

Tabla de Contenido

1. Introducción	1
1.1. Motivación	1
1.1.1. Aplicación móvil y origen del conjunto de datos	2
1.1.2. Acercamiento al problema mediante ciencia de datos	3
1.2. Objetivos	4
1.2.1. Objetivo General	4
1.2.2. Objetivos Específicos	4
1.2.3. Alcances	4
2. Marco Teórico	5
2.1. Aprendizaje de Máquinas	6
2.1.1. Inteligencia Artificial	6
2.1.1.1. Contexto	6
2.1.1.2. Origen	6
2.1.2. Aprendizaje de Máquinas	7
2.1.3. Modelos de clasificación	7
2.1.4. Redes neuronales artificiales	8
2.1.5. Métricas de desempeño	9
2.1.6. Balanceo del conjunto de datos	11
2.2. Predicción del crimen	12
2.2.1. <i>Hotspots</i>	12
2.2.1.1. Grillas	12
2.2.1.2. Análisis de densidad	13
2.2.1.3. Agrupación distrital	13
2.2.2. Tipos de modelos de predicción del crimen	13
2.2.3. Tipos de datos utilizados	14
2.2.4. Estado del arte y contribución	15
2.3. Aspectos éticos	17
2.3.1. Sesgos presentes en el conjunto de datos	17
2.4. Caracterización de la empresa	19
2.5. Contexto geográfico y división político-administrativa	20
3. Metodología	21
3.1. Análisis exploratorio de datos	21
3.1.1. Categorías presentes en el conjunto de datos	22
3.1.2. Distribución espacial de los reportes	24
3.1.3. Distribución temporal de los reportes	25

3.1.4.	Calidad de los datos	25
3.2.	Modelamiento del conjunto de datos	26
3.2.1.	Modelamiento espacial del conjunto de datos	26
3.2.2.	Modelamiento temporal del conjunto de datos	26
3.2.2.1.	Diseño del nuevo conjunto de datos	26
3.2.2.2.	Construcción del nuevo conjunto de datos	28
3.2.3.	Categorías	29
3.2.4.	Determinación de la variable objetivo	31
3.2.5.	Filtro por región y estadísticas por comuna	31
3.2.5.1.	Filtro espacial de los datos	31
3.2.5.2.	Estadísticas por comuna	32
3.2.6.	Configuraciones previas a la predicción	33
3.2.6.1.	Incorporación de un componente cíclico estacional	33
3.2.6.2.	Escalamiento de los datos numéricos	33
3.2.6.3.	Subconjunto de entrenamiento y subconjunto de prueba	34
3.2.6.4.	Balanceo del conjunto de datos	34
3.3.	Implementación y evaluación de los modelos	35
3.3.1.	Algoritmos de clasificación binaria	35
3.3.2.	Métricas de desempeño	35
3.3.3.	Interpretabilidad de los resultados	35
4.	Resultados y discusión	36
4.1.	Resultados de los algoritmos empleados	37
4.2.	Relevancia de los atributos	43
4.2.1.	Relevancia de atributos para la configuración de dos–semanas	44
4.2.2.	Relevancia de atributos para la configuración semanal	45
4.2.3.	Relevancia de atributos para la configuración de tres–días	46
4.2.4.	Relevancia de atributos para la configuración diaria	47
4.3.	Discusión	48
4.3.1.	Sesgos propios de los datos	48
4.3.2.	Modelamiento temporal y espacial de los datos	49
4.3.2.1.	Discusiones sobre el modelamiento temporal de los datos	49
4.3.2.2.	Discusiones sobre el modelamiento espacial de los datos	51
4.3.2.3.	Incorporación de comunas y variables complementarias	51
4.3.3.	Implementación de los modelos	53
4.3.4.	Escalabilidad de los modelos implementados	53
4.3.5.	Recomendaciones para la organización	54
5.	Conclusiones	55
6.	Trabajos Futuros	57
	Bibliografía	59
	Anexos	62
A.	Categorías	62
B.	Modelamiento del conjunto de datos	68
C.	Tiempos de procesamiento	70