



전소미
(JYP)

528 772



최유정
(판타지오)

392 773

최저 누적
문현주
(넥스타)

16 736

PRODUCE 101 ANALYSIS & ESTIMATE

BY SAMKUGJI

230

전소미 (JYP)
215

최저 분량
이채린
MIDAS
6

52
순위상승

김주나
(뮤직케이)

2-3회
33
순위상승

응 씨 카이
(초록뱀주나)

43
순위상승

유연정
(스타쉽)

서바이벌 프로그램

'프로듀스 101'의 생존자

머신러닝으로

80% 이상 추정 (예측) 성공

WHAT?

- Estimate future
(6 weeks)



“
당신의 소녀에게
투.표.하.라!
”

월드 클래스로 자리 잡은 대한민국의 KPOP 아이돌!
이제 Mnet에서 아시아를 대표할
차세대 초대형 걸그룹이 탄생한다!

대한민국 50여 개의 엔터테인먼트에서 모인
101명의 소녀 연습생!
이들의 프로듀서는 바로 TV를 보고 있는 당신!

국민 프로듀서들이 지켜보는 가운데,
혹독한 트레이닝을 받게 될 101명의 연습생 소녀들!
유닛 걸그룹의 최종 멤버는
100% 국민 투표를 통해 결정된다!

국가대표 걸그룹이 되기 위한
치열한 연습과 잔인한 방출!

당신은 어느 소녀에게 한 표를 던질 것인가!
당신의 한 표가 소녀들의 운명을 결정한다!

WHY?

- Business Insight
- Difficult
- Interesting



HOW?

- Machine Learning
- OLS
(Ordinary Least Squares)
- See what I do
- Control what I do
- Interpret what I do

GOODNESS OF FIT

$$R^2 = \frac{ESS}{TSS} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum e_i^2 \quad TSS = ESS + RSS$$

On coefficients are chosen in such a way as to minimize the sum of the squares of the residuals. Thus it automatically follows that

OLS Regression

```

Dep. Variable:          vote_1
Model:                  OLS
Method:                 Least Squares
Date:                  Fri, 01 Apr 2016
Time:                  21:14:53
No. Observations:      97
Df Residuals:          90
Df Model:               7
Covariance Type:       nonrobust
  
