

ESCUELA COLOMBIANA DE INGENIERÍA
JULIO GARAVITO

PROYECTO DE GRADO

Advanced Natural Language Processing Techniques to Profile Cybercriminals

Autor:

Alejandro ANZOLA ÁVILA

Director:

Dr. Daniel Orlando DÍAZ LÓPEZ

Programa de Ingeniería de Sistemas

BOGOTÁ, COLOMBIA



30 de marzo de 2019

Índice general

1. Introducción	1
1.1. Objetivo general	1
1.2. Objetivos específicos	1
2. Cronograma	2
3. Marco teórico	4
3.1. Análisis de vínculos (Link Analysis)	4
3.2. Agentes de software (Software Agents)	5
3.3. Sistemas Basados en Conocimiento (Knowledge Based Systems)	5
3.3.1. Fuzzy Knowledge Based Systems	6
3.4. Minería de datos (Data Mining)	6
3.4.1. Minería de texto (Text Mining)	6
3.5. Redes Neuronales (Neural Networks)	6
3.5.1. Sistemas de Detección de Anomalías (Anomaly Detection Systems)	6
3.5.2. Mapa autoorganizado (Self-organizing Maps)	6
3.6. Aprendizaje de maquina (Machine Learning)	8
4. Estado del arte	10
5. Propuesta	12
5.1. Análisis	12
5.2. Diseño	13
5.3. Resultados	14
Glosario	17
Bibliografía	19

Índice de figuras

2.1. Diagrama Gantt de actividades de 1 ^{er} periodo	2
2.2. Diagrama Gantt de actividades de 2 ^{do} periodo	2
3.1. Arquitectura KBS	5
3.2. Proceso de adaptación de SOM	8
3.3. Ejemplo de salida de SOM	9
3.4. Ejemplo de uso de SOM en aplicaciones de perfilado	9

Índice de cuadros

2.1. Detalle de cronograma de actividades	3
-----------------------------------------------------	---

Capítulo 1

Introducción

1.1. Objetivo general

El objetivo de este proyecto es generar herramientas y estrategias para el perfilado de cibercriminales con ayuda de metodologías de *Natural Language Processing (NLP)* aplicado a datos recolectados de comunicaciones y redes sociales.

1.2. Objetivos específicos

- Diseñar e implementar una solución de lenguaje natural para realizar el perfilado de sospechosos.
- Identificar el estado del arte en sistemas que usan NLP para apoyar agencias de seguridad del Estado.
- Construcción de artefactos para la construcción de *Datasets* con información recolectada de medios privados como de fuentes abiertas.
- Validar la solución desarrollada frente a un escenario real.
- Modelado de diferentes metodologías, heurísticas y meta–heurísticas para NLP.

