Fusión Semántico-Espacio-Temporal de Datos de Redes Sociales Y Gubernamentales para el Análisis de Crimen

Leon Felipe Davis Coropuna

Problema

Crimes - 2001 to Present

Actualmente, los enfoques existentes para el análisis de datos criminales enfrentan limitaciones debido a la fragmentación y heterogeneidad de la información proveniente de diversas fuentes. Esta dispersión genera registros superpuestos, incompletos o inconsistentes, lo que dificulta consolidar una visión unificada de fenómenos delictivos.

Objetivo

Desarrollar un enfoque para el análisis de datos criminales urbanos que integre de manera conjunta las dimensiones semántica, espacial y temporal, a partir de la fusión de información heterogénea proveniente de redes sociales y fuentes oficiales, con el fin de generar una visión unificada y coherente que contribuya a la toma de decisiones en seguridad pública.

Datasets

Crimes - 2001 to Present

Este dataset proporciona un registro de incidentes delictivos en la ciudad de Chicago desde el año 2001 hasta la actualidad. Los datos son proporcionados por el sistema CLEAR (Citizen Law Enforcement Analysis and Reporting) del Departamento de Policía de Chicago, y son actualizados diariamente. En total, el conjunto de datos contiene más de 8.3 millones de registros y 22 columnas, incluyendo datos espaciales. Como muestra se tomó datos desde 2020.

CHICAGO DATA PORTAL

Link dataset

Datasets

Crimes - 2001 to Present

Se generarán dos tipos de embeddings: espacio-temporales y textuales. Los embeddings espacio-temporales se construyeron a partir de variables derivadas de fecha y ubicación, transformadas mediante técnicas de codificación y normalización.

```
24  X_scaled
25  Y_scaled
26  Cluster
27  Rot30_X
28  Rot30_Y
29  Rot45_X
30  Rot45_Y
31  Rot60_X
32  Rot60_Y
33  Month
34  dayOfWeek
35  dayOfMonth
36  dayOfYear
37  weekOfMonth
38  weekOfYear
39  Hour
40  Minute
41  Hour_Zone
42  BusinessHour
43  Weekend
44  Season
45  Holiday
```

Datasets

Crimes - 2001 to Present

En cuanto a los embeddings textuales, estos se obtuvieron a partir de las descripciones y otras variables de tipo texto presentes en el dataset, como: Location Description (descripción del lugar), Category (tipo de crimen, se toman las mismas categorías que en [1][2]), etc.

Class	# Articles	Proportion
Gambling	249	2.91%
Murder	2,557	29.85%
Sexual Abuse	673	7.86%
Theft/Burglary	774	9.03%
Drug	1,039	12.13%
Battery/Assault	1,889	22.05%
Accident	721	8.42%
Non-Crime	1,406	16.41%
All	8,567	100.00%

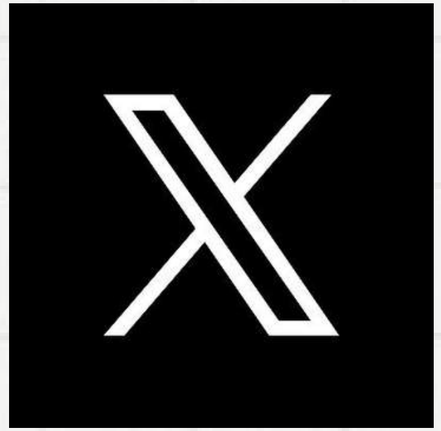
Table 2 Statistics of the annotated news articles for model evaluation

Class	# Samples	Proportion
Gambling	249	2.91%
Murder	2,557	29.85%
Sexual abuse	673	7.86%
Theft/Burglary	774	9.03%
Drug	1,039	12.13%
Battery/Assault	1,889	22.05%
Accident	721	8.42%
Non-Crime/Accident	1,406	16.41%

Datasets

Tweets entre enero y abril de 2020

El dataset fue obtenido mediante técnicas de scraping en la plataforma X.com (antes Twitter), utilizando palabras clave relacionadas con crímenes. Como resultado, se recolectaron aproximadamente 101630 registros, cada uno representando un tweet individual, junto con sus respectivos metadatos. La información fue almacenada inicialmente en archivos JSON, organizados cronológicamente. Posteriormente, estos archivos fueron unificados y convertidos en un único archivo CSV con 106 atributos por registro.



x.com

Datasets

Tweets entre enero y abril de 2020

En este conjunto de datos se generan embeddings tanto espacio-temporales como textuales. Para los embeddings textuales se empleó el modelo XLM-RoBERTa en tareas de clasificación y reconocimiento de entidades nombradas (NER), con el objetivo de extraer metadatos relevantes. Por su parte, los embeddings espacio-temporales se derivaron de información de ubicación y fechas, también procesada mediante XLM-RoBERTa, dada su eficacia demostrada en [1][2].

Label
Criminal
Victim
Police
Date/Time
Location
Item
Action
Worth
Root Cause
Trigger

Datasets

Tweets entre enero y abril de 2020

"Last night, John Doe shot Maria Lopez near 5th Ave. Stole her purse and fled in a black car."


```
{
 "classification": ["Theft/Burglary", "Battery/Assault"],
 "metadata": {
 "Criminal": ["John Doe"],  "Victim": ["Maria Lopez"],  "Police": [],
 "Date/Time": ["Last night"],  "Location": ["near 5th Ave"],
  "Item": ["her purse"],   "Action": ["shot", "Stole her purse and fled in a
black car"],  "Worth": [],  "Root Cause": [],  "Trigger": []  }
}
```

Trabajos Relacionados

Dataset

Estructura del dataset:

- title: título del artículo.
- introduction: introducción del artículo.
- body_text: contenido completo del artículo.
- labels: lista binaria o multiclase con 0/1 para cada tipo de crimen (multi-label).

