

# Law 통합 분석 리포트

데이터 드리프트 분석 보고서

생성일시: 2025년 08월 07일 15시 06분

## Dataset Information & Statistics

### 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

	Null Count	Dtype

Non-Null Count		
780	0	category
780	0	object

## 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 사정들을 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 상실하였거나 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 것을 확인하고 피해자가 죽었을 것이라고 생각하여 목 조르기를 멈춘 후, 112로 전화하였다는 취지로 진술하였다.
w	그 인식이나 예견은 확정적인 것은 물론 불확정적인 것이라도 다른바 미필적 고의로 인정되는 것인 바,

### Description

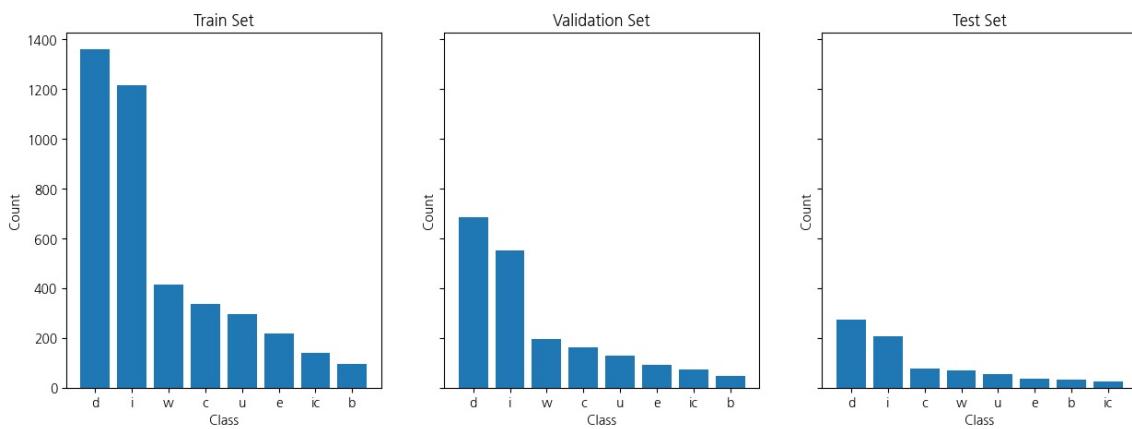
class	text
1935	1935
8	1852
d	위 주장은 받아들이지 아니한다.
687	6

### Info

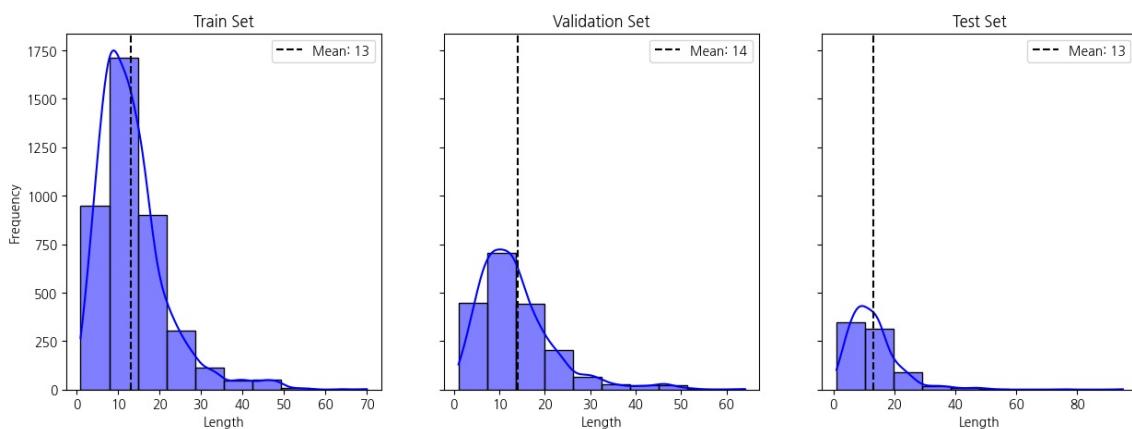
Non-Null Count	Null Count	Dtype
1935	0	category
1935	0	object

