

# Law Dataset Report

## Train Dataset

### Preview

class	text
i	2) 또한 자동차 열쇠에 대한 배타적인 점유를 취득하는 이상 인근에 주차된 해당 자동차를 사실상 지배하는 것은 매우 용이하고,
w	자기의 행위로 인하여 타인의 사망이라는 결과를 발생시킬 만한 가능성 또는 위험이 있음을 인식하거나 예견하면 족한 것이며 그 인식이나 예견은 확정적인 것은 물론 불확정적인 것이라도 이른바 미필적 고의로 인정되는 것인바,
u	이 사건의 경우 앞서 든 증거들에 의하여 인정되는 다음과 같은 사실들을 고려하여보면
c	나. 피고인은 본건 범행 당시 술에 취하여 심신상실 또는 심신미약의 상태에 있었다.
d	② 피해자가 피고인에게 먼저 시비를 걸며 부엌에 있던 식칼을 꺼내들고 피고인을 위협하자 피고인은 이에 대응하여 피해자를 밀어 넘어뜨리고 판시와 같이 피해자를 식칼로 찔렀는데,
i	그 과정에서 피해자가 찔린 것일 뿐,
d	고개를 숙이자 다시 등 부위를 한번 더 찔렀다고 진술하였고
i	이 사건 음료수 병에서 피고인의 지문이 나오지 않은 점 등에 비추어
i	사고(명품) 장애가 없고,
i	그리고 결국 피해자는 경부압박질식으로 사망하였다.

### Description

class	text
4078	4078
8	3832
d	살피건대,
1360	11

### Info

Non-Null Count	Null Count	Dtype
4078	0	category
4078	0	object

## Validation Dataset

### Preview

class	text
i	피해자가 입은 상해의 부위 및 정도는 사망에 이를 수 있는 정도의 위험한 수준이었고,
c	이 사건 범행 당시 피고인에게 미필적으로나마 살인의 범의가 있었음을 충분히 인정할 수 있다.
e	이러한 손상은 둔기에 의해 발생 가능한 손상으로 보인다는 소견을 밝히고 있는 점,
c	피고인 A에게 살인의 고의는 없었고,
c	피고인이 이 사건 범행 당시 음주로 인하여 사물을 변별할 능력이나 의사를 결정할 능력을 상실하였다거나 그 능력이 미약한 상태에 이르렀다고는 보이지 아니하므로,
i	피고인은 누군가로부터 피해자가 피고인을 부른다는 이야기를 듣고 피해자를 만나러 갔는데,
d	2) 이후 피고인은 C에게 수시로 문자메시지 등을 통해 욕설을 하였고,
i	엘리베이터 안에서 피해자가 완강히 저항하고
d	피해자의 코에 귀를 대보니 피해자가 숨을 안 쉬는 것을확인하고 피해자가 죽었을 것이라고 생각하여 목 조르기를 멈춘 후, 112로 전화하였다는 취지로 진술하였다.

w	그 인식이나 예견은 확정적인 것은 물론 불확정적인 것이라도 이른바 미필적 고의로 인정되는 것인바,
---	--

## Description

class	text
1935	1935
8	1852
d	위 주장은 받아들이지 아니한다.
687	6

## Info

Non-Null Count	Null Count	Dtype
1935	0	category
1935	0	object

## Test Dataset

### Preview

class	text
u	이 법원이 적법하게 채택하여 조사한 증거에 의하여 인정되는 아래와 같은 사정 즉,
u	위와 같은 사정들을 모아보면,
u	앞서 살핀 증거에 의하여 인정되는 다음과 같은 사실 또는 사정을 종합하면,
u	피고인과 변호인은,
d	(증거기록 제178쪽).
u	먼저, 피고인이 면도칼로 피해자의 왼쪽 얼굴 방향에서 목 방향으로 그은 사실이 있는지 여부를 보건대,
d	F은 당시 현장에 없었다"는 취지로 진술하였으나,
i	그 당시 피고인이나 피해자의 움직임에 따라 칼날이 조금 더 안쪽으로 들어갔다면 경동맥이 손상되어 그 자리에서 사망하였을 수도 있었다.
i	위 칼로 옆구리나 등을 찌를 경우 사망의 위험성이 크다고 보이는 점,
e	피고인에 대한 정신감정결과 및 이 법원이 채택하여 조사한 증거에 의하면 피고인이 범행 당시 술을 마신 사실,