Capítulo 2

Cronograma

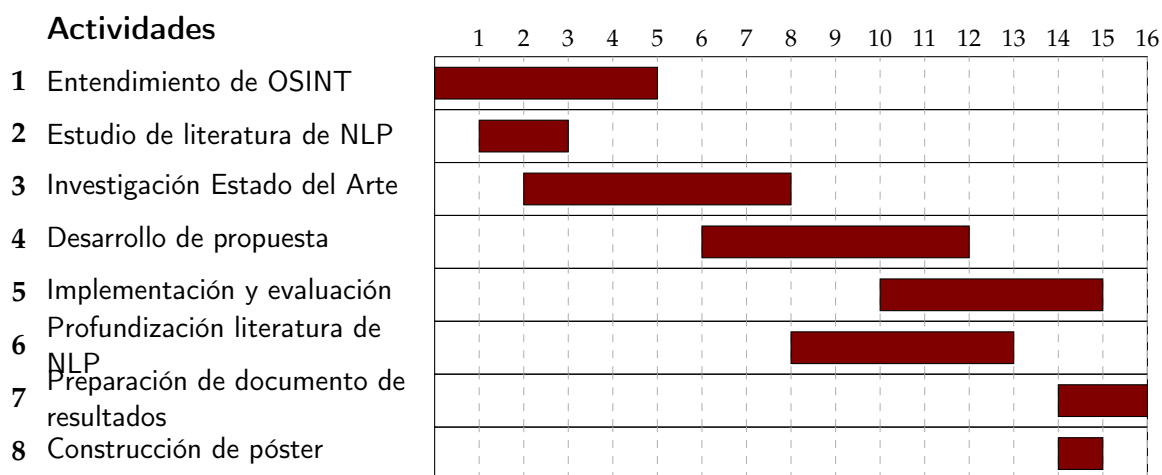


FIGURA 2.1: Diagrama Gantt de actividades de 1^{er} periodo

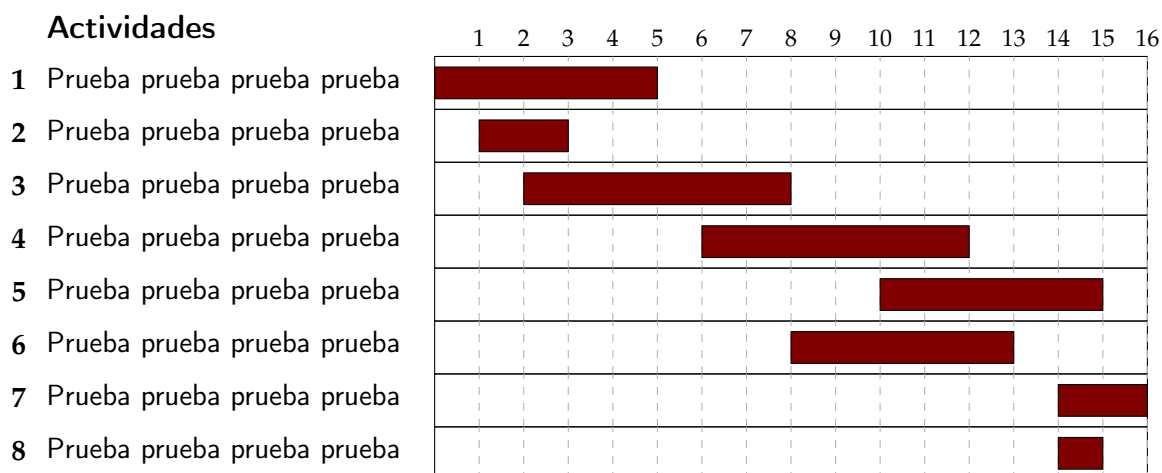


FIGURA 2.2: Diagrama Gantt de actividades de 2^{do} periodo

	Detalle
1	Entendimiento de NLP y proyecto <i>Open Source Intelligence (OSINT)</i>
2	Estudio de la literatura
3	Investigación del Estado del Arte
4	Desarrollo de la propuesta de investigación
5	Desarrollo de la implementación y pruebas
6	Realización de curso de NLP de National Research University Higher School of Economics de <i>Coursera</i>
7	Preparación final de documento de libro de proyectos y artículo de investigación
8	Construcción de póster para presentación en la Vitrina Académica

CUADRO 2.1: Detalle de cronograma de actividades

Capítulo 3

Marco teórico

La Web contiene una gran cantidad de opiniones respecto a productos, políticos, y mucho mas, expresado en forma de noticias, sitios de opinión, reseñas en tiendas online, redes sociales. Como resultado, el problema de “Minería de opinión” ha obtenido una atención creciente en las ultimas dos décadas y es un factor decisivo para las nuevas organizaciones (como es mencionado en [5]). De esto mismo partimos que el análisis de textos para extraer el significado y demás componentes extraíbles del texto componen un factor que debe considerarse al momento de realizar decisiones, de manera que los avances hechos hasta ahora tienen como meta una aplicación practica de lo que se conoce como Natural Language Processing.

Luego de los ataques terroristas del 11 de Septiembre de 2001 en Estados Unidos, se realizaron fuertes criticas respecto a la inteligencia, donde el director del FBI *Robert S. Mueller* indico que el principal problema que la agencia tuvo fue que se enfocaba demasiado en lidiar con el crimen luego de que fue cometido y ponía muy poco énfasis en prevenirlo (adaptado de [3]). Es por esto que el uso de NLP para temas de seguridad como también de metodologías de Machine Learning y Deep Learning han sido ampliamente utilizadas en ámbito de seguridad luego de estos eventos.

Para obtener una mejor inteligencia se necesito de mejores tecnologías a las que se tenían entonces (véase [3, pág 2]):

- Integración de datos (ó *Data Integration (DI)* en inglés)
- Análisis de vínculos (ó *Link Analysis (LA)* en inglés)
- Agentes de software (ó *Software Agents (SA)* en inglés)
- Minería de texto (ó *Text Mining (TM)* en inglés)
- Redes neuronales (ó *Neural Networks (NN)* en inglés)
- Algoritmos de Machine Learning (ó *Machine Learning Algorithms (MLA)* en inglés)

3.1. Análisis de vínculos (Link Analysis)

Es la visualización de asociaciones entre entidades y eventos, por lo general involucran una visualización por medio de una gráfica o un mapa que muestre las relaciones entre sospechosos y ubicaciones, sea por medio físico o por comunicaciones en la red.

