

# TABLA DE CONTENIDO

RESUMEN .....	I
DEDICATORIA .....	II
AGRADECIMIENTOS .....	III
TABLA DE CONTENIDO .....	IV
ÍNDICE DE TABLAS Y CUADROS .....	VI
ÍNDICE DE ILUSTRACIONES .....	VII
1 Contexto .....	1
2 Objetivos .....	3
2.1 Objetivo general .....	3
2.2 Objetivos específicos .....	3
3 Justificación del proyecto .....	4
4 Marco Conceptual .....	5
4.1 El uso de la ciencia de datos para la generación de evidencia que oriente la toma de decisiones .....	5
4.2 Minería de texto y sus aplicaciones .....	6
4.2.1 Text mining .....	7
5 Metodología .....	14
5.1 Explorar .....	14
5.1.1 Descripción de las fuentes de datos .....	14
5.2 Preparar Datos .....	14
5.2.1 Descripción de los subsets de datos a utilizar para el análisis .....	14
5.3 Planificar modelo: .....	15
5.3.1 Gensim-LDA .....	15
5.3.2 Mallet .....	16
5.3.3 Análisis de las salidas de los modelos .....	16
5.4 Elaborar modelo y comunicar sus resultados .....	16
5.4.1 Normalización de los datos .....	17
5.4.2 Procesamiento de los datos .....	22
5.4.3 Aplicación de modelos .....	22
5.4.4 Análisis de outputs .....	23
5.4.5 Estimación de los tiempos del proceso .....	25
6 Construcción de una implementación con foco en el autoservicio .....	26

7	Resultados .....	27
7.1	Implementación del modelo seleccionado en la detección de grupos de consumidores en eventuales casos colectivos.....	27
7.2	Implementación del modelo seleccionado para el análisis expedito de grandes volúmenes de reclamos .....	30
7.3	Implementación del modelo en los reclamos de un mercado, para observar tendencias generales. ....	32
8	Conclusiones .....	35
9	Bibliografía.....	38
	Anexos .....	40
	Anexo A: Tabla de resumen de los tópicos dominantes asociados a los reclamos de la empresa de retail 1 .....	40
	Anexo B: Tabla de resumen de los tópicos dominantes asociados a los reclamos de la empresa de retail 2 .....	42
	Anexo C: Tabla de resumen del lenguaje y bibliotecas utilizadas en el proceso .....	43