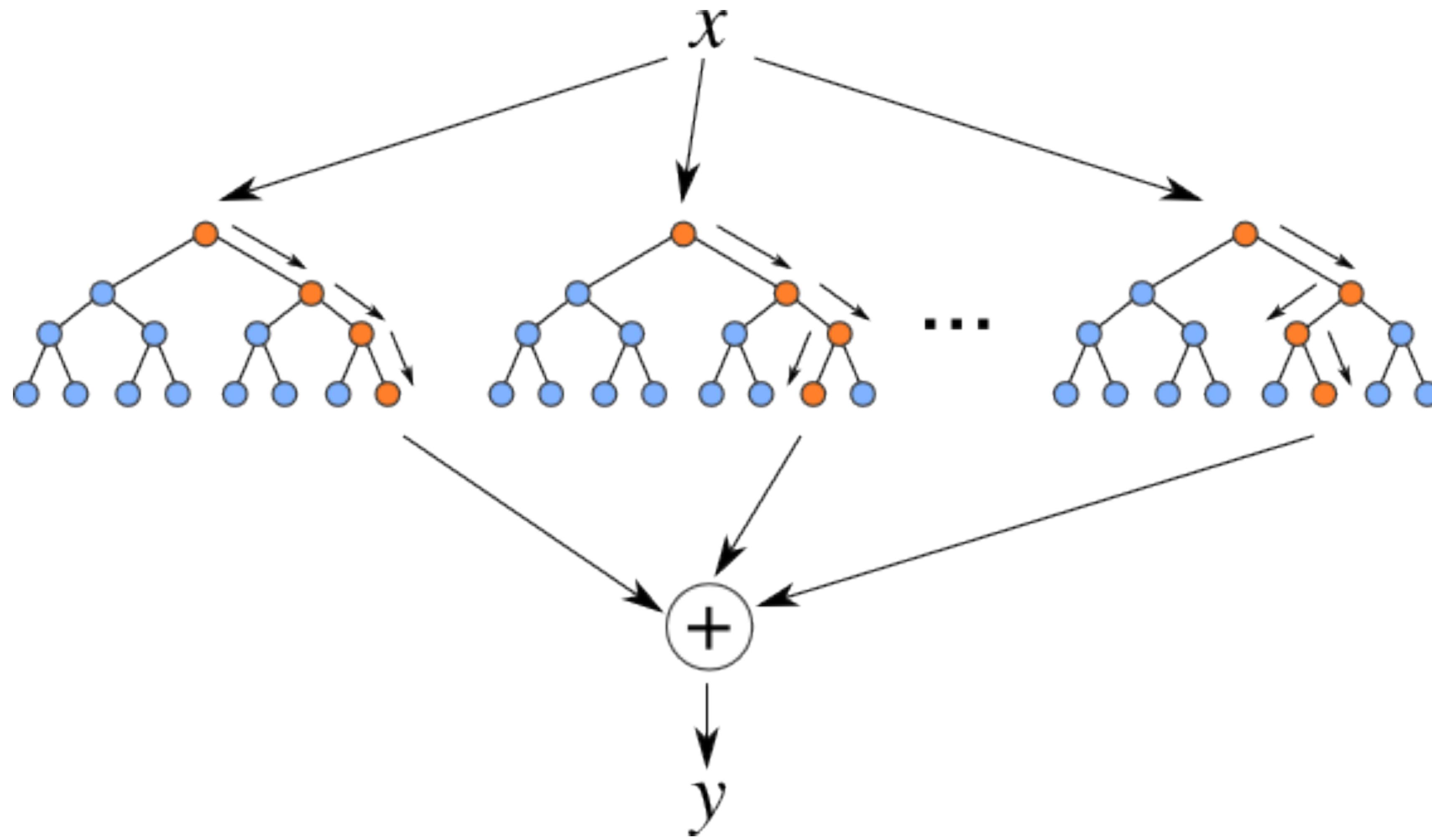
학습하는 데이터에 맞춰서

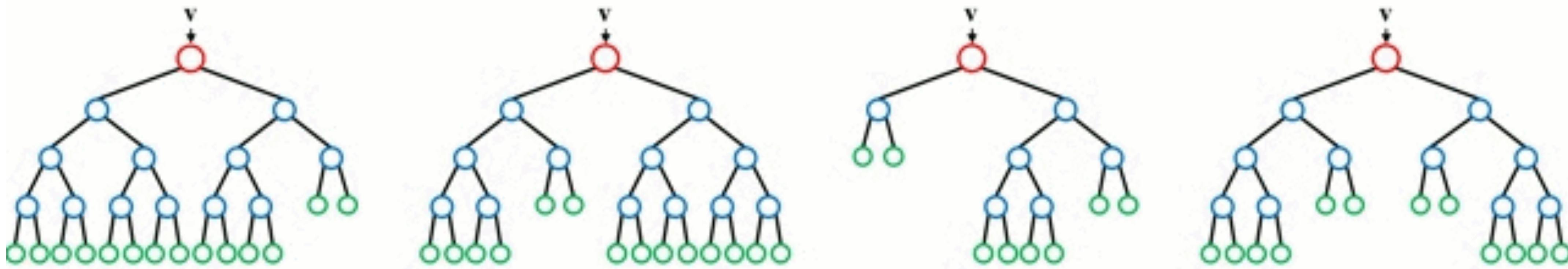
트리의 깊이를 계속해서 깊게 하고

가지를 쳐나간다면

새로운 데이터에 적용하기 어려울 수 있음

데이터를 쪼개고
여러 개의 나무를 만들어서
의사결정나무를 만들자!