	News Title	News_Intro	News_Desc	News_All	Gambling	Murder	Sexual Abuse	Theft/Burglary	Drug	Battery/Assault	Ac
0	ตร.นนท์จับหนุ่มแท็กซี่ขโมยเนื้อหมูของร้านตามส้...	ตำรวจ สภ.รัตนาธิเบศร์จับกุมโชเฟอร์แท็กซี่ตามห...	ตำรวจ สภ.รัตนาธิเบศร์จับกุมโชเฟอร์แท็กซี่ตามห...	ตร.นนท์จับหนุ่มแท็กซี่ขโมยเนื้อหมูของร้านตามส้...	0	0	0	1	0	0	

Dataset

Se definieron diez categorías de etiquetas que permiten estructurar la información relevante: Criminal, que identifica a la persona que cometió el crimen; Victim, la persona afectada; Police, oficiales involucrados en el caso; DateTime, la fecha y hora del evento; Location, el lugar donde ocurrió el crimen; Item, objetos relevantes como armas o drogas; Action, la descripción del acto criminal; Worth, el daño económico o físico provocado; RootCause, que representa causas subyacentes como pobreza o adicción; y Trigger, los detonantes inmediatos como celos o provocación.

```
{
  "tokens": ["El", "sospechoso", "Juan", "Pérez", "disparó", "a", "María", "López", "."],
  "ner_tags": ["O", "O", "B-Criminal", "I-Criminal", "B-Action", "O", "B-Victim", "I-Victim", "O"]
}
```

CAMELON: A System for Crime Metadata Extraction and Spatiotemporal Visualization from Online News Articles

Ref [1]

CAMELON

Contexto:

En los países en desarrollo, el crimen sigue siendo un problema grave, agravado por el crecimiento urbano acelerado. El monitoreo oportuno y detallado de los crímenes podría ayudar tanto a la policía como a los responsables de políticas públicas a actuar proactivamente. Si bien las noticias en línea se han convertido en una fuente útil y actualizada de información sobre crímenes, los sistemas actuales que utilizan estas noticias solo se enfocan en clasificar tipos de crímenes y ubicarlos en mapas, sin proporcionar suficiente detalle para una toma de decisiones informada.

CAMELON

Problema:

Los sistemas existentes de monitoreo de crímenes que se basan en noticias en línea:

- Se limitan a clasificar el tipo de crimen y visualizar la ubicación.
- No extraen metadatos relevantes como víctimas, criminales, armas, causas o acciones.
- No permiten un análisis profundo de patrones delictivos ni evaluación de severidad regional.
- No están diseñados para ser adaptables a diferentes idiomas o países.

CAMELON

Objetivo:

Desarrollar CAMELON, un sistema inteligente para:

- Recoger automáticamente artículos de noticias.
- Clasificarlos en tipos de crímenes más detallados (como asesinato, drogas, robo, etc.).
- Extraer metadatos relevantes del contenido (como víctima, criminal, causa, acción, valor de daño).
- Visualizar la información espacial y temporalmente.

CAMELON

Metodología:

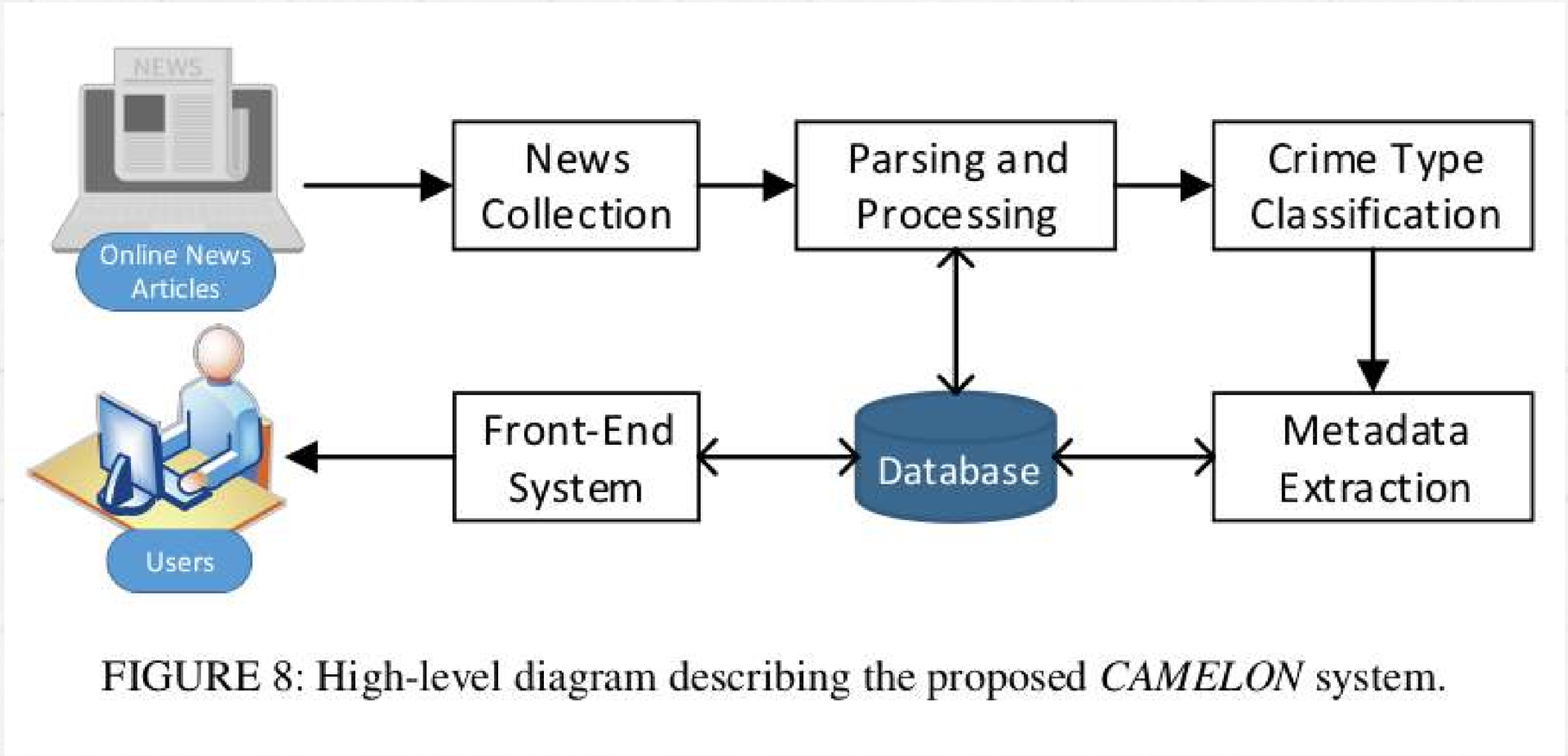


FIGURE 8: High-level diagram describing the proposed *CAMELON* system.

CAMELON

Metodología:

1- Recolección de datos:

- Se recopilan automáticamente noticias de medios en línea mediante un parser HTML específico para cada sitio web.
- Se extraen metadatos básicos: fecha/hora de publicación, título, introducción y cuerpo del texto.

CAMELON

Metodología:

2- Clasificación automática de tipos de crimen:

- Las noticias son clasificadas en siete categorías: apuestas, asesinato, abuso sexual, robo/asalto a viviendas, drogas, agresión física, y accidentes.
- La clasificación se aborda como una tarea de clasificación multietiqueta usando modelos de deep learning:
- Modelos utilizados: BiLSTM, WangchanBERTa, Multilingual BERT (MBERT) y XLM-RoBERTa (XLMR).
- El modelo XLMR obtuvo los mejores resultados.

CAMELON

Metodología:

3- Extracción de metadatos del crimen:

- Tarea formulada como un problema de reconocimiento de entidades nombradas (NER).
- Se extraen atributos como: criminal, víctima, policía, fecha/hora, ubicación, evidencia, acción, daño, causa raíz y detonante.
- Se incluye la extracción de entidades estándar (persona, lugar, fecha) y también de atributos en texto libre (acción, causa, motivación).
- Anotaciones realizadas manualmente con Doccano.
    a. Modelos probados: CRF, BiLSTM-CRF, WangchanBERTa, y XLM-RoBERTa.

CAMELON

Metodología:

4- Almacenamiento en base de datos:

- Se usa MySQL para almacenar los datos procesados.
- Incluye representaciones espaciales (coordenadas geográficas obtenidas mediante Google Geocoding API), temporales y semánticas (metadatos y tipo de crimen).

CAMELON

Metodología:

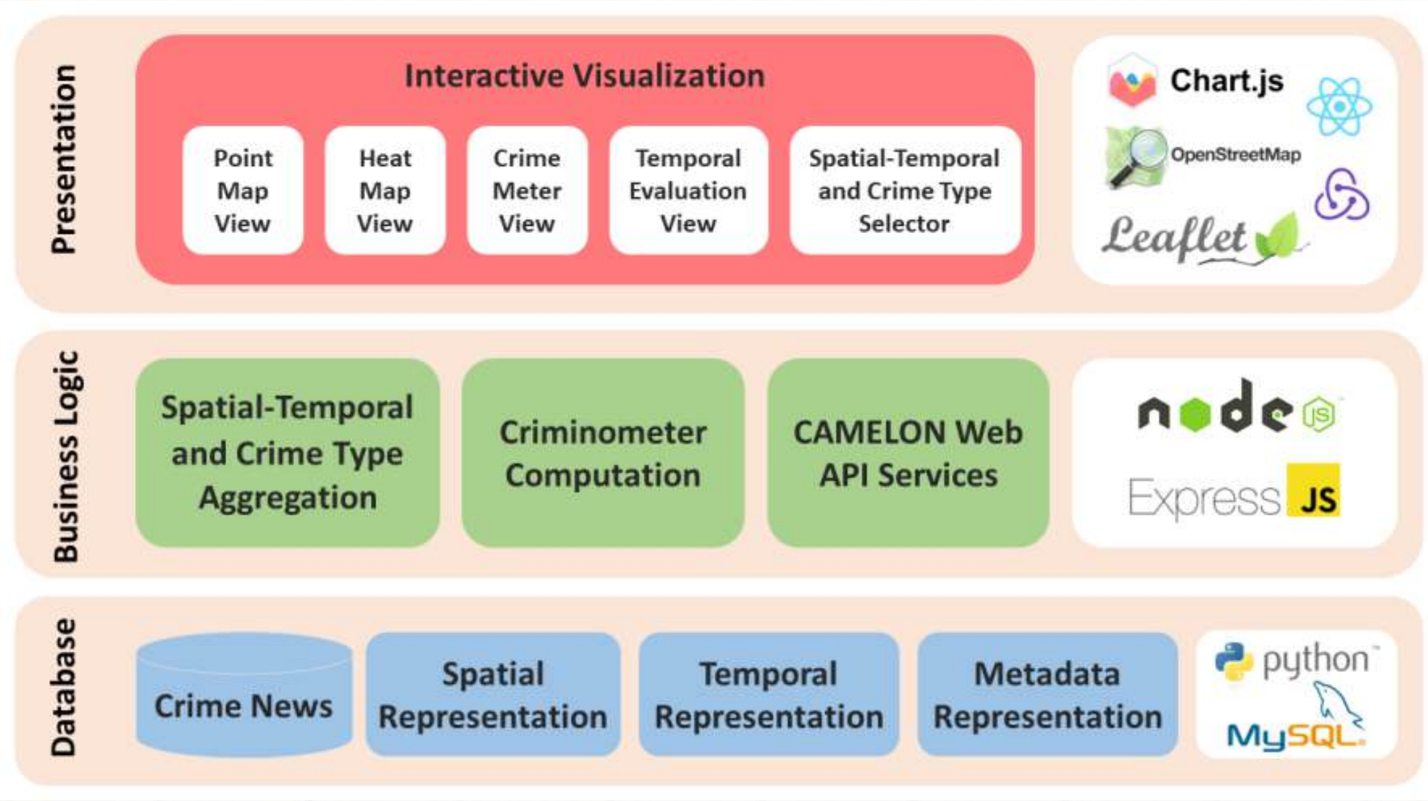
5- Tareas del sistema

- Tarea #1 - Identificar patrones delictivos a nivel nacional, regional y provincial.
- Tarea #2 - Selección interactiva de aspectos espaciales, temporales y del tipo de crimen de interés.
- Tarea #3 - Visualización de los metadatos de crimen extraídos por el extractor de metadatos
- Tarea #4 - Evaluación del crimen mediante el índice Criminómetro a nivel provincial.

CAMELON

Metodología:

6- Visualización web interactiva



CAMELON

Resultados:

Estadísticas de los artículos de noticias anotados para la tarea de clasificación por tipo de crimen. Cabe señalar que un artículo puede estar anotado con múltiples tipos de crimen.

Class	# Articles	Proportion
Gambling	249	2.91%
Murder	2,557	29.85%
Sexual Abuse	673	7.86%
Theft/Burglary	774	9.03%
Drug	1,039	12.13%
Battery/Assault	1,889	22.05%
Accident	721	8.42%
Non-Crime	1,406	16.41%
All	8,567	100.00%

CAMELON

Resultados:

Comparación del rendimiento de clasificación en la tarea de clasificación por tipo de crimen en términos de F1.

Classifier	Gambling	Murder	Sexual Abuse	Theft	Drug	Battery/ Assault	Accident	Non-Crime	Average
BiLSTM	0.707	0.833	0.711	0.611	0.750	0.606	0.724	0.755	0.712
WangchanBERTa	***0.888***	0.905	0.889	0.778	0.907	0.720	0.845	0.809	0.843
MBERT	0.008	0.854	0.753	0.658	0.849	0.641	0.733	0.789	0.661
XLMR	0.887	***0.917***	***0.904***	***0.818***	***0.916***	***0.753***	***0.846***	***0.839***	***0.860***

CAMELON

Resultados:

Estadísticas de las entidades de metadatos de crimen anotadas.

Label	# Entities	# Tokens	Avg. Entity Length
Criminal	3,775	37,958	11.47
Victim	2,697	27,003	11.20
Police	3,894	31,784	8.38
Date/Time	2,062	13,634	6.74
Location	2,858	32,821	12.94
Item	3,211	31,329	10.38
Action	4,467	26,271	6.26
Worth	3,140	17,977	6.08
Root Cause	793	4,693	6.00
Trigger	1,357	7,982	5.94

CAMELON

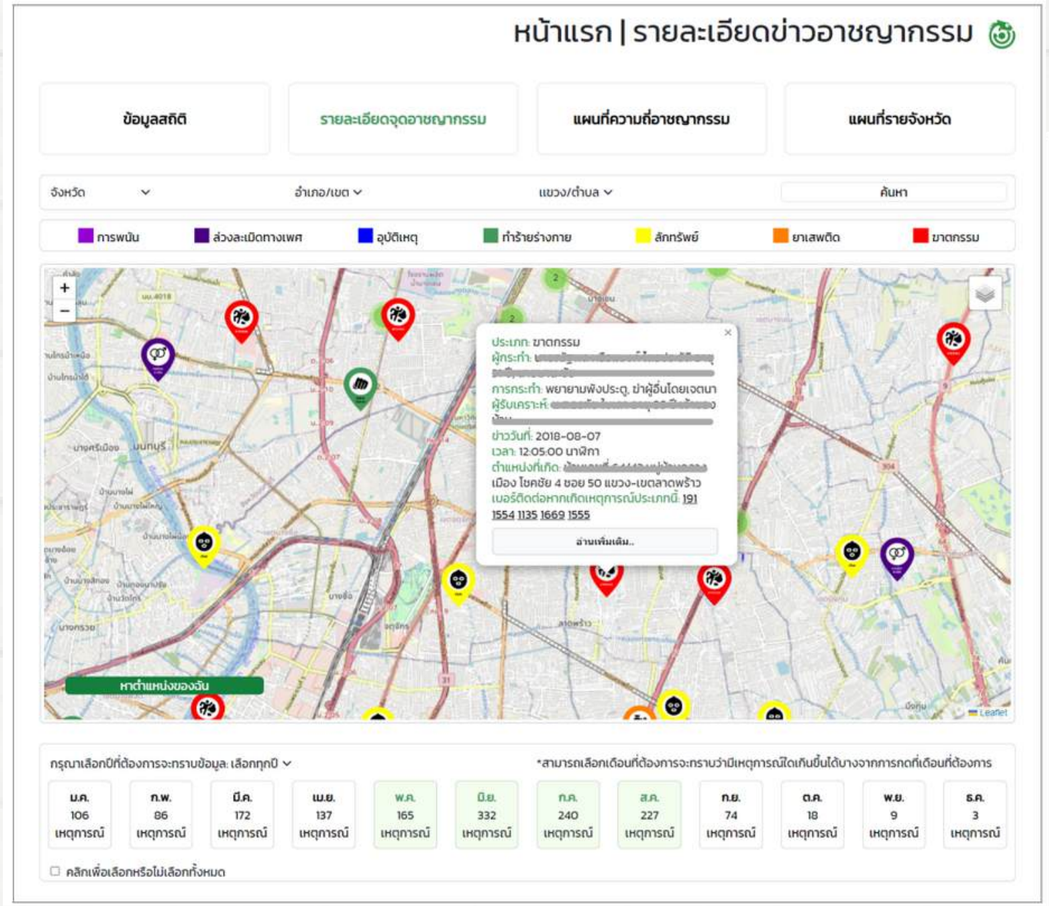
Resultados:

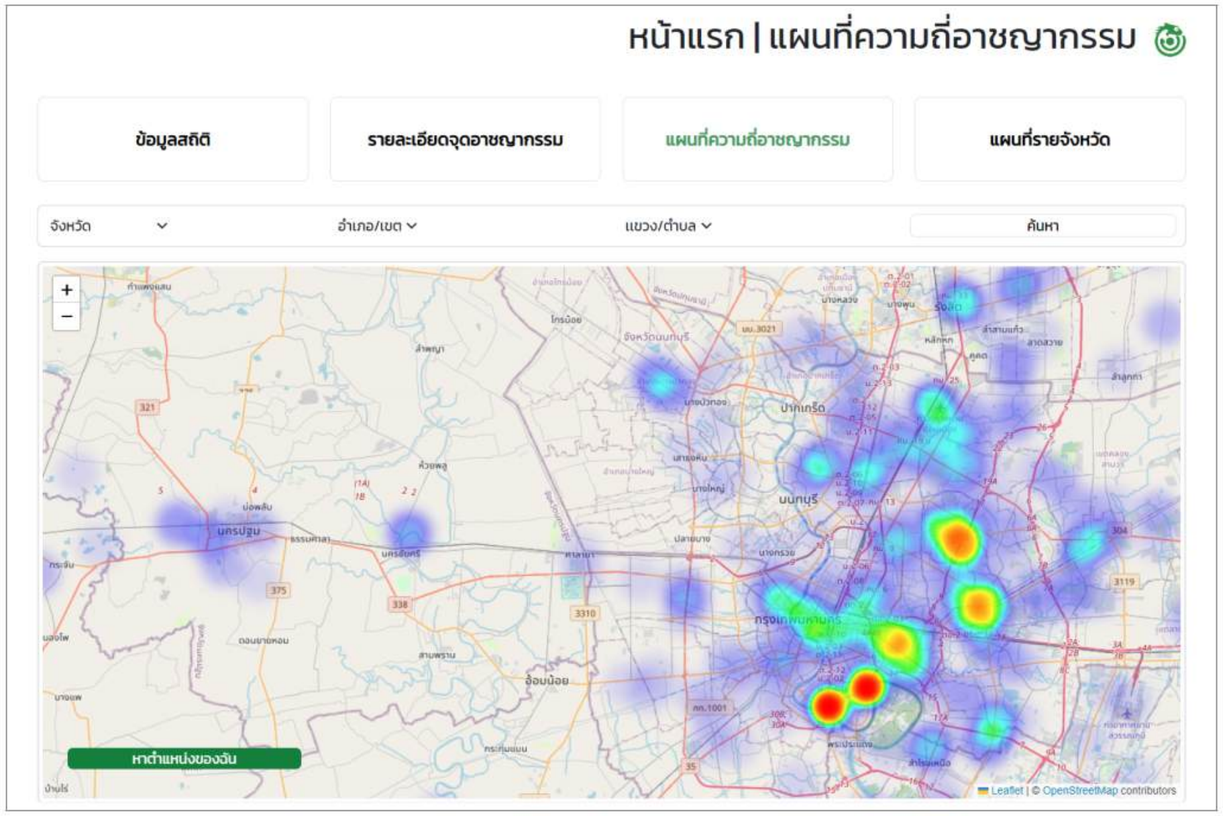
Comparación del rendimiento entre diferentes algoritmos de clasificación en la tarea de extracción de metadatos de crimen.

Model	Label	Precision	Recall	F1	MCC	Accuracy
CRF	Criminal	0.63	0.58	0.60	0.60	0.98
	Victim	0.52	0.39	0.45	0.45	0.99
	Police	0.46	0.53	0.49	0.48	0.98
	Date/Time	0.47	0.57	0.51	0.51	0.99
	Location	0.71	0.65	0.67	0.67	0.99
	Item	0.57	0.68	0.62	0.62	0.98
	Action	0.09	0.10	0.09	0.08	0.98
	Worth	0.55	0.37	0.44	0.45	0.99
	Root Cause	0.30	0.11	0.16	0.17	0.99
	Trigger	0.23	0.04	0.06	0.09	0.99
	Average	0.45	0.40	0.41	0.41	0.99
BiLSTM -CRF	Criminal	0.66	0.62	0.64	0.63	0.99
	Victim	0.61	0.52	0.56	0.56	0.99
	Police	0.49	0.61	0.54	0.54	0.98
	Date/Time	0.52	0.23	0.32	0.34	0.99
	Location	0.82	0.61	0.70	0.70	0.99
	Item	0.66	0.64	0.65	0.64	0.99
	Action	0.18	0.03	0.05	0.07	0.99
	Worth	0.61	0.32	0.42	0.44	0.99
	Root Cause	0.33	0.02	0.05	0.09	0.99
	Trigger	0.45	0.00	0.01	0.04	0.99
	Average	***0.53***	0.36	0.39	0.40	0.99
Wangchan BERTa	Criminal	0.65	0.65	0.65	0.64	0.99
	Victim	0.58	0.61	0.60	0.59	0.99
	Police	0.50	0.70	0.59	0.58	0.97
	Date/Time	0.54	0.64	0.58	0.58	0.99
	Location	0.78	0.88	0.81	0.80	0.99
	Item	0.55	0.84	0.67	0.67	0.98
	Action	0.23	0.21	0.22	0.21	0.97
	Worth	0.49	0.51	0.50	0.49	0.98
	Root Cause	0.31	0.15	0.20	0.20	0.99
	Trigger	0.33	0.16	0.21	0.22	0.99
	Average	0.50	***0.54***	0.50	0.50	0.98
XLMR	Criminal	0.66	0.66	0.66	0.65	0.99
	Victim	0.59	0.64	0.62	0.61	0.99
	Police	0.50	0.69	0.58	0.57	0.98
	Date/Time	0.58	0.69	0.63	0.63	0.99
	Location	0.73	0.88	0.80	0.80	0.99
	Item	0.56	0.84	0.67	0.68	0.98
	Action	0.25	0.21	0.23	0.22	0.98
	Worth	0.49	0.49	0.49	0.48	0.98
	Root Cause	0.33	0.16	0.22	0.22	0.99
	Trigger	0.33	0.15	0.21	0.22	0.99
	Average	0.50	***0.54***	***0.51***	***0.51***	0.99