## Description

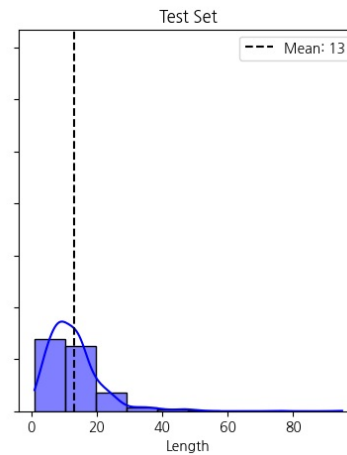
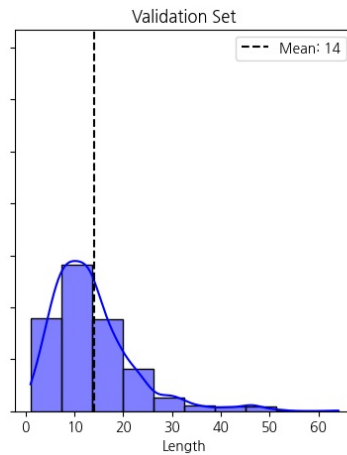
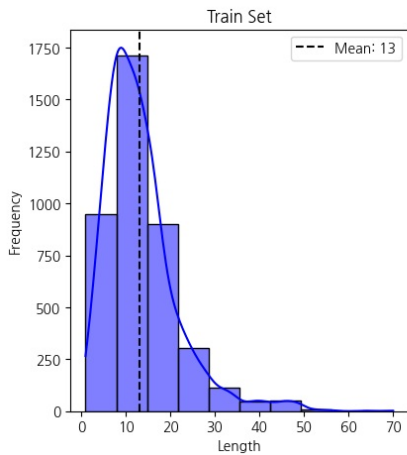
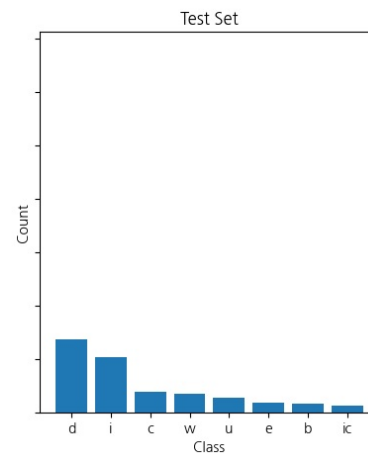
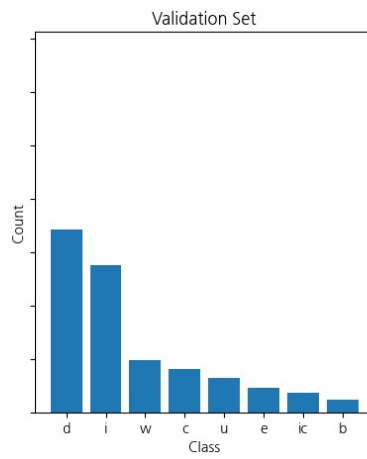
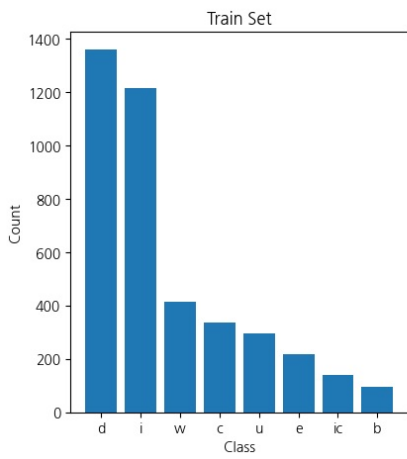
class	text
780	780
8	755
d	(대법원 2009. 2. 26. 선고 2008도9867 판결 등 참조).
274	4

## Info

Non-Null Count	Null Count	Dtype
780	0	category
780	0	object

## Visualizations

### Class Distribution



Dataset	Longest Sentence	Shortest Sentence	Mean Sentence Length	Sum of Sentences
Train	70	1	13	4078
Validation	64	1	14	1935
Test	95	1	13	780

## Word Cloud



### 통계 요약 코멘트:

- 데이터는 4,078개의 문서가 있으며 평균 문장 길이는 대략 13단어입니다.
- 주요 키워드는 피고인, 피해자, 범행, 당시, 사건으로 범죄 법정 기록에서 흔히 사용되는 용어입니다.

### 드리프트 관련 해석:

- 문서 수:
- 총 문서 수가 4078개로, 이는 데이터 세트가 상당히 크다는 것을 나타냅니다(3). 이 숫자는 분석에 충분한 데이터를 제공하며, 데

이터 분포의 변화를 감지할 수 있는 가능성을 높여줍니다.

- 평균 문장 길이:
- 평균 문장 길이가 13단어로, 일관된 형식으로 구성된 비교적 잘 구조화된 문서임을 시사합니다(4). 이는 분석이 일관되고 관련성 있는 정보를 제공할 가능성이 크다는 것을 의미하며, 데이터 드리프트를 평가하는 데 있어 신뢰성을 더욱 향상시킵니다.
- 주요 키워드:
- 주요 키워드는 피고인, 피해자, 범행, 당시, 사건으로 나타났으며, 이 단어들은 법적 문서에서 흔히 발견되는 핵심 용어입니다(5). 이는 데이터가 일관된 주제를 유지하고 있으며, 데이터의 관련성을 높이고 데이터 분포의 변화를 식별하는 데 도움을 줍니다.
- 결론적으로, 분석에 충분한 데이터를 제공하며 구조화되고 일관된 형식을 가진 크기와 주요 키워드는 법적 문서에서 데이터 분포 변화 가능성을 효과적으로 평가할 수 있게 해줍니다.