3.2. Agentes de software (Software Agents)

Es el software que realiza tareas asignadas por el usuario de manera autónoma, donde sus habilidades básicas son:

- **Realización de tareas:** Hacen obtención de información, filtrado, monitoreo y reporte.
- **Conocimiento:** Pueden usar reglas programadas, o pueden aprender reglas nuevas (véase 3.3).
- **Habilidades de comunicación:** Reportar a humanos e interactuar con otros agentes.

3.3. Sistemas Basados en Conocimiento (Knowledge Based Systems)

Según [6], los *Knowledge Based Systems (KBS)* son uno de los mayores miembros de la familia de *Artificial Intelligence (AI)*. El KBS consiste de una *Knowledge Base (KB)* y un programa de búsqueda llamado *Inference Engine (IE)* representado en la figura 3.1. La KB puede ser usado como un repositorio de conocimiento de varias formas.

Existen 5 tipos de KBS, donde uno de ellos es conocido como *Expert Systems (ES)*, donde su uso se da como *Rule-based Systems (RBS)*, donde su KB esta dado como reglas y su IE esta dado por algo llamado *Working Memory (WM)*, que representa los hechos que se conocen inicialmente del sistema junto con los hechos que se van dando como inferencia de las reglas.

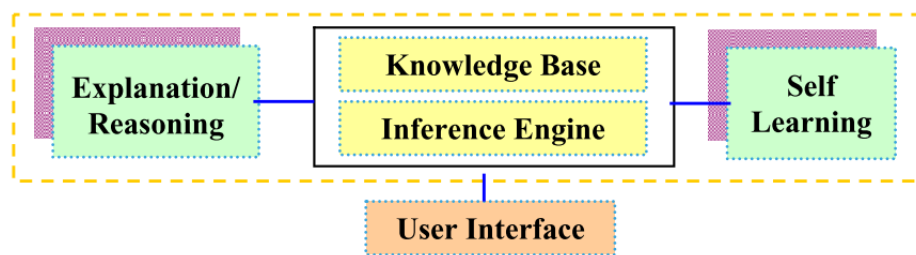


FIGURA 3.1: Arquitectura KBS. Tomado de [6]

Estas reglas pueden resumirse como una colección de condicionales de la forma **IF/ELSE** que se componen de un *antecedente* y un *consecuente*.

Existen dos tipos de RBS, definidos como *Deductive Systems (DS)* y *Reactive Systems (RS)*, donde el DS tiene como objetivo realizar una conclusión en base a los hechos iniciales en la WM, por el otro lado se tienen los RS, los cuales de igual manera a los DS, toman los hechos de la WM y realizan sea una acción interactiva con su entorno o bien una modificación de los hechos que se encuentran en la WM tal como la adición o eliminación de hechos. Tómese el ejemplo de la ecuación 3.1 tomada de [4], donde x es la temperatura y AC es aire acondicionado.

$$\begin{cases} \text{IF } x \text{ es moderado,} & \text{THEN } y = \text{ajustar AC a bajo} \\ \text{IF } x \text{ es alto,} & \text{THEN } y = \text{ajustar AC a moderado a alto} \\ \text{IF } x \text{ es muy alto,} & \text{THEN } y = \text{ajustar AC a alto} \end{cases} \quad (3.1)$$

3.3.1. Fuzzy Knowledge Based Systems

Pendiente

3.4. Minería de datos (Data Mining)

Según [7], la minería de datos se define como el proceso de descubrir información útil en repositorios grandes de datos. Las técnicas de minería de datos son desplegadas para limpiar grandes bases de datos para encontrar patrones nuevos y útiles que podrían permanecer desconocidos de lo contrario. También ofrecen capacidades para predecir la salida de observaciones futuras, tales como predecir si un cliente nuevo gastara mas de \$100 en una tienda.

No todas las tareas de descubrimiento de información son considerados como *Data Mining (DM)*. Por ejemplo, realizar una consulta de campos individuales usando un sistema de base de datos o encontrar una pagina web por medio de una búsqueda en Internet son tareas relacionadas con *adquisición de información*.

