

# 킵스타터 프로젝트 성공여부 분류



패스트캠퍼스 신경철

# 목차

1. 킥스타터란?
2. 분석 절차
3. 분석 종류
4. 모형 검정 방법
5. 데이터 탐색
6. 최적의 모형 선택
7. 최상의 예측 정확도 도출

# 킵스타터란?

개인 또는 기업이 자신이 구현하고자 하는 상품, 혹은 서비스 등의 세부사항을 업로드하면  
그 프로젝트를 지지하는 킵스타터 회원이 후원자로 나서는 시스템

# 분석 절차

데이터 탐색

프로젝트 성공과 실패의 분포의 차이 확인

최적의  
모형 선택

다양한 분석을 적용하여 최적의 모형 선택

최상의  
예측 정확도  
도출

변수 투입 및 제거를 통해 최상의 예측 도출

# 분석 종류

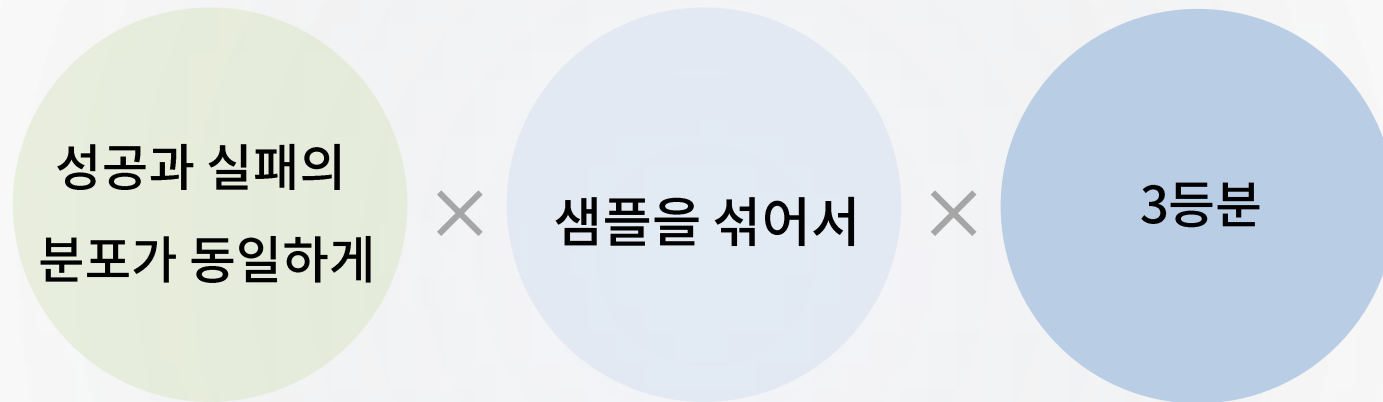
Logistic Regression

Random Forest

Perceptron

SVM

# 모형 검증 방법

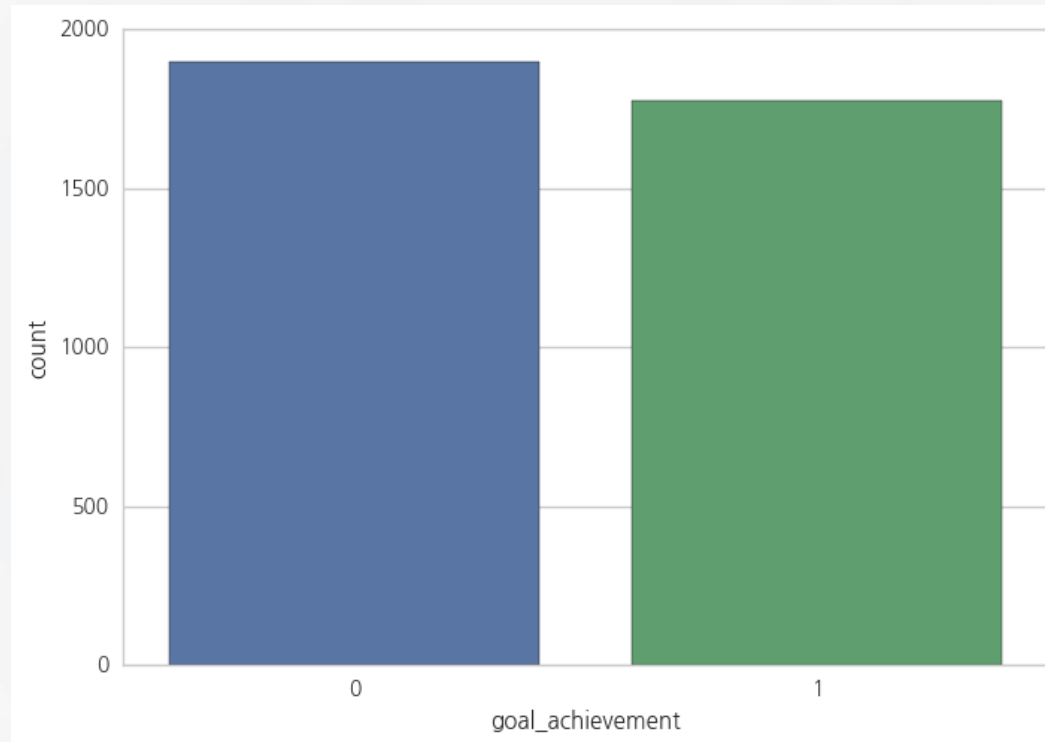


**Stratified Shuffle 3Folds**

# 데이터 탐색

	period	img	text	backed	created	goal_ex	tier_count	tier_min_ex	tier_median_ex
count	3673.000000	3673.000000	3673.000000	3673.000000	3673.000000	3.673000e+03	3673.000000	3673.000000	3673.000000
mean	35.375715	11.740811	4282.832562	7.703240	1.781105	2.590829e+04	9.087395	19.667091	118.864271
std	12.342400	14.902819	3820.912875	26.281276	2.676555	1.909164e+05	6.120009	126.547070	286.494142
min	-251.000000	0.000000	3.000000	0.000000	1.000000	7.559722e+00	1.000000	0.000000	1.000000
25%	30.000000	1.000000	1620.000000	0.000000	1.000000	2.750000e+03	5.000000	1.070205	32.500000
50%	30.000000	7.000000	3186.000000	1.000000	1.000000	7.500000e+03	8.000000	5.000000	60.000000
75%	40.000000	17.000000	5644.000000	5.000000	2.000000	2.000000e+04	12.000000	10.000000	109.000000
max	61.000000	131.000000	30945.000000	755.000000	58.000000	1.057714e+07	69.000000	5000.000000	6000.000000