CAMELON

UI:


หน้าแรก | รายละเอียดข่าวอาชญากรรม
ข้อมูลสถิติ
รายละเอียดจุดอาชญากรรม
แผนที่ความถี่อาชญากรรม
แผนที่รายจังหวัด
จังหวัด
อำเภอ/เขต
แขวง/ตำบล
ค้นหา
การพนัน
ล่วงละเมิดทางเพศ
อุบัติเหตุ
ทำร้ายร่างกาย
ลักทรัพย์
ยาเสพติด
ฆาตกรรม
ประเภท: ฆาตกรรม
การกระทำ: พยายามพังประตู, ฆ่าผู้อื่นโดยเจตนา
ข่าววันที่: 2018-08-07
เวลา: 12:05:00 นาฬิกา
เมือง โชคชัย 4 ซอย 50 แขวง-เขตลาดพร้าว
เบอร์ติดต่อหากเกิดเหตุการณ์ประเภทนี้: 191 1554 1135 1669 1555
อ่านเพิ่มเติม..
หาตำแหน่งของฉัน
Leaflet
กรุณาเลือกปีที่ต้องการจะทราบข้อมูล: เลือกทุกปี
*สามารถเลือกเดือนที่ต้องการจะทราบว่ามีเหตุการณ์ใดเกิดขึ้นบ้างจากการกดที่เดือนที่ต้องการ
ม.ค. 106 เหตุการณ์
ก.พ. 86 เหตุการณ์
มี.ค. 172 เหตุการณ์
เม.ย. 137 เหตุการณ์
พ.ค. 165 เหตุการณ์
มิ.ย. 332 เหตุการณ์
ก.ค. 240 เหตุการณ์
ส.ค. 227 เหตุการณ์
ก.ย. 74 เหตุการณ์
ต.ค. 18 เหตุการณ์
พ.ย. 9 เหตุการณ์
ธ.ค. 3 เหตุการณ์
คลิกเพื่อเลือกหรือไม่เลือกทั้งหมด


หน้าแรก | แผนที่ความถี่อาชญากรรม
ข้อมูลสถิติ
รายละเอียดจุดอาชญากรรม
แผนที่ความถี่อาชญากรรม
แผนที่รายจังหวัด
จังหวัด
อำเภอ/เขต
แขวง/ตำบล
ค้นหา
หาตำแหน่งของฉัน
Leaflet | © OpenStreetMap contributors

CAMELON

UI:

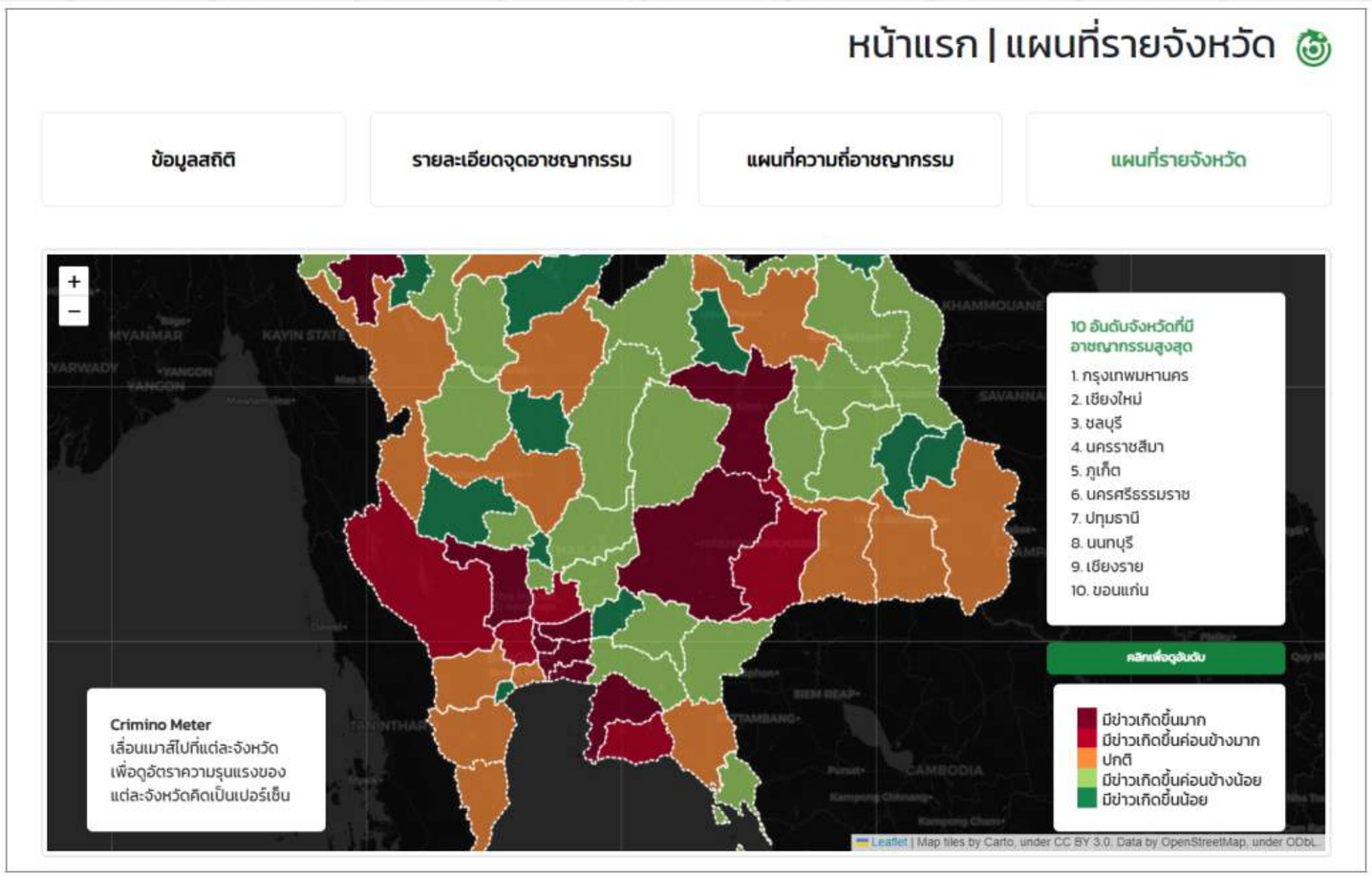


FIGURE 14: Example snapshot of the *Criminometer Map View* component.

Beyond administrative reports: a deep learning framework for classifying and monitoring crime and accidents leveraging large-scale online news

Ref: [2]

CRIMSON

Contexto:

Los crímenes y accidentes afectan gravemente la economía y la salud mental de las comunidades, especialmente en países en desarrollo. Sin embargo, los métodos tradicionales de monitoreo, basados en reportes administrativos, suelen ser lentos y poco accesibles. Aunque se han considerado las redes sociales como una fuente alternativa, presentan problemas de lenguaje informal, ruido y baja fiabilidad. En contraste, las noticias en línea ofrecen contenido más confiable y bien redactado, lo que las convierte en una fuente prometedora para la extracción precisa de información sobre estos eventos.

CRIMSON

Problema:

- **E**xiste una falta de herramientas efectivas que permitan clasificar automáticamente noticias sobre crímenes y accidentes y monitorear estos eventos en tiempo real, usando fuentes confiables como noticias en línea.
- No se ha validado rigurosamente si los datos extraídos de noticias pueden representar fielmente las estadísticas oficiales de crímenes y accidentes.

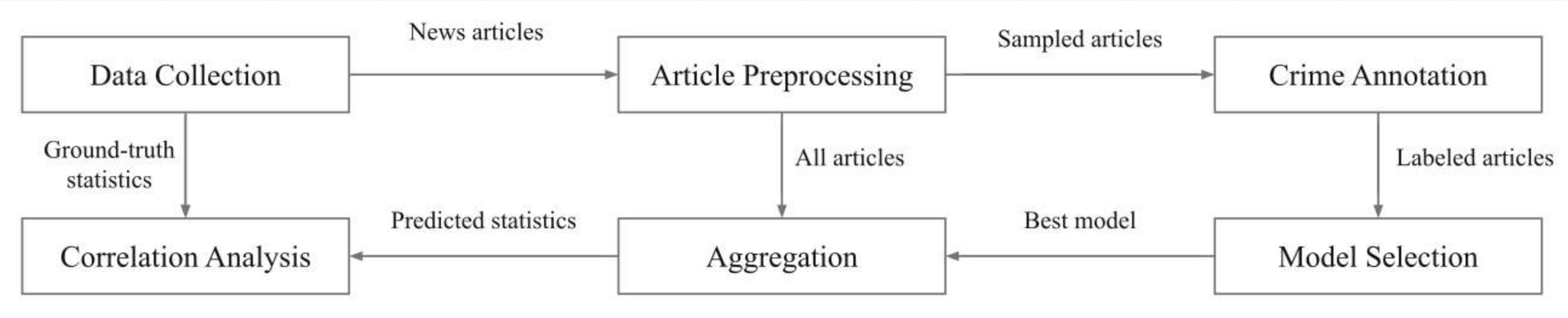
CRIMSON

Objetivo:

- Proponer CRIMSON, un framework inteligente basado en aprendizaje profundo para:
  - Clasificar automáticamente artículos de noticias en múltiples tipos de crímenes/accidentes (multi-label classification).
- Realizar validaciones cruzadas con estadísticas oficiales de crímenes y accidentes en Tailandia para medir la correlación entre las noticias clasificadas y los datos reales.
- Demostrar que las noticias online pueden ser una fuente confiable y oportuna para complementar o incluso mejorar los métodos tradicionales de monitoreo de crímenes.

CRIMSON

Metodología:



CRIMSON

Metodología:

1- Recolección de datos

- Se recolectaron aproximadamente 1.5 millones de artículos desde 2009 hasta 2021 de dos fuentes de noticias tailandesas.
- También se recopilaron estadísticas oficiales de delitos y accidentes para validación:
  - Estadísticas anuales de delitos como asesinato, agresión, violación, y robo
  - Estadísticas mensuales de accidentes de tránsito.

CRIMSON

Metodología:

2- Preprocesamiento de datos

- Se eliminaron elementos irrelevantes de los HTML (publicidad, menús, etc.).
- Se extrajo solo: fecha de publicación, título, introducción y cuerpo del artículo.
- Se conservaron únicamente los elementos textuales necesarios para la clasificación.

CRIMSON

Metodología:

3- Anotación de datos

- Se definieron 8 categorías: juegos de azar, asesinato, abuso sexual, robo, drogas, agresión, accidente, y no crimen.
- Un conjunto de 8,567 artículos fue anotado manualmente por voluntarios de forma multi-etiqueta (una noticia puede tener más de una clase).