## Visualizations

### Class Distribution

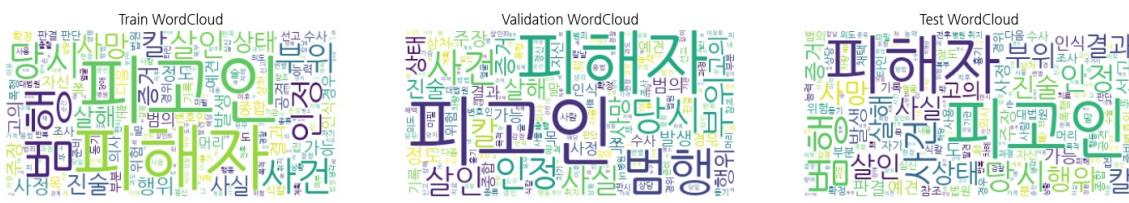


## Document Length Distribution



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

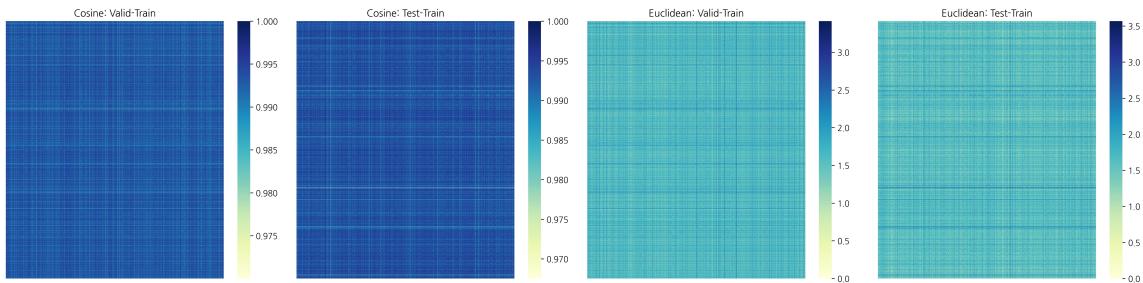


## Data Drift Analysis Results

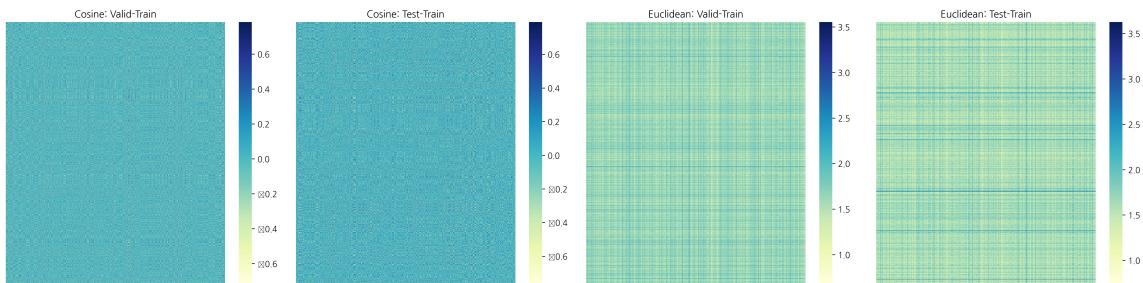
Train: (4078, 768), Valid: (1935, 768), Test: (780, 768)

**PCA Reduced Dimension:** 300.0

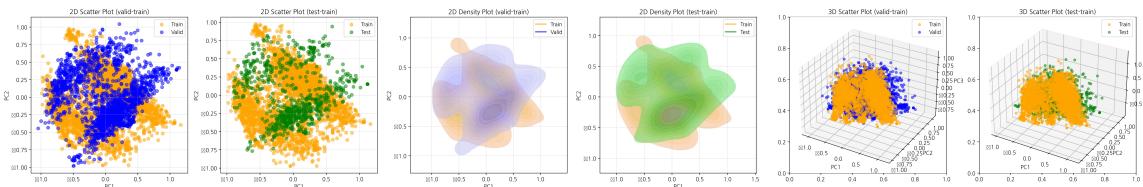
## Embedding Distance (Original Dimension)



## Embedding Distance after PCA



## Embedding Visualization after PCA



## Drift Score Summary

- MMD: score = 0.0010, drift = False
- Wasserstein Distance: score = 0.3233, drift = True
- KL Divergence: score = 0.0100, drift = False
- JensenShannon Divergence: score = 0.1533, drift = True
- Energy Distance: score = 0.0033, drift = False

## LLM Explanation

### 기술적 분석 MMD와 KL Divergence의 스코어가 낮은 값(0.0010, 0.0100)을 보이며 드리프트가 발생하지 않았다는 결과를 나타내지만, Wasserstein Distance와 JensenShannon Divergence는 상대적으로 높은 스코어(0.3233, 0.1533)를 보여 드리프트가 발생한 것으로 나타났다. 이는 데이터 분포에서 일부 차이가 존재하지만, 모든 지표가 일치하는 것은 아님을 시사한다. ### 현 상황 분석 현재 드리프트 상황은 모델의 성능에 영향을 미칠 수 있다. 특히 Wasserstein Distance와 JensenShannon Divergence의 높은 스코어는 데이터 분포의 변동이 모델의 예측 정확도에 악영향을 미칠 수 있음을 의미한다. 따라서 모델의 유지보수와 업데이트가 필요할 수 있다. ### 시각적 분석 PCA 시각화 결과를 통해 데이터 분포의 변화를 관찰할 수 있다. 임베딩 차원(768)의 고차원 데이터를 저차원으로 투영하여 시각적으로 분석하면, 드리프트로 인해 발생한 클러스터링 또는 분산의 변화를 확인할 수 있을 것이다. 이는 모델이 학습한 패턴과 실제 데이터의 패턴 사이에 거리가 있는지 여부를 직관적으로 파악하는 데 도움이 된다. ### 권장사항 즉시 조치로서, Wasserstein Distance와 JensenShannon Divergence의 높은 스코어에 주목하여 모델을 재학습하거나 업데이트하는 것을 고려해

야 한다. 또한, 지속적인 모니터링을 통해 데이터 분포의 변화를 실시간으로 추적하고, 이러한 변화가 모델의 성능에 미치는 영향을 평가할 수 있는 시스템을 구축하는 것이 필요하다. 이는 미래의 드리프트를 예방하고 모델의 안정성을 보장하는 데 중요한 역할을 할 것이다.

### DataDrift Dataclinic System

@KETI Korea Electronics Technology Institute, 2025