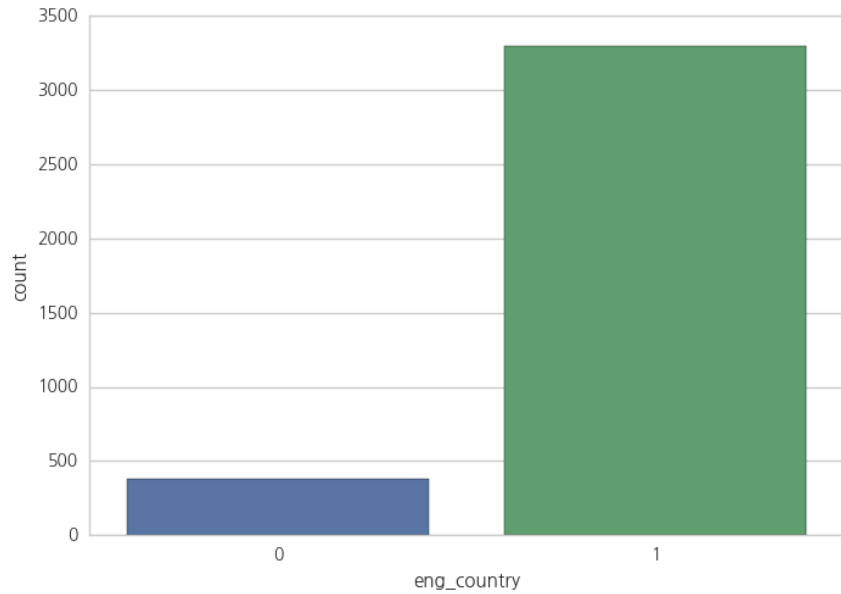
# 데이터 탐색



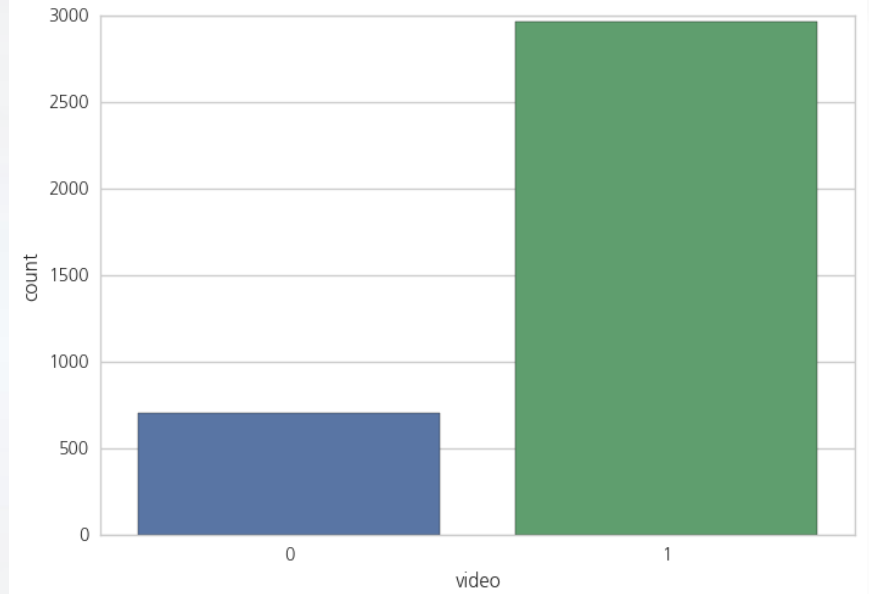
프로젝트 성공 실패 여부(타겟 변수)



