

ÍNDICE

| | |
|--|-----------|
| INTRODUCCIÓN | 1 |
| CAPÍTULO 1 | 2 |
| 1 DESCRIPCIÓN DEL CONTEXTO | 3 |
| 1.1 MERCADO DE TELEFONÍA MÓVIL | 5 |
| CAPÍTULO 2 | 9 |
| 2 DESCRIPCIÓN DEL PROBLEMA | 10 |
| 2.1 OBJETIVO GENERAL | 13 |
| 2.2 OBJETIVOS ESPECÍFICOS..... | 13 |
| CAPÍTULO 3 | 14 |
| 3 MARCO TEÓRICO | 15 |
| 3.1 MINERÍA DE DATOS..... | 15 |
| 3.2 DESCUBRIMIENTO DE CONOCIMIENTO EN BASE DE DATOS..... | 17 |
| 3.2.1 Preparación de los datos..... | 19 |
| 3.2.2 Recopilación..... | 19 |
| 3.2.3 Limpieza..... | 20 |
| 3.2.4 Tipo de Variables | 21 |
| 3.3 MATRIZ DE DATOS..... | 22 |
| 3.3.1 Transformación..... | 23 |
| 3.3.2 Binarización de la matriz de datos..... | 24 |
| 3.3.3 Transformación y Normalización..... | 25 |
| 3.4 SELECCIÓN DE ATRIBUTOS..... | 27 |
| 3.4.1 Clasificación para variables individuales..... | 28 |
| 3.4.2 Generación de subconjuntos | 29 |
| 3.4.3 Selección de subconjuntos de variables | 30 |
| 3.4.4 Selección de variables no supervisadas | 35 |
| 3.5 SEGMENTACIÓN..... | 36 |
| 3.5.1 Características de Segmentación..... | 37 |
| 3.5.2 Método Jerárquico..... | 38 |
| 3.5.3 Limitaciones de la agrupación jerárquica..... | 45 |
| 3.5.4 K-Medias..... | 46 |
| 3.5.5 Validación de clusters..... | 49 |

| | |
|--|-----------|
| CAPÍTULO 4 | 50 |
| 4 METODOLOGÍA | 51 |
| 4.1 PREPARACIÓN DE LOS DATOS..... | 51 |
| 4.1.1 <i>Pre-procesamiento</i> | 55 |
| 4.1.2 <i>Transformación de datos</i> | 61 |
| 4.2 SELECCIÓN DE ATRIBUTOS | 63 |
| 4.2.1 <i>Test Chi-Cuadrado</i> | 64 |
| 4.2.2 <i>Test Kolmogorov-Smirnov (K-S)</i> | 67 |
| 4.3 RESULTADOS | 73 |
| 4.3.1 <i>Selección de Grupos óptimos</i> | 73 |
| 4.3.2 <i>Segmentación mediante K-Medias</i> | 80 |
| CAPÍTULO 5 | 84 |
| 5 ANÁLISIS DE RESULTADOS Y CONCLUSIONES | 85 |
| 5.1 ANÁLISIS DE RESULTADOS..... | 85 |
| 5.2 CONCLUSIONES | 88 |
| 5.2.1 <i>Limitaciones y trabajos futuros</i> | 90 |
| REFERENCIAS | 91 |
| BIBLIOGRÁFICA | 91 |
| ANEXO | 96 |

TABLAS

| | |
|---|----|
| TABLA 3.1: TIPO DE VARIABLES..... | 22 |
| TABLA 3.2: MATRIZ DE DATOS..... | 23 |
| TABLA 3.3: MATRIZ DE DATOS BINARIOS..... | 25 |
| TABLA 3.4: TABLA DE DISTANCIAS..... | 43 |
| TABLA 3.5: TABLA DE DISTANCIAS, TRAS PRIMERA FUSIÓN..... | 43 |
| TABLA 4.1: GRUPO DE CLIENTES..... | 51 |
| TABLA 4.2: VARIABLES DE ESTUDIO..... | 52 |
| TABLA 4.3: VARIABLES CATEGÓRICAS..... | 54 |
| TABLA 4.4: VARIABLES CONTINUAS..... | 55 |
| TABLA 4.5: DISTRIBUCIÓN MESES ESTUDIO..... | 56 |
| TABLA 4.6: DISTRIBUCIÓN VARIABLE ESTADO..... | 57 |
| TABLA 4.7: DISTRIBUCIÓN VARIABLE TRÁFICO_MENSUAL..... | 58 |
| TABLA 4.8: DISTRIBUCIÓN VARIABLE PLAN_DATOS..... | 59 |
| TABLA 4.9: DISTRIBUCIÓN VARIABLE TRÁFICO_EXCEDIDO..... | 60 |
| TABLA 4.10: CLASIFICACIÓN DE VARIABLES..... | 61 |
| TABLA 4.11: VARIABLE DESCRIPCIÓN..... | 62 |
| TABLA 4.12: VARIABLES BINARIAS..... | 62 |
| TABLA 4.13: VARIABLES CONTINUAS..... | 62 |
| TABLA 4.14: TABLA DE FRECUENCIA INICIAL..... | 64 |
| TABLA 4.15: TABLA DE FRECUENCIA FINAL..... | 65 |
| TABLA 4.16: TEST CHI-CUADRADO VAR_OBJ X SI_METROPOLITANA..... | 65 |
| TABLA 4.17: TABLA FRECUENCIA VARIABLE DESCRIPCIÓN..... | 66 |
| TABLA 4.18: TEST K-S VARIABLES DE TRÁFICO..... | 67 |
| TABLA 4.19: AGRUPACIÓN CON DISTANCIA MÍNIMA..... | 74 |
| TABLA 4.20: AGRUPACIÓN CON DISTANCIA MÁXIMA..... | 76 |
| TABLA 4.21: AGRUPACIÓN CON DISTANCIA MEDIA..... | 77 |
| TABLA 4.22: AGRUPACIONES POR K-MEDIAS..... | 81 |
| TABLA 4.23: CARACTERÍSTICA DE LOS GRUPOS..... | 82 |
| TABLA 4.24: MEDIA DE LOS ATRIBUTOS..... | 83 |
| TABLA 5.1: MIGRACIÓN DE CLIENTES POR GRUPO..... | 86 |
| TABLA 5.2: CARACTERÍSTICA DE LOS GRUPOS..... | 86 |