3.4.1. Minería de texto (Text Mining)

Es un subcampo de Inteligencia Artificial conocida como Natural Language Processing, en donde las herramientas de minería de datos pueden capturar rasgos críticos del contenido de un documento basado en el análisis de sus características lingüísticas.

La mayoría de los crímenes son electrónicos por naturaleza, por lo que se dejan rastros textuales que investigadores pueden seguir y analizar. Estas se enfocan en el descubrimiento de relaciones en texto no-estructurado y pueden ser aplicados al problema de *búsqueda y localización de palabras clave*.

3.5. Redes Neuronales (Neural Networks)

Pendiente

3.5.1. Sistemas de Detección de Anomalías (Anomaly Detection Systems)

Pendiente

3.5.2. Mapa autoorganizado (Self-organizing Maps)

El objetivo principal de los *Self-organizing Maps (SOM)* es de transformar una patrón de entrada m -dimensional en un mapa discreto uni- o bi-dimensional, donde sus principales características es que es un algoritmo que se basa en *Unsupervised Learning*, es *Feedforward*, es de una sola capa de neuronas donde su propósito es de realizar

Clustering y realizar una reducción de dimensionalidad sobre los datos de una forma topológicamente ordenada.

Los SOM tienen tres características distintivas:

- **Competencia:** por cada patrón de entrada, las neuronas en la red competirán entre ellas para determinar un ganador.
- **Cooperación:** la neurona ganadora determina la ubicación espacial (vecinos) alrededor de donde otras vecinas también se verán estimuladas.
- **Adaptación:** la neurona ganadora como también sus vecinas tendrán sus pesos asociados actualizados, y se tiene que los vecinos entre mas cerca estén del ganador, mayor es el grado de adaptación.

El algoritmo de aprendizaje de SOM parte de primero inicializar los pesos de las o neuronas con pesos aleatorios pequeños de una distribución de probabilidad aleatoria o uniforme, donde cada vector de entrada se define como $x = [x_1, \dots, x_m]^T \in \mathbb{R}^m$ y la entrada general de N patrones como $\mathbf{X}^{m \times N}$, el vector de pesos de la neurona i es $\mathbf{w}_i = [w_{i1}, \dots, w_{im}] \in \mathbb{R}^{1 \times m}$, con la matriz de pesos $\mathbf{W}^{o \times m}$.

Para alcanzar el objetivo de *competencia*, se realiza por cada patrón de entrada x_i una comparación con cada uno de los pesos de las o neuronas y se establece la de menor distancia respecto x_i (típicamente la distancia Euclidiana), dejando un ganador *winner*, tal como en la ecuación 3.2.

$$winner = \operatorname{argmin}_j \|x_i - w_j\|; j = 1, \dots, o \quad (3.2)$$

Luego de establecer la neurona ganadora, se realiza el paso para alcanzar la *cooperación*, que consiste en que por medio de una función kernel h (típicamente una distribución gaussiana), que permite establecer un área de afectación de las otras neuronas según su ubicación física en el mapa, definidos como r_{winner} y r_j que son la ubicación de la neurona ganadora y la neurona vecina j , en el cual el grado de afectación de la neurona vecina depende de la distancia de la que esta de la neurona ganadora, definido en la ecuación 3.3.

$$h_{j,winner}(t) = \exp\left(\frac{-\|r_j - r_{winner}\|^2}{2\sigma(t)^2}\right) \quad (3.3)$$

Parte importante del proceso de convergencia del SOM es que a medida que avanzan las iteraciones t del algoritmo el área de afectación se va reduciendo como parte del proceso de adaptación, por lo que definimos $\sigma(t) = \sigma_0 \exp(-t/\tau_1)$, donde τ_1 es una constante heurística y σ_0 la dimensión del mapa SOM.

Finalmente para alcanzar la *adaptación* se realiza una actualización de los pesos de la matriz \mathbf{W} en base a la influencia de área $\sigma(t)$ y de una tasa de aprendizaje $\alpha(t) = \alpha_0 \exp(-t/\tau_2)$, donde τ_2 es otra constante heurística y α_0 es una constante de aprendizaje inicial, que debe ser $0 \leq \alpha_0 \leq 1$, la actualización se describe por la ecuación 3.4 y el proceso puede ser visto gráficamente en la figura 3.2, tanto de forma uni- como bi-dimensional.