# 데이터 탐색

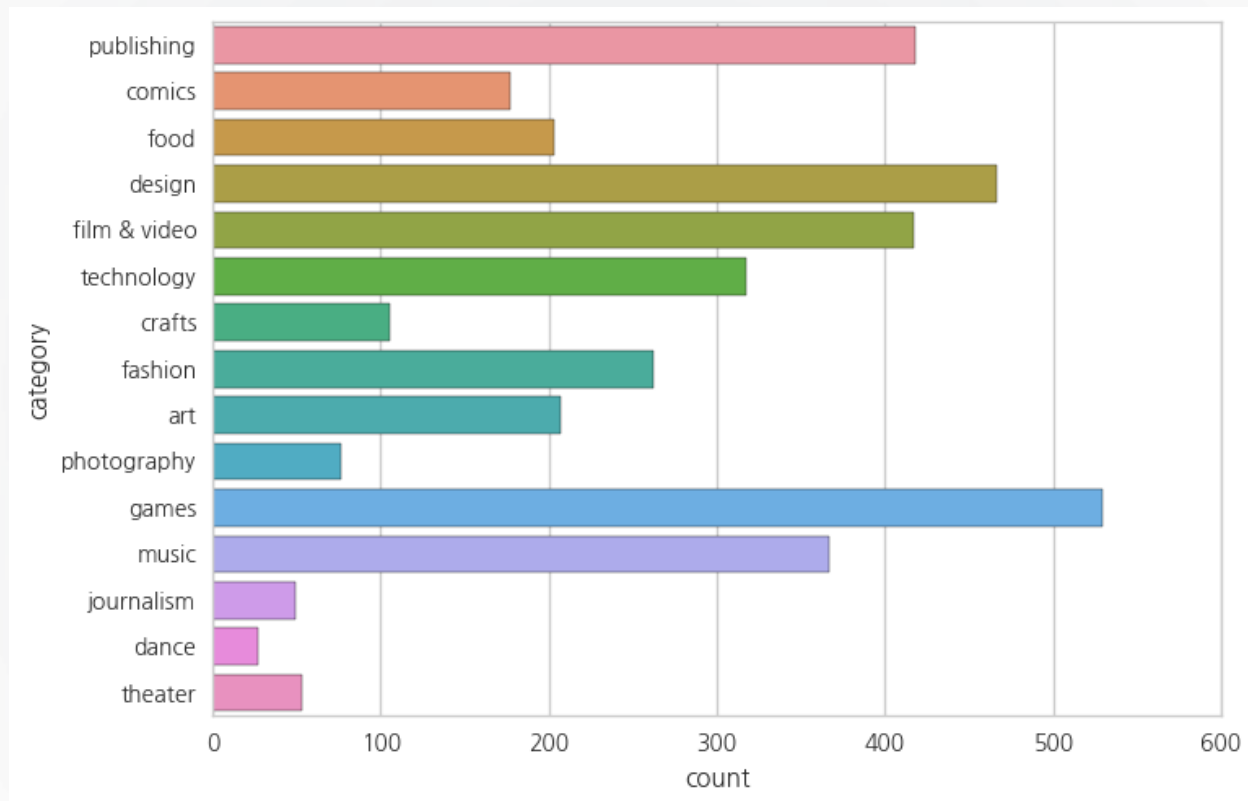


프로젝트 개설 국가 영어권 여부