FIGURAS

| | |
|---|----|
| FIGURA 1.1: FLUJO DEL SERVICIO DE TRANSMISIÓN DE DATOS | 8 |
| FIGURA 2.1: BENEFICIO GENERADO POR UN CLIENTE A LO LARGO DEL TIEMPO..... | 11 |
| FIGURA 2.2: RESULTADOS DE ENCUESTA A CLIENTES..... | 12 |
| FIGURA 3.1: ETAPAS DE PROYECTO DE MINERÍA DE DATOS | 16 |
| FIGURA 3.2: CARGA DE TRABAJO EN LAS FASES DE UN PROYECTO DE MINERÍA DE DATOS..... | 17 |
| FIGURA 3.3: DIAGRAMA PROCESO KDD | 18 |
| FIGURA 3.4: INGRESOS DE HOGARES DE UNA POBLACIÓN | 26 |
| FIGURA 3.5: DISTRIBUCIÓN K-S..... | 34 |
| FIGURA 3.6: DENDROGRAMA CLUSTER JERÁRQUICO..... | 38 |
| FIGURA 3.7: COMPORTAMIENTO CLUSTER JERÁRQUICO..... | 39 |
| FIGURA 3.8: EJEMPLO DE UN ÁRBOL DE AGRUPAMIENTO (DENDROGRAMA)..... | 41 |
| FIGURA 3.9: EJEMPLO DE DISTRIBUCIÓN DE PUNTOS BIDIMENSIONAL | 42 |
| FIGURA 3.10: AGRUPAMIENTOS POSIBLES MEDIANTE UN MÉTODO JERÁRQUICO | 44 |
| FIGURA 3.11: GRUPOS FORMADOS CON EL MÉTODO K-MEDIAS | 48 |
| FIGURA 5.1: GRUPOS ORDENADOS SEGÚN COMPORTAMIENTO..... | 87 |

GRÁFICOS

| | |
|--|----|
| GRÁFICO 1.1: PENETRACIÓN DE DIFERENTES SEGMENTOS DEL SECTOR TELECOMUNICACIONES | 3 |
| GRÁFICO 1.2: INVERSIÓN EN TELECOMUNICACIONES DESDE 2004-2008 (MILLONES DE PESOS) | 4 |
| GRÁFICO 1.3: PARTICIPACIÓN DE MERCADO | 6 |
| GRÁFICO 1.4: PARTICIPACIÓN DE MERCADO DE LAS COMPAÑÍAS DE TELEFONÍA MÓVIL..... | 6 |
| GRÁFICO 1.5: PENETRACIÓN DE TELEFONÍA MÓVIL DURANTE EL PERÍODO 2000-2008..... | 7 |
| GRÁFICO 3.1: CAMBIOS DE UNA DISTRIBUCIÓN..... | 26 |
| GRÁFICO 4.1: HISTOGRAMA DE TRÁFICO PROMEDIO DE LA MUESTRA | 68 |
| GRÁFICO 4.2: HISTOGRAMA DE TRÁFICO MÍNIMO DE LA MUESTRA..... | 69 |
| GRÁFICO 4.3: HISTOGRAMA DE TRÁFICO PROMEDIO CORREGIDO DE LA MUESTRA..... | 70 |
| GRÁFICO 4.4: HISTOGRAMA DE TRÁFICO MÍNIMO CORREGIDO DE LA MUESTRA | 70 |
| GRÁFICO 4.5: HISTOGRAMA DE TRÁFICO PROMEDIO NORMALIZADO DE LA MUESTRA..... | 71 |
| GRÁFICO 4.6: HISTOGRAMA DE TRÁFICO MÍNIMO NORMALIZADO DE LA MUESTRA | 72 |
| GRÁFICO 4.7: DISTANCIA MÍNIMA ENTRE GRUPOS..... | 75 |
| GRÁFICO 4.8: DISTANCIA MÁXIMA ENTRE GRUPOS..... | 76 |
| GRÁFICO 4.9: DISTANCIA MEDIA ENTRE GRUPOS..... | 78 |
| GRÁFICO 4.10: SEGMENTO DE CLIENTES MEDIANTE DISTANCIA MÍNIMA..... | 79 |
| GRÁFICO 4.11: SEGMENTO DE CLIENTES MEDIANTE DISTANCIA MÁXIMA..... | 79 |
| GRÁFICO 4.12: SEGMENTO DE CLIENTES MEDIANTE DISTANCIA MÁXIMA..... | 80 |
| GRÁFICO 4.13: SEGMENTO DE CLIENTES MEDIANTE K-MEDIA | 81 |