

# 어떤 의안이 법률이 되는가: 데이터 불균형을 고려한 의안 결과 예측

박은정  
epark@dm.snu.ac.kr

강필성  
hong@korea.ac.kr

조성준  
zoon@snu.ac.kr

April 14, 2015

## Abstract

법의 제정은 각 개인과 여러 이익집단에 크고 작은 영향을 미친다. 이 때 개개인 또는 이익집단이 법을 수정하거나 제안하는 의안의 결과를 사전에 예측할 수 있다면, 의안의 조속한 통과 내지 폐기를 위해 적극적인 행동을 취하거나, 의안의 집행에 앞서 사전에 대비를 할 수 있다. 본 연구에서는 발의자, 의안 제안 환경 등 의안과 관련한 각종 메타 변수를 도출하여, 대한민국 국회를 대상으로 의안의 결과를 예측해본다. 사용된 알고리즘은 로지스틱 회귀분석(logistic regression), SVM, k-NN, 나이브 베이즈(Naive Bayes), 의사결정나무(decision tree) 등 다섯 가지이다. 법률안이 무조건 통과되지 않는다는 기준 모델에 대해 다섯 개의 알고리즘 모두 정확도가 4% 이상 향상 되었지만, 로지스틱 회귀분석과 SVM 등 특정 알고리즘의 경우 재현율이 낮았다. 한편 데이터의 불균형을 해소해준 후에는 이전 모델에 비해 정확도와 정밀도를 희생함으로써 재현율을 향상시킬 수 있었다.

## 1 서론

텍스트는 길고 자유롭게 입력할 수 있다. 들여쓰기나 띄어쓰기를 어떻게 해도  $\text{\LaTeX}$ 이 알아서 깔끔하게 출력해줄 것이니, 원고를 편리하게 포맷팅할 수 있다. 보통은 한 문장 혹은 한 구절을 한 줄에 입력한다.

다만 엔터키를 두 번 입력해서 빈칸을 입력하면 단락이 나뉘진다는 것은 알아두자. 텍스트를 강조하는 방법은 아주 다양하지만 주로 *이탤릭체*나 **볼드체**로 표기하는 것이 일반적이다.

‘작은 따옴표’나 “큰 따옴표”를 쓸 때는 조심해야 한다. 왼쪽과 오른쪽에 사용하는 문자가 다르기 때문이다. 누군가를 인용하거나 들여쓰기를 할 때는 quotation 환경을 쓰면 된다:

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), an interdisciplinary subfield of computer science, is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. [1]

### 1.1 이것은 절입니다

지금까지 제목과 장/절 입력, 단락 나누기, 텍스트 강조하기, 인용 등을 살펴보았다. 다음으로 목록과 표, 그림 입력에 관해 알아보자. <sup>1</sup>

<sup>1</sup>이 문서의 코드는 <http://github.com/snudm/dmtr>에서 찾으실 수 있습니다.

Places to Go Backpacking		
Name	Driving Time (hours)	Notes
Big	1.5	Very nice overnight to Berry Creek Falls from either Headquarters or oean side.
Sunol	1	Technicolor green in the spring. Watch out for the cows.
Henry	1.5	Large wilderness nearby suitable for multi-day treks.

표 1: Places to go backpacking

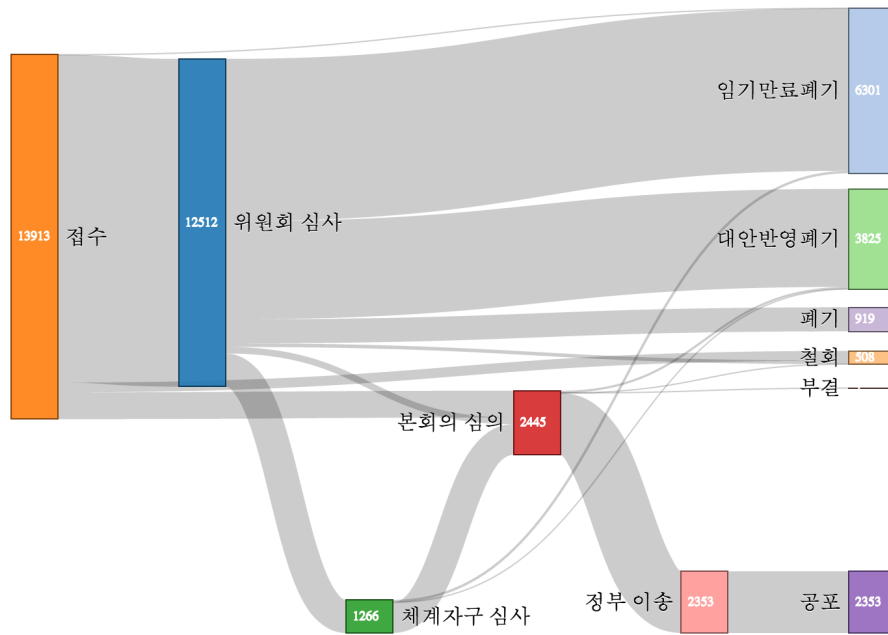


그림 1: 하하하

1. 이렇게
2. 숫자 목록을 생성하거나

- 이렇게
- 점 목록을 생성할 수 있다.

표는 아래의 표 1과 같이 입력한다.

그림은 아래의 그림 1과 같이 이미지 파일로 입력하거나 tikz를 이용해서 L<sup>A</sup>T<sub>E</sub>X으로 직접 조판할 수 있다.

## 2 수식 입력

$x^y$  또는  $x_n = \sqrt{a+b}$ 과 같이 간단한 수식은 달러 기호를 이용해서 인라인(inline)으로 바로 입력할 수 있다. 실제로 \$2000 이렇게 달러 기호를 조판하고 싶다면 \\$2000와 같이 달러 기호를 escaping해주면

된다.

더 복잡한 수식을 입력하는 경우에는 두 가지 방법이 있다. 먼저, 수식에 번호를 달지 않고 수식을 입력하는 법:

$$z \left( 1 + \sqrt{\omega_{i+1} + \zeta - \frac{x+1}{\Theta+1}y+1} \right) = 1$$

다음으로 수식에 번호를 달고 입력하는 법:

$$\left[ \mathbf{X} + \mathbf{a} \geq \hat{a} \sum_i^N \lim_{x \rightarrow k} \delta C \right] \quad (1)$$

끝으로, 참고문헌을 불러오고 문서를 닫아주는 것을 잊지 말자.

## 참고문헌

- [1] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.