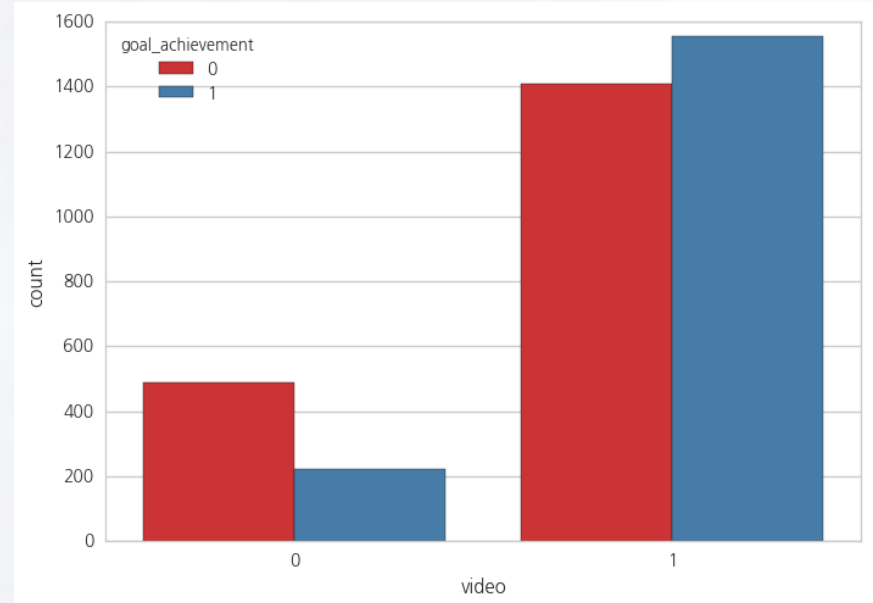
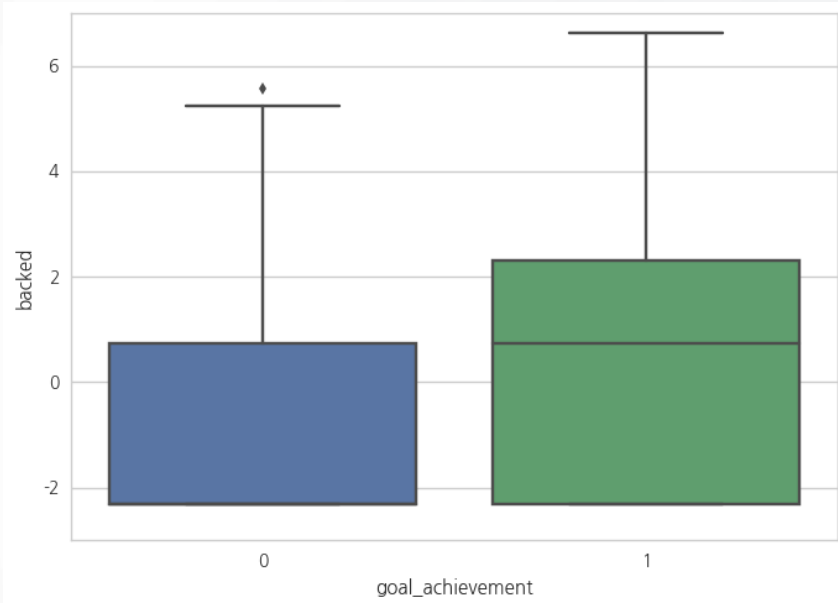
동영상 업로드 여부