CRIMSON

Metodología:

4- Clasificación de noticias

- Se implementó una tarea de clasificación de texto multi-etiqueta, usando diferentes técnicas:
  - Modelos clásicos: Naive Bayes, SVM, XGBoost (con TF-IDF).
  - Embeddings: BiLSTM con Thai2Vec.
  - Modelos preentrenados: WangchanBERTa (Thai), mBERT (multilingüe), XLM-R (cross-lingual).
- Se entrenaron usando la pérdida de entropía cruzada binaria y técnicas de ajuste fino estándar.

CRIMSON

Metodología:

5- Validación cruzada con datos reales

- Se realizó análisis de correlación de Pearson entre:
  - Estadísticas extraídas del modelo,
  - Datos oficiales del gobierno tailandés.
- Se encontró alta correlación para tipos como agresión, abuso sexual y accidentes.

CRIMSON

Resultados:

Estadísticas de los artículos de noticias anotados para la tarea de clasificación por tipo de crimen. Cabe señalar que un artículo puede estar anotado con múltiples tipos de crimen.

Class	# Samples	Proportion
Gambling	249	2.91%
Murder	2,557	29.85%
Sexual abuse	673	7.86%
Theft/Burglary	774	9.03%
Drug	1,039	12.13%
Battery/Assault	1,889	22.05%
Accident	721	8.42%
Non-Crime/Accident	1,406	16.41%

CRIMSON

Resultados:

Comparación del rendimiento de clasificación en la tarea de clasificación por tipo de crimen en términos de F1.

Classifier	Gambling	Murder	Sexual abuse	Theft/ Burglary	Drug	Battery/ assault	Accident	Non-crime/ Accident	Avg. crime/ Accident	Avg. all classes
NB	0.335	0.760	0.744	0.598	0.690	0.611	0.682	0.733	0.631	0.644
SVM	0.883	0.886	0.864	0.776	0.872	0.716	0.816	0.816	0.831	0.829
XGB	0.875	0.887	0.877	0.756	0.867	0.689	0.781	0.803	0.819	0.817
BiLSTM	0.707	0.833	0.711	0.611	0.750	0.606	0.724	0.755	0.706	0.712
WangchanBERTa	0.888	0.905	0.889	0.778	0.907	0.720	0.845	0.809	0.848	0.843
MBERT	0.008	0.854	0.753	0.658	0.849	0.641	0.733	0.789	0.642	0.661
XLMR-Base	***0.903***	0.907	0.889	0.784	0.903	0.727	0.824	0.823	0.848	0.845
XLMR-Large	0.887	***0.917***	***0.904***	***0.818***	***0.916***	***0.753***	***0.846***	***0.839***	***0.863***	***0.860***

CRIMSON

Resultados:

Coeficientes de correlación entre los casos reales de agresión/asalto, asesinato, abuso sexual y accidentes

Normalization	Crime/Accident type	Ground-truth statistics	Period	# of Articles	
				S1	S2
None	Battery/Assault	Battery/Assault	2016–2020 (Yearly)	0.981***	0.858**
	Murder	Murder	2016–2020 (Yearly)	0.367	0.025
	Sexual abuse	Rape	2016–2020 (Yearly)	0.570*	0.564*
	Theft/Burglary	Theft/Burglary	2016–2020 (Yearly)	0.217	0.001
	Accident	Accident-Death	Jan–Dec 2020 (Monthly)	0.173	0.035
	Accident	Accident-Injure	Jan–Dec 2020 (Monthly)	0.360	0.153
	Accident	Accident-Total	Jan–Dec 2020 (Monthly)	0.358	0.151
	Accident	Accident-Death	Jan–Oct 2020 (Monthly)	0.623*	0.463*
	Accident	Accident-Injure	Jan–Oct 2020 (Monthly)	0.620*	0.401*
	Accident	Accident-Total	Jan–Oct 2020 (Monthly)	0.621*	0.402*
Normalized	Battery/Assault	Battery/Assault	2016–2020 (Yearly)	−0.518	0.569*
	Murder	Murder	2016–2020 (Yearly)	0.419*	−0.137
	Sexual abuse	Sexual abuse	2016–2020 (Yearly)	−0.642	−0.641
	Theft/Burglary	Theft/Burglary	2016–2020 (Yearly)	0.313	0.509*
	Accident	Accident-Death	Jan–Dec 2020 (Monthly)	−0.061	−0.146
	Accident	Accident-Injure	Jan–Dec 2020 (Monthly)	−0.020	−0.051
	Accident	Accident-Total	Jan–Dec 2020 (Monthly)	−0.020	−0.053
	Accident	Accident-Death	Jan–Oct 2020 (Monthly)	0.246	0.223
	Accident	Accident-Injure	Jan–Oct 2020 (Monthly)	0.159	0.181
	Accident	Accident-Total	Jan–Oct 2020 (Monthly)	0.161	0.182

Referencias

[1] S. Pongpaichet et al., “CAMELON: A System for Crime Metadata Extraction and Spatiotemporal Visualization From Online News Articles,” IEEE Access, vol. 12, pp. 22778–22802, 2024, doi: 10.1109/ACCESS.2024.3363879.

[2] S. Tuarob, P. Tatiyamaneekul, S. Pongpaichet et al., “Beyond administrative reports: a deep learning framework for classifying and monitoring crime and accidents leveraging large-scale online news,” Neural Computing and Applications, vol. 37, pp. 7183–7205, 2025. [Online]. Available: https://doi.org/10.1007/s00521-024-10833-8

[3] İlgün, E.G., Dener, M. Exploratory data analysis, time series analysis, crime type prediction, and trend forecasting in crime data using machine learning, deep learning, and statistical methods. Neural Comput & Applic (2025). https://doi.org/10.1007/s00521-025-11094-9