$$w_j(t+1) = w_j(t) + \alpha(t)h_{j,winner}(t)[x_i - w_j(t)] \quad (3.4)$$

Luego de que el algoritmo de aprendizaje termina de realizar las iteraciones, la salida de este es la matriz de pesos \mathbf{W} , en la figura 3.3 se puede apreciar una aproximación del

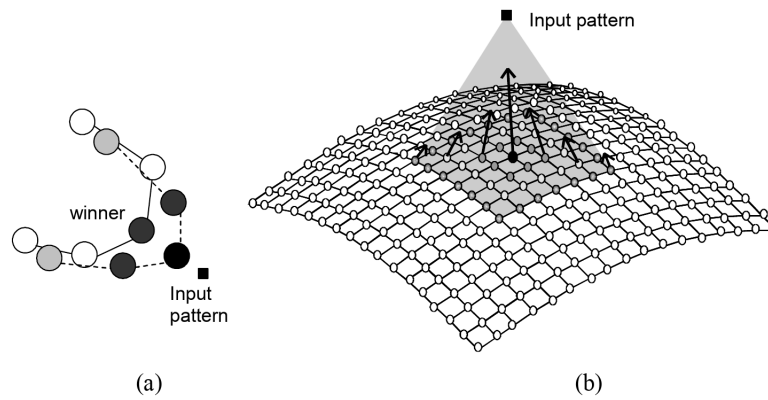


FIGURA 3.2: Proceso de adaptación de SOM, (a) uni-dimensional, (b) bi-dimensional. Tomado de [1]

algoritmo con un mapa uni-dimensional tratando de aproximar una función sinusoidal con ruido adicionado en un gráfico 2D.

3.6. Aprendizaje de maquina (Machine Learning)

Pendiente

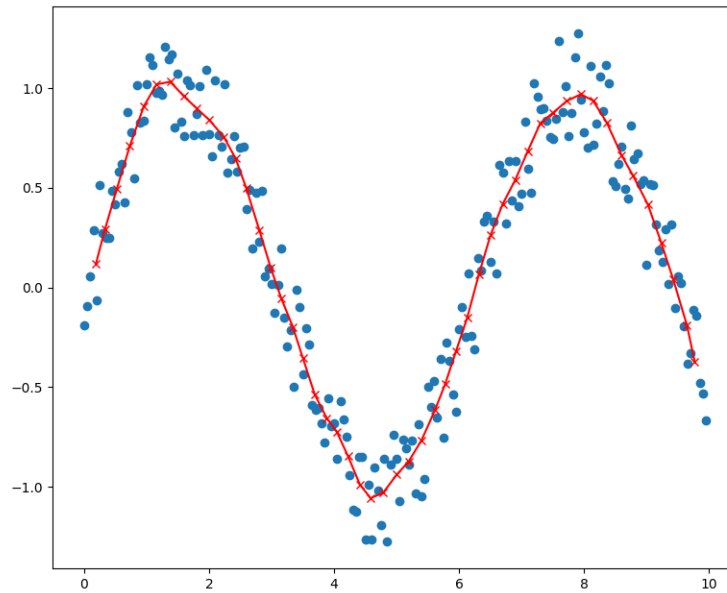


FIGURA 3.3: Ejemplo de salida de SOM. Implementación propia.

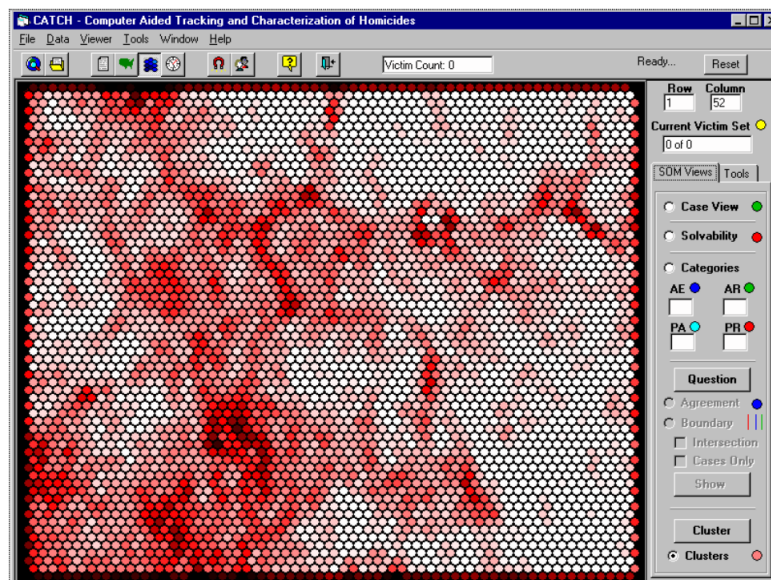


FIGURA 3.4: Ejemplo de uso de SOM en aplicaciones de perfilado. Tomado de [3]

Capítulo 4

Estado del arte

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus

eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Capítulo 5

Propuesta

5.1. Análisis

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

5.2. Diseño

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante.

Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

5.3. Resultados

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac,

nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia

nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Glosario

Símbolos

Clustering

Pendiente. 7

Coursera

Sitio web de cursos de aprendizaje en <https://www.coursera.org>. 3

Dataset

Pendiente. 1

Feedforward

Pendiente. 6

Unsupervised Learning

Pendiente. 6

Artificial Intelligence (AI)

Pendiente. 5

Data Integration (DI)

Para acceder a múltiples y diversas fuentes de información. 4

Data Mining (DM)

Proceso de descubrir automáticamente información útil en repositorios grandes de datos. 6

Deductive Systems (DS)

Pendiente. 5

Expert Systems (ES)

Pendiente. 5

Inference Engine (IE)

Pendiente. 5

Knowledge Based Systems (KBS)

Pendiente. 5

Knowledge Base (KB)

Pendiente. 5

Link Analysis (LA)

Para visualizar asociaciones y relaciones criminales y terroristas. 4

Machine Learning Algorithms (MLA)

Para extraer perfiles de perpetradores y mapas gráficos de crímenes. 4

Neural Networks (NN)

Para predecir la probabilidad de crímenes y nuevos ataques terroristas. 4

Natural Language Processing (NLP)

Rama de la inteligencia artificial que lidia con la interacción entre computadores y humanos usando el lenguaje natural. 1–4

Open Source Intelligence (OSINT)

Disciplina responsable de la adquisición, procesamiento y posterior transformación en inteligencia de información obtenida de fuentes públicas como prensa, radio, televisión, internet, informes de diferentes sectores y, en general, cualquier recurso de acceso público (Tomado de [2]). 3

Reactive Systems (RS)

Pendiente. 5

Rule-based Systems (RBS)

Pendiente. 5

Software Agents (SA)

Para el monitoreo, obtención, análisis y actuación sobre la información. 4

Self-organizing Maps (SOM)

Pendiente. 6, 7

Text Mining (TM)

Búsqueda sobre terabytes de información en documentos, páginas web y correos electrónicos. 4

Working Memory (WM)

Pendiente. 5

Bibliografía

- [1] Leandro Nunes De Castro. *Fundamentals of natural computing: basic concepts, algorithms, and applications*. Chapman y Hall/CRC, 2006.
- [2] Martín José Hernández Medina, Ricardo Andrés Pinto Rico y Cristian Camilo Pinzón Hernández. *Inteligencia de fuentes abiertas para el contexto colombiano*. Inf. téc. Escuela Colombiana de Ingeniería Julio Garavito, 2018.
- [3] J. Mena. *Investigative Data Mining for Security and Criminal Detection*. Elsevier Science, 2003. ISBN: 9780080509389. URL: <https://books.google.com.co/books?id=3mDlrtJuZv4C>.
- [4] Jerry M Mendel. *Uncertain Rule-Based Fuzzy Systems*. ISBN: 9783319513690. DOI: [10.1007/978-3-319-51370-6](https://doi.org/10.1007/978-3-319-51370-6).
- [5] Ana Maria Popescu y Orena Etzioni. «Extracting product features and opinions from reviews». En: *Natural Language Processing and Text Mining* (2007), págs. 9-28. ISSN: 08247935. DOI: [10.1007/978-1-84628-754-1_2](https://doi.org/10.1007/978-1-84628-754-1_2). arXiv: [0309034](https://arxiv.org/abs/0309034) [cs].
- [6] Priti Srinivas Sajja y Rajendra Akerkar. «Knowledge-based systems for development». En: *Advanced Knowledge Based Systems: Model, Applications & Research 1* (2010), págs. 1-11.
- [7] Pang-Ning Tan, Michael Steinbach y Vipin Kumar. *Introduction to Data Mining*. US ed. Addison Wesley, mayo de 2005. ISBN: 0321321367. URL: <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0321321367>.
- [8] Patrick Henry Winston. *Artificial Intelligence (3rd Ed.)* Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1992. ISBN: 0-201-53377-4.