

Tabla de Contenido

1. Introducción	1
1.1. Proyecto OpinionZoom	1
1.1.1. Aporte del Proyecto de Grado	2
1.2. Objetivos	2
1.2.1. Objetivo General	2
1.2.2. Objetivos Específicos	2
1.3. Hipótesis de Investigación	3
1.4. Solución Tecnológica	3
1.5. Resultados Esperados	3
1.6. Estructura de la Tesis	3
2. Marco Conceptual	5
2.1. La Web y Redes Sociales en Chile	5
2.2. Data Mining	8
2.3. Antecedentes de la Disciplina	8
2.4. Text Opinion Mining	10
2.5. Detección de Tópicos	11
2.5.1. Latent Dirichlet Allocation	13
2.6. Enfoque a Procesos	14
3. Antecedentes y Alcances	17
3.1. Contexto de Mercado	17
3.2. Antecedentes de OpinionZoom	20
3.2.1. Origen del Proyecto	20
3.2.2. Misión, Visión y Valores de la Organización	21
3.2.3. Objetivos	22
3.2.4. Nuevo Paradigma: de Investigación a Investigación Aplicada	22
3.3. Análisis de Estrategia	24
3.3.1. Fuerzas de Porter	24
3.3.2. Modelo Delta de Hax	29
3.4. Lineamientos del Proyecto	32
3.4.1. Detección de Necesidades de Mercado	33
3.4.2. Levantamiento de Requerimientos de Negocio	35
3.4.3. Levantamiento de Requerimientos Técnicos	37
4. Propuesta de Valor	40

4.1.	Modelo de Negocio	40
4.2.	Interés Complementario	42
4.2.1.	Concepto de la Propuesta	43
4.2.2.	Valor Agregado	51
4.3.	Caracterización de Clientes	52
4.3.1.	Segmentación de Clientes	52
4.3.2.	Propuesta General a Clientes	55
5.	Valoración del Negocio	61
5.1.	Cuantificación del Mercado	61
5.1.1.	Afinidad de Segmentos de Clientes	62
5.1.2.	Extrapolación al Mercado Chileno	64
5.2.	Pricing	66
5.3.	Estructura Organizacional	68
5.3.1.	Personal Requerido	70
5.3.2.	Carga Laboral	70
5.4.	Cuantificación Económica del Negocio	73
5.4.1.	Estructura de Costos	73
5.4.2.	Modalidad de Ingresos	81
5.4.3.	Evaluación Económica del Proyecto	83
5.4.4.	Evaluación Económica del Servicio Inteligencia de Clientes	88
6.	Diseño de la Solución	92
6.1.	Arquitectura de Macroprocesos	92
6.2.	Procesos Comprendidos	95
6.3.	Lógica de Negocio	100
6.4.	Diseño de la Herramienta	101
6.4.1.	Sistema de Crawling	104
6.4.2.	Sistema de Creación de Modelo Parametrizado	110
6.4.3.	Proceso de Extracción, Transformación y Carga	119
7.	Implementación y Puesta en Marcha	132
7.1.	Infraestructura Tecnológica	132
7.1.1.	Hardware de Trabajo	132
7.1.2.	Softwares de Trabajo y Apoyo	132
7.2.	Modelo de Datos	135
7.2.1.	Modelo para Creación del Modelo Parametrizado	136
7.2.2.	Modelo para Sistema de Detección de Intereses: ETL	136
7.3.	Desarrollo de la Solución	138
7.3.1.	Desarrollo Sistema de Creación del Modelo Parametrizado	138
7.3.2.	Desarrollo del Proceso de ETL	139
7.4.	Análisis Preliminares y Desempeño	141
7.4.1.	Limpieza del Corpus	141
7.4.2.	Muestra Aleatoria	142
7.4.3.	Tiempo de Procesamiento de Creación del Modelo	144
7.5.	Cantidad Óptima de Tópicos	145
7.5.1.	Análisis de Coherencia	145

7.5.2.	Análisis Exploratorio y Factorial	148
7.6.	Caracterización de Tópicos	151
7.6.1.	Clasificación	151
7.7.	Tiempos del Proceso del ETL	154
8.	Evaluación y Análisis de la Propuesta	155
8.1.	Análisis Económico: Sensibilidad del Proyecto	155
8.1.1.	Metodología y Supuestos	155
8.1.2.	Resultados del Análisis de Sensibilidad	157
8.2.	Resultados Sistema Desarrollado	161
8.2.1.	Resultado del Etiquetado	161
8.2.2.	Desempeño de la Herramienta de ETL	161
8.2.3.	Capacidad de Clasificación de la Herramienta	163
9.	Conclusiones	166
9.1.	De Negocio	166
9.2.	De Procesos	167
9.3.	De Investigación	167
9.4.	Propuesta de Mejora y Trabajo Futuro	168
	Bibliografía	169
	Appendices	173
	A. Modelo de Datos	174
	B. Limpieza de Corpus de Entrenamiento	176
	C. Listado de Tópicos	183