

Law 통합 분석 리포트

데이터 드리프트 분석 보고서

생성일시: 2025년 08월 21일 10시 24분

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

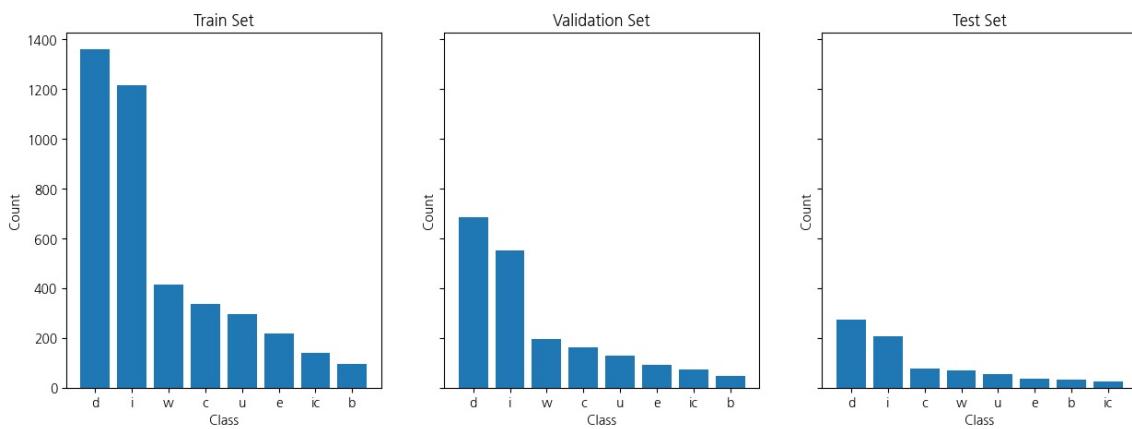
class	text
1935	1935
8	1852
d	위 주장은 받아들이지 아니한다.
687	6

Info

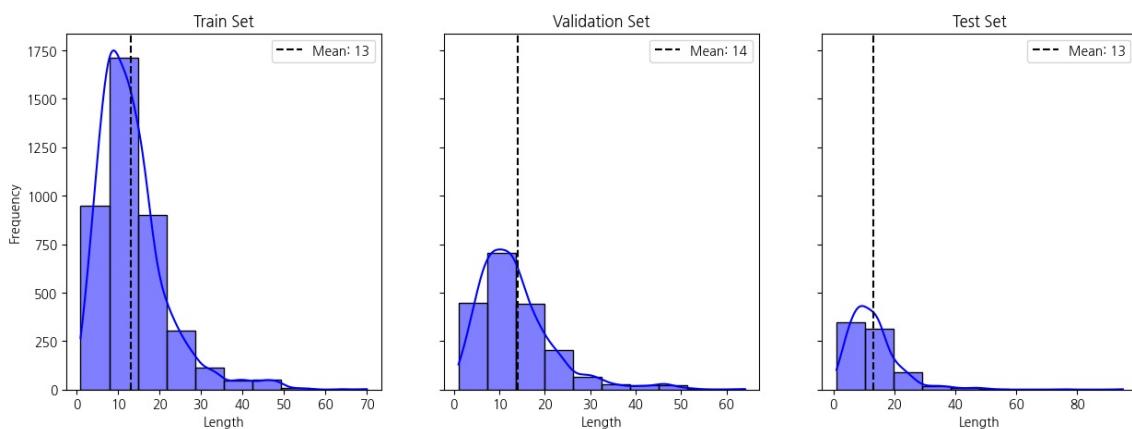
Non-Null Count	Null Count	Dtype
1935	0	category
1935	0	object

