

# Tabla de Contenido

Índice de Tablas	ix
Índice de Ilustraciones	x
<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes generales . . . . .	1
1.1.1. Caracterización del problema . . . . .	2
1.1.2. Justificación de este trabajo . . . . .	2
1.2. Objetivos . . . . .	3
1.2.1. Obejtivo general . . . . .	3
1.2.2. Objetivos específicos . . . . .	3
1.3. Hipótesis del trabajo . . . . .	4
1.4. Metodología . . . . .	4
1.4.1. CRISP-DM . . . . .	5
1.5. Resultados esperados . . . . .	6
1.6. Alcance . . . . .	7
1.7. Estructura del documento . . . . .	7
<b>2. Comprensión del delito, el actuar criminal y su aprehensión</b>	<b>9</b>
2.1. Interpretación económica del delito . . . . .	9
2.2. La criminología y el comportamiento antisocial . . . . .	10
2.2.1. Triángulo del delito . . . . .	12
2.3. Interpretación jurídica del delito y su castigo . . . . .	13
2.3.1. La importancia de la pena y su función . . . . .	13
2.4. La vinculación criminal y el uso de la tecnología . . . . .	14
2.4.1. Modelos de vinculación criminal . . . . .	15
<b>3. Modelos y conceptos de text mining</b>	<b>16</b>
3.1. Conceptos básicos . . . . .	16
3.1.1. Preprocesamiento de texto . . . . .	16
3.1.2. Term-Document Matrix . . . . .	17
3.1.3. Neural Networks . . . . .	17
3.1.4. Deep Learning . . . . .	19
3.2. Recuperación de documentos . . . . .	21
3.2.1. Latent Semantic Indexing . . . . .	22
3.3. Word embeddings . . . . .	24
3.3.1. Word2Vec . . . . .	24

3.3.2.	GloVe . . . . .	26
3.3.3.	FastText . . . . .	28
3.3.4.	Comparación Word Embeddings . . . . .	30
3.4.	Modelos de similitud semántica textual . . . . .	30
3.4.1.	Doc2Vec . . . . .	30
3.4.2.	Word Mover’s Distance . . . . .	32
3.4.3.	WMD vs Doc2vec . . . . .	33
3.5.	Latent Dirichlet Allocation . . . . .	33
3.6.	Métricas de desempeño . . . . .	34
3.6.1.	Similitud Coseno . . . . .	34
3.6.2.	Precision . . . . .	35
3.6.3.	Recall . . . . .	36
3.6.4.	F1-Measure . . . . .	36
3.6.5.	Perplexity . . . . .	36
3.6.6.	Topic Coherence . . . . .	37
<b>4.</b>	<b>Desarrollo metodológico</b>	<b>38</b>
4.1.	Entendimiento del contexto . . . . .	38
4.1.1.	Estadísticas . . . . .	40
4.2.	Comprensión los datos . . . . .	40
4.2.1.	Registro de robos en lugar habitado Macrozona San Antonio . . . . .	41
4.2.2.	Foco de investigación Fiscalía Local San Antonio . . . . .	42
4.2.3.	Discusión . . . . .	45
4.3.	Preparación de los datos . . . . .	45
4.4.	Modelamiento . . . . .	45
4.4.1.	Recuperación de documentos . . . . .	47
4.4.2.	Computo de similitud semántica textual . . . . .	48
4.4.3.	Implementación Word Mover’s Distance . . . . .	48
4.4.4.	Implementación Word embeddings . . . . .	48
4.4.5.	Implementación Doc2Vec . . . . .	49
4.4.6.	Implementación Latent Dirichlet Allocation . . . . .	49
4.5.	Evaluación de resultados . . . . .	50
4.5.1.	Implementación de visualizaciones y resumen de contenido . . . . .	50
4.5.2.	Representación de los documentos a través de grafos . . . . .	50
4.5.3.	Tablas de resumen de contenido . . . . .	51
4.5.4.	Implementación de una visualización interactiva para tópicos . . . . .	51
4.5.5.	Juicio de expertos . . . . .	52
<b>5.</b>	<b>Resultados</b>	<b>54</b>
5.1.	Asociación de causas por similitud semántica textual . . . . .	55
5.1.1.	Análisis exploratorio de los datos . . . . .	55
5.1.2.	Elección de los términos de búsqueda . . . . .	57
5.1.3.	Recuperación de documentos . . . . .	58
5.1.4.	Computo de similitud semántica textual: Word Mover’s Distance . . . . .	60
5.1.5.	Computo de similitud semántica textual: Doc2Vec . . . . .	60
5.1.6.	Interpretación de los resultados . . . . .	63
5.2.	Seguimiento del Foco de Investigación . . . . .	70

5.2.1. Selección de la muestra . . . . .	70
5.2.2. Resultados del seguimiento al Foco de Investigación . . . . .	71
5.2.3. Visualización de los resultados . . . . .	71
<b>6. Conclusiones</b>	<b>74</b>
6.1. Recomendaciones para trabajos futuros . . . . .	75
<b>Bibliografía</b>	<b>77</b>
<b>Anexos</b>	<b>82</b>