```

R-
Ad
F-
Pr
Lo
AI
BI



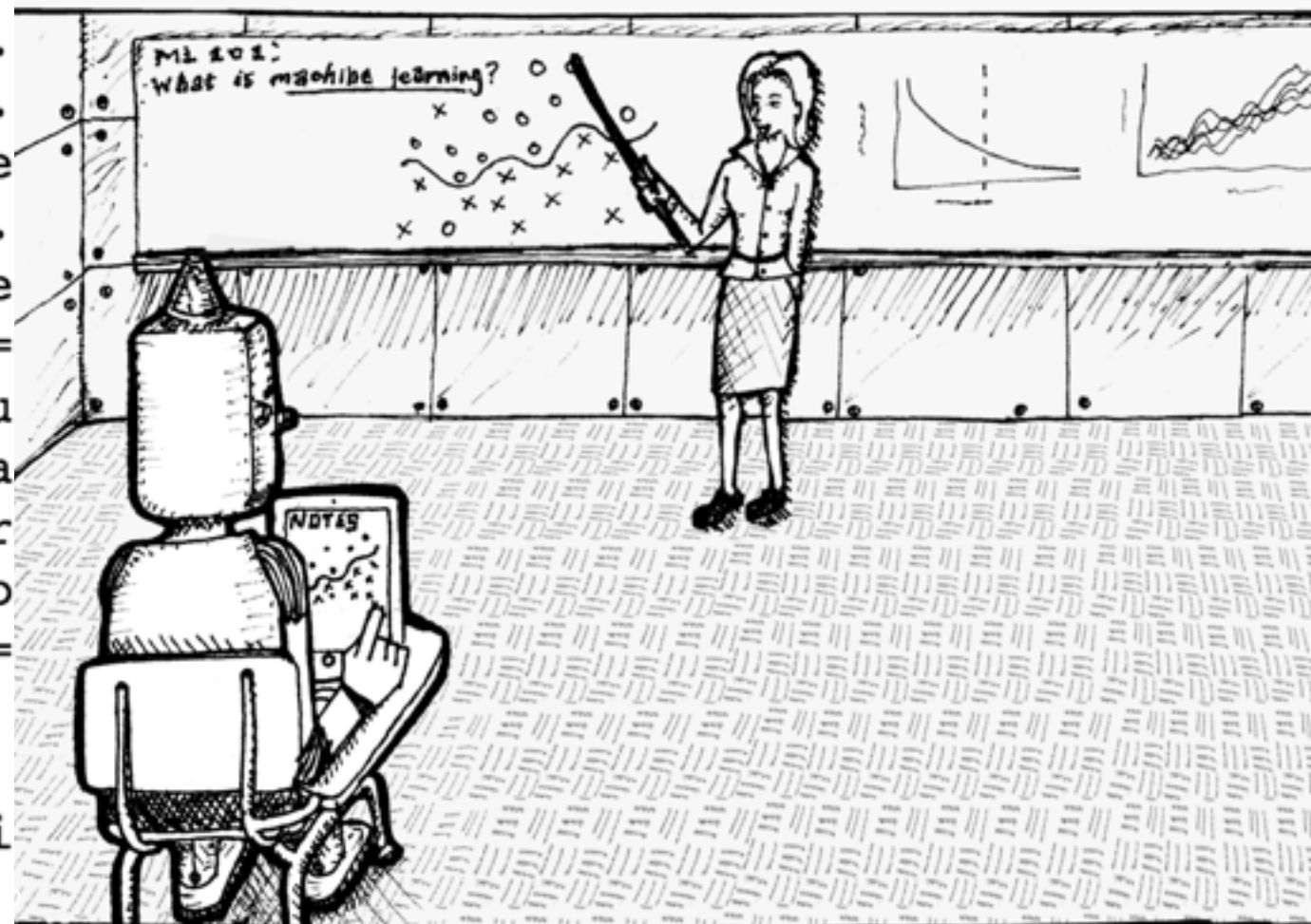
	coef	std
expo_1	2.305e+04	7144.
expo_2	3.812e+04	7212.
level_2	1.748e+04	6799.
agency_category	3.051e+04	3449.
visual	9.87e+04	2.05e
agency_member_count	-1.471e+04	6847.
is_ranker_same_agency_2	4.636e+04	1.88e

```

Omnibus:              5.245    Du
Prob(Omnibus):        0.073    Ja
Skew:                 0.139    Pr
Kurtosis:             4.298    Co
  
```

Warnings:

[1] Standard Errors assume that the covari



ENTERTAINMENT BUSINESS INSIGHT

- 내가 만약 연예기획사 대표라면? (프로듀스101 시즌2 대비)
 - 비주얼은 무시할 수 없다!
 - 소속사의 인지도와 기대치가 의외로 중요!
 - 방송노출 시간이 전부다! 첫방부터 카메라 샷을 잡아라!
 - 최대 2~3명까지만 같이 출연이 적절하다!
 - 각 연습생의 외모, 특기, 성격의 개성이 겹치지 않는 스타일을 함께 내보내라!
 - 예능의 캐릭터가 성립될 개성이 있다면 금상첨화
(스톤나영 - 김나영, 다이어터 - 강미나)
 - 실력은 아주 잘하거나 못해서 급성장 가능
(댄스A - 김청하, 보컬A - 유연정, F - 김소혜)
- 내가 만약 사업가라면?
 - 예측 11위안에 있는 기획사 주식을 구입한다! (판타지오 94% 증가, JYP 10%)

81.81%

ACCURACY FOR FINAL
IDOL IN 94 GIRLS

90.90%

ACCURACY FOR FINAL
AGENCY IN 48 AGENCIES

I'M SAMKUGJI

- <https://github.com/samkugji>
- <https://www.linkedin.com/in/samkugji>
- DS(Dae-Sung) Yu
- DS(DataScientist)
- DS(Definite Sixth-sense)

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