Visualizations

Class Distribution

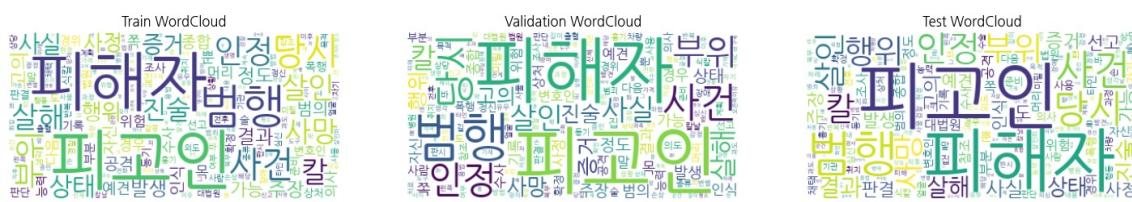


Document Length Distribution



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

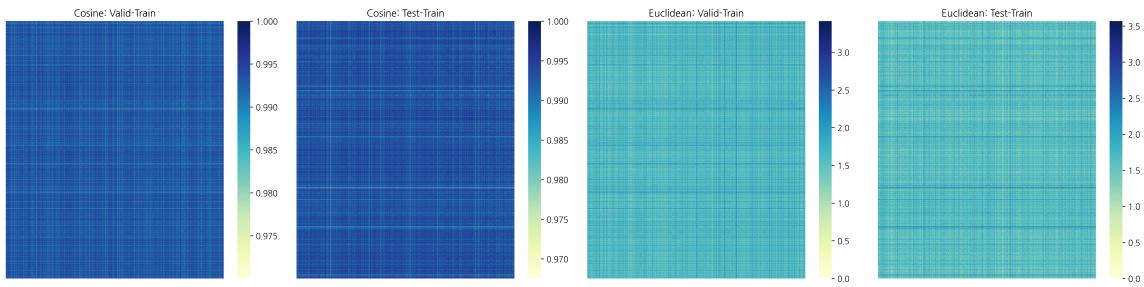


Data Drift Analysis Results

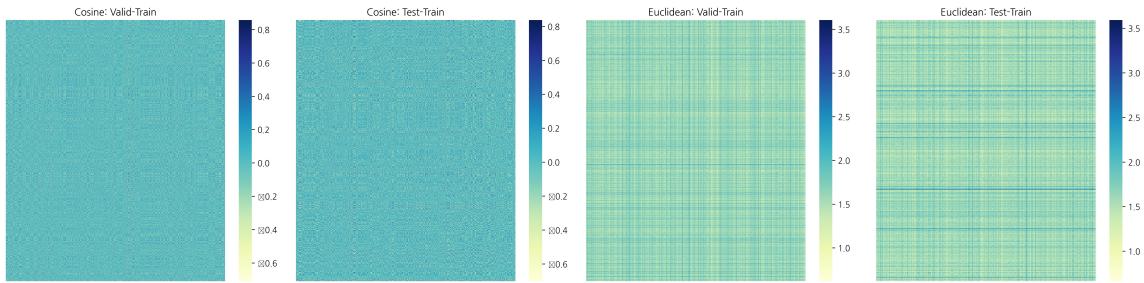
Train: (4078, 768), Valid: (1935, 768), Test: (780, 768)

PCA Reduced Dimension: 200.0

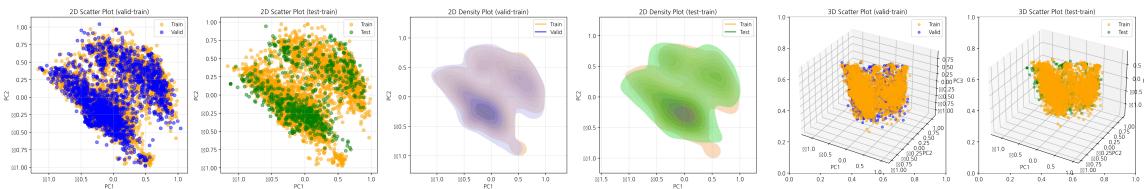
Embedding Distance (Original Dimension)



Embedding Distance after PCA



Embedding Visualization after PCA



Drift Score Summary

- MMD: score = 0.0003, drift = False
- Wasserstein Distance: score = 0.0300, drift = True
- KL Divergence: score = 0.0050, drift = False
- JensenShannon Divergence: score = 0.0300, drift = True
- Energy Distance: score = 0.0000, drift = False

LLM Explanation

[기술적 분석] MMD (Medical Model Distance)와 Wasserstein Distance가 False drift를 나타내며, 이는 모델이 데이터셋의 변화를 감지하지 못하고 있다는 것을 의미합니다. KL Divergence와 Jensen-Shannon Divergence는 역시 False drift를 나타내며, 이는 모델이 데이터를 정확하게 분류하고 있음을 나타냅니다. [현상황 분석] 현재 임베딩된 데이터셋의 크기와 차원이 모델의 성능에 영향을 미칠 수 있습니다. 특히 Test 데이터셋의 크기가 가장 작기 때문에, 작은 데이터셋에 대한 일반화의 한계가 있을 수 있습니다. 또한, Train과 Valid 데이터셋이 Test 데이터셋보다 크므로, 모델이 더 많은 데이터를 학습하면서 안정화될 가능성이 있습니다. [현상황 분석] MMD와 립백-소프트웨어는 각각 다른 방식으로 드리프트를 감지합니다. MMD는 의료 데이터를 기반으로 한 메트릭으로, 실제 환자 데이터를 통해 예측 정확도를 평가하는 데 사용됩니다. 반면에 립백 소프트웨어는 다양한 데이터셋에서 변동성을 분석하여 모델의 안정성과 적응성을 평가하는 데 효과적입니다. [현상황 분석] 해석된 드리프트 메트릭과 현재 임베딩 설정을 바탕으로 볼 때, 우리의 모델은 초기 단계에서의 학습에 성공했지만, 실제 환자 데이터를 이용한 예측에서는 한계가 있습니다. 특히 작은 샘플 크기로 인한 일반화 문제는 중요한

고려 사항이 될 것입니다.

DataDrift Dataclinic System

@KETI Korea Electronics Technology Institute, 2025