# 데이터 탐색



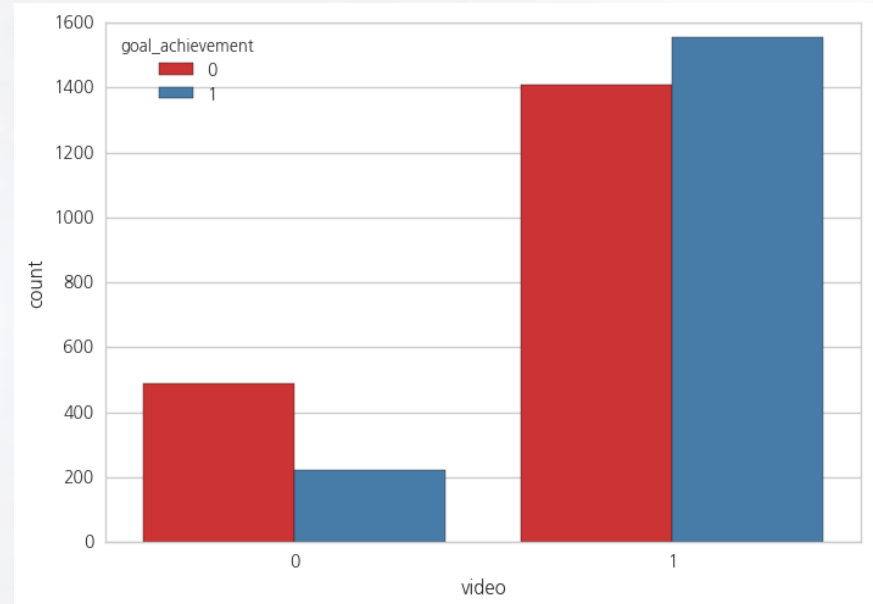
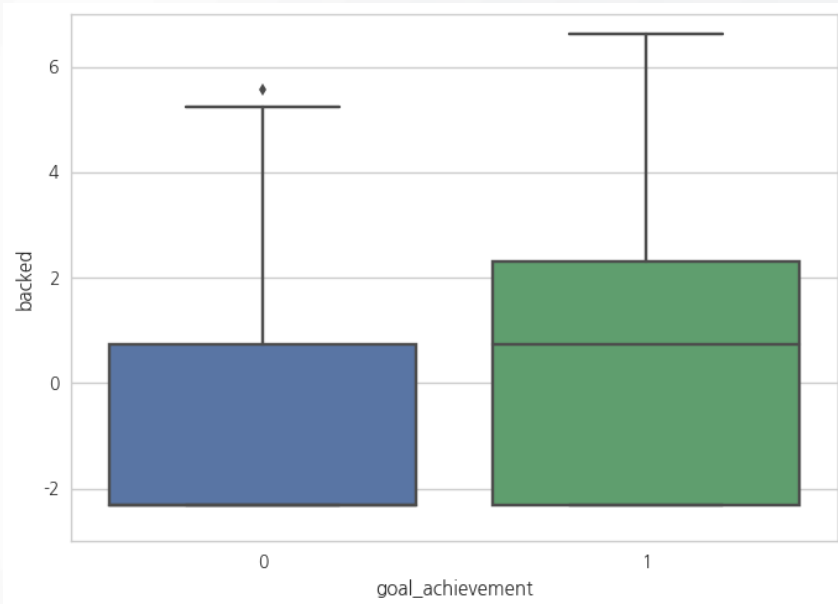
프로젝트 카테고리 변수

# 데이터 탐색



성공 실패에 따라 분포가 다른 변수들

# 데이터 탐색



성공 실패에 따라 분포가 다른 변수들

# 최적의 모형 선택

## Logistic Regression

투입 변수

'period', 'backed', 'created', 'img', 'video',  
'goal\_ex', 'tier\_count', 'eng\_country'

Accuracy

70.0

Precision

70.0

Recall

70.0

# 최적의 모형 선택

## Random Forest

투입 변수

'period', 'slug', 'backed', 'img', 'text', 'goal\_ex',  
'tier\_count', 'tier\_min\_ex', 'tier\_median\_ex'

Accuracy

66.2

Precision

66.3

Recall

66.0

# 최적의 모형 선택

## Perceptron

투입 변수

'period', 'backed', 'created',  
'goal\_ex', 'tier\_count'

Accuracy

68.0

Precision

69.0

Recall

68.3

# 최적의 모형 선택

## Support Vector Machine

투입 변수

'period', 'backed', 'created',  
'goal\_ex', 'tier\_count'

Accuracy

72.6

Precision

72.7

Recall

72.7



# 최적의 모형 선택

각 모형의 Accuracy, Precision, Recall 확인 결과,  
**Support Vector Machine**이 최적의 퍼포먼스를 보여줄 수 있음을 확인!

그리하여 다양한 변수 투입, 제거를 반복하여 분석한 결과

# 최상의 예측 정확도 도출

## Support Vector Machine

투입 변수

'period', 'slug', 'backed', 'created', 'img',  
'goal\_ex', 'tier\_count', 'tier\_median\_ex'

Accuracy

77.1

Precision

77.0

Recall

77.0