<http://blog.heartcount.io/random-forest-ver-10>

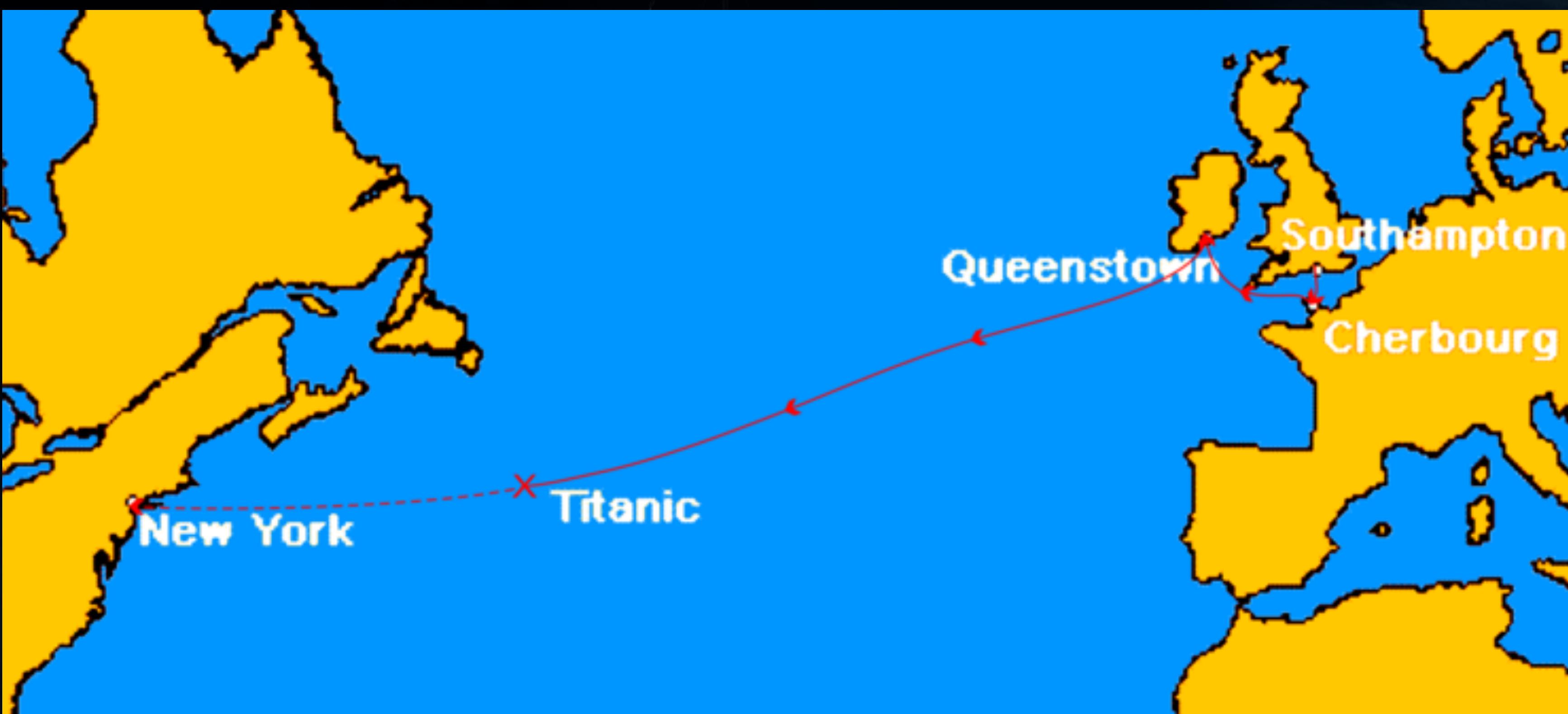
내용은 어렵지만
하고 싶은 말은 여기에 다 있다.

**랜덤 포레스트는
진짜 랜덤이다.**

오늘의 데이터는



Titanic



탑승자 종류	총원	구조율	사망율	구조인원	사망인원
1등실, 어린이	6	83%	17%	5	1
2등실, 어린이	24	100%	0%	24	0
3등실, 어린이	79	34%	66%	27	52
1등실, 여성	144	97%	3%	140	4
2등실, 여성	93	86%	14%	80	13
3등실, 여성	165	46%	54%	76	89
1등실, 남성	175	33%	67%	57	118
2등실, 남성	168	8%	92%	14	154
3등실, 남성	462	16%	84%	75	387
승무원, 여성	23	87%	13%	20	3
승무원, 남성	885	22%	78%	192	693
어린이 총합	109	51%	49%	56	53
여성 총합	425	74%	26%	316	109
남성 총합	1690	20%	80%	338	1352
전체 총합	2224	32%	68%	710	1514

PassengerID	승객 식별자
Pclass	1등석, 2등석, 3등석 정보
Name	이름
Sex	성별
Age	나이
SibSp	함께 탑승한 형제, 배우자의 수
Parch	함께 탑승한 부모, 자녀의 수
Ticket	티켓 번호
Fare	티켓 요금
Cabin	선실 번호
Embarked	탑승한 곳
Survived	생존 여부



데이터 탐색 데이터 전처리

새로운 변수 생성
결측값 다루기

HOW TO EVALUATE?

$$\text{Accuracy} = \frac{\text{올바른 예측}}{\text{전체 관측치 수}}$$

CHILD

어린이의 구조율

MANNERS MAKETH MAN

사회적 위치에 따른 구조율

FAMILY SIZE

가족 구성원의 수

**그동안
고생 많으셨습니다!**

감사합니다! :)