

## TABLA DE CONTENIDO

CAPÍTULO 1. INTRODUCCIÓN .....	1
1.1 Antecedentes generales .....	1
1.2 Antecedentes de la empresa .....	3
CAPÍTULO 2. DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN .....	5
CAPÍTULO 3. OBJETIVOS .....	9
3.1 Objetivo General .....	9
3.2 Objetivos Específicos.....	9
CAPÍTULO 4. ALCANCES.....	10
CAPÍTULO 5. MARCO CONCEPTUAL .....	12
5.1 Metodología <i>Knowledge Discovery in Databases</i> .....	12
5.2 Modelos de Regresión .....	12
5.2.1 Modelo de Regresión Lineal.....	12
5.2.1.1 Algoritmo <i>Stepwise</i> .....	13
5.2.1.2 Evaluación del modelo de Regresión Lineal .....	14
5.2.2 Modelo de Regresión de Fronteras .....	14
5.2.2.1 Modelo de frontera de producción estocástica.....	15
5.3 Método para estimar <i>Customer Lifetime Value</i> .....	17
5.3.1 Método RFM.....	17
CAPÍTULO 6. METODOLOGÍA .....	18
CAPÍTULO 7. DESARROLLO METODOLÓGICO .....	19
7.1 Análisis y selección de datos .....	19
7.1.1 Recopilación datos .....	19
7.1.2 Pre procesamiento de datos.....	20
7.1.3 Análisis descriptivo .....	22
7.1.4 Creación de variables.....	24
7.1.5 Selección de variables.....	27
7.1.6 Selección muestra de datos .....	30
7.2 Modelo predictivo de potencial de compra del cliente .....	31
7.2.1 Modelo simple: promedio general.....	31

7.2.1.1	Resultados por región .....	31
7.2.1.2	Resultados por zona .....	33
7.2.1.3	Resultados categorías principales .....	33
7.2.2	Modelo de regresión lineal .....	34
7.2.2.1	Modelo (I): Efectos de las ventas 2014 .....	34
7.2.2.2	Modelo (II): Categorías principales .....	43
7.2.3	Modelo de Regresión de Fronteras .....	49
7.2.3.1	Modelo (I): Efectos de las ventas 2014 .....	49
7.2.3.2	Modelo (II): Categorías principales .....	49
7.2.3.3	Cálculo de potencial de compra .....	53
7.2.4	Resultados complementarios .....	54
7.2.5	Elección del modelo de potencial de compra .....	54
7.3	Ranking de clientes .....	58
7.4	Propuestas de líneas de acción .....	63
CAPÍTULO 8. CONCLUSIONES .....		66
CAPÍTULO 9. TRABAJOS PROPUESTOS .....		69
BIBLIOGRAFÍA .....		70
ANEXOS .....		73

## ÍNDICE DE GRÁFICOS

<b>Gráfico 1:</b> Crecimiento del sector <i>retail</i> .....	1
<b>Gráfico 2:</b> Número de clientes y ventas según tipo de minorista 2014.....	10
<b>Gráfico 3:</b> Modelo de regresión de frontera estocástica.....	16
<b>Gráfico 4:</b> Histograma ventas 2014 de bodegas .....	21
<b>Gráfico 5:</b> <i>Box plot con y sin outliers</i> .....	21
<b>Gráfico 6:</b> Ventas totales y bodegas 2013 .....	23
<b>Gráfico 7:</b> Ejemplo – Bodegas similares .....	23
<b>Gráfico 8:</b> Ventas totales 2013 y 2014 por zona .....	26
<b>Gráfico 9:</b> Ventas propias vs. socios comerciales según categoría 2013 .....	29
<b>Gráfico 10:</b> Modelo (I) – Efecto cliente: comparación pesos de los parámetros betas en la regresión lineal 2 .....	36
<b>Gráfico 11:</b> Modelo (I) – Efecto contexto: Comparación pesos de los parámetros betas en la regresión lineal 2 .....	37
<b>Gráfico 12:</b> Modelo (I) – Efecto tipo de vendedor: comparación pesos de los parámetros betas en la regresión lineal 2.....	39
<b>Gráfico 13:</b> Modelo (I) – 3 Efectos: Comparación de parámetros betas de la regresión lineal 2.....	40
<b>Gráfico 14:</b> Categoría: Aceites – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas .....	51
<b>Gráfico 15:</b> Categoría aceites – Venta real vs. Venta potencial 2013 .....	58
<b>Gráfico 16:</b> Número de clientes por zonas y cuadrantes.....	61
<b>Gráfico 17:</b> Número de vendedores por tipo y cuadrantes .....	62
<b>Gráfico 18:</b> Categoría aceites – Venta real vs. Venta estimada por el modelo de frontera 2013 .....	63
<b>Gráfico 19:</b> Frontera de posibilidades de producción .....	78
<b>Gráfico 20:</b> Comparación de los modelos de regresión de fronteras .....	81
<b>Gráfico 21:</b> Categoría: Detergentes – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas .....	95
<b>Gráfico 22:</b> Categoría: Galletas – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas .....	96
<b>Gráfico 23:</b> Categoría: Pastas – Estadísticos principales de los modelos de regresión de fronteras según número de variables incluidas .....	97

