

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes generales . . . . .	2
1.1.1. Características de la empresa . . . . .	2
1.1.2. Desempeño y crecimiento . . . . .	3
1.1.3. Área de trabajo . . . . .	4
<b>2. Planteamiento del problema y justificación</b>	<b>5</b>
2.1. Oportunidad identificada . . . . .	5
2.2. Propuesta de valor de la solución planteada . . . . .	8
<b>3. Objetivos y alcance</b>	<b>9</b>
3.1. Objetivo general . . . . .	9
3.2. Objetivos específicos . . . . .	9
3.3. Alcances y resultados esperados . . . . .	9
<b>4. Marco conceptual</b>	<b>11</b>
4.1. Clustering . . . . .	11
4.1.1. Métodos de clustering . . . . .	12
4.1.2. Algoritmos jerárquicos . . . . .	12
4.1.3. Clustering jerárquico . . . . .	13
4.2. Elección de medidas de distancia . . . . .	14
4.2.1. Medidas de distancia basadas en correlación . . . . .	15
4.3. Aplicación de una medida de distancia basada en correlación . . . . .	16
4.4. Medidas de evaluación de clúster . . . . .	17
4.4.1. Coeficiente de silueta . . . . .	17
4.4.2. Índices de validación externos . . . . .	18
4.5. Modelo global de optimización . . . . .	19
4.5.1. Métricas de evaluación para modelo global de optimización . . . . .	20
4.6. Cross Industry Standard Process For Data Mining (CRISP-DM) . . . . .	23
<b>5. Metodología</b>	<b>25</b>
5.1. CRISP-DM aplicado al caso de estudio . . . . .	25
5.1.1. Comprensión del negocio . . . . .	26
5.1.2. Comprensión de los datos . . . . .	26
5.1.3. Preparación de los datos . . . . .	27
5.1.4. Modelado . . . . .	27
5.1.5. Evaluación . . . . .	27
5.1.6. Implementación . . . . .	28

<b>6. Desarrollo metodológico</b>	<b>29</b>
6.1. Obtención y comprensión de los datos . . . . .	29
6.1.1. Análisis exploratorio . . . . .	31
6.1.2. Análisis de PLP a través de sus posiciones, por categoría . . . . .	32
6.1.3. Análisis de PLP a través de sus posiciones, por número de página . . . . .	35
6.2. Preparación de los datos . . . . .	37
6.2.1. Reestructuración de los datos . . . . .	37
6.3. Modelado . . . . .	39
6.4. Resultados del modelado . . . . .	40
6.4.1. Linkage Complete . . . . .	40
6.4.2. Linkage Average . . . . .	41
6.4.3. Linkage Single . . . . .	42
6.4.4. Linkage Weighted . . . . .	43
6.4.5. Elección de un número K de clúster . . . . .	44
6.5. Evaluación de la clusterización . . . . .	46
6.5.1. Análisis clúster 9 . . . . .	48
6.5.2. Análisis clúster 1 . . . . .	51
6.5.3. Análisis 14 . . . . .	53
6.6. Evaluación . . . . .	55
6.6.1. Construcción de features y target . . . . .	55
6.6.2. Construcción del modelo por clúster . . . . .	56
6.6.3. Entrenamiento para cada modelo . . . . .	56
<b>7. Resultados</b>	<b>59</b>
7.1. Resultados por clúster . . . . .	59
7.2. Análisis de NDCG . . . . .	62
7.2.1. Análisis de NDCG para modelo por clúster . . . . .	63
<b>8. Conclusiones</b>	<b>71</b>
8.1. Trabajos futuros . . . . .	73
<b>Bibliografía</b>	<b>73</b>
<b>Anexo A. Antecedentes generales</b>	<b>76</b>
A.1. Anexo 1 . . . . .	76
<b>Anexo B. Oportunidad identificada</b>	<b>77</b>
B.1. Anexo 2 . . . . .	77
<b>Anexo C. Marco teórico</b>	<b>78</b>
C.1. Anexo 3 . . . . .	78
C.1.1. Algoritmos de partición . . . . .	78
C.1.2. K-Means . . . . .	78
C.2. Anexo 4 . . . . .	79
C.2.1. Medidas de distancia geométrica . . . . .	79
C.3. Anexo 5 . . . . .	80
C.3.1. Métricas de evaluación para modelo global de optimización . . . . .	80
C.4. Anexo 6 . . . . .	80

C.4.1. Estado del arte en clusterización de distribuciones . . . . .	80
<b>Anexo D. Desarrollo metodológico</b>	<b>82</b>
D.1. Anexo 7 . . . . .	82
D.1.1. Funnel de compra . . . . .	82
D.2. Anexo 8 . . . . .	83
D.2.1. Elección de un número k de clúster . . . . .	83
D.3. Anexo 9 . . . . .	86
D.3.1. Resultados de la clusterización . . . . .	86
<b>Anexo E. Resultados</b>	<b>98</b>
E.1. Anexo 10 . . . . .	98
E.1.1. Resultados por clúster . . . . .	98
E.2. Anexo 11 . . . . .	104
E.2.1. Análisis de NDCG . . . . .	104
E.2.2. Comparación de NDCG entre modelos . . . . .	106