<b>Gráfico 24:</b> Porcentaje de clientes por zona y cuadrantes .....	102
<b>Gráfico 25:</b> Porcentaje de vendedores por tipo y cuadrantes .....	103

## ÍNDICE DE TABLAS

<b>Tabla 1:</b> Análisis descriptivo de variables transaccionales .....	22
<b>Tabla 2:</b> Clasificación del tipo de vendedor según ventas 2013 (S/.) .....	24
<b>Tabla 3:</b> Datos según región.....	25
<b>Tabla 4:</b> Análisis descriptivo de variables RFM .....	27
<b>Tabla 5:</b> Matriz de correlación de Pearson de las variables transaccionales principales de los clientes.....	28
<b>Tabla 6:</b> Matriz de correlación de Pearson de las variables de región.....	28
<b>Tabla 7:</b> Análisis descriptivo de la venta 2013 (S/.) de las categorías propias .....	30
<b>Tabla 8:</b> Resultados preliminares de potencial de compra por región .....	32
<b>Tabla 9:</b> Resultados preliminares de potencial de compra por zonas.....	33
<b>Tabla 10:</b> Resultados preliminares de potencial de compra por categorías .....	34
<b>Tabla 11:</b> Modelo (I) – Efecto cliente: Resultados regresión lineal 1 (parámetros betas) .....	35
<b>Tabla 12:</b> Modelo (I) – 3 Efectos: Resultados regresión lineal 2.....	41
<b>Tabla 13:</b> Estadísticos principales de las regresiones lineales de los efectos .....	41
<b>Tabla 14:</b> Estadísticos del modelo de regresión lineal de los 3 efectos en conjunto .....	42
<b>Tabla 15:</b> Estimación del potencial de compra por zonas basado en el modelo de regresión lineal de los 3 efectos en conjunto .....	43
<b>Tabla 16:</b> Parámetros betas estimados por categoría de las regresiones lineales .....	46
<b>Tabla 17:</b> Estadísticos de las regresiones lineales de cada categoría.....	47
<b>Tabla 18:</b> Estimación potencial de compra según categorías.....	48
<b>Tabla 19:</b> Categoría: Aceites – Variables incluidas en los modelos de regresión de fronteras .....	50
<b>Tabla 20:</b> Parámetros betas estimados por categoría del modelo de fronteras estocásticos.....	52
<b>Tabla 21:</b> Eficiencia técnica de cada categoría del modelo de fronteras estocásticos ..	53
<b>Tabla 22:</b> Cálculo de potencial de compra según modelos de regresión de fronteras de cada categoría.....	53
<b>Tabla 23:</b> Comparación de los modelos de regresión lineal .....	55
<b>Tabla 24:</b> Comparación de los modelos de regresión de fronteras estocásticos.....	56
<b>Tabla 25:</b> Resumen de ventas de los cuadrantes de la categoría aceites.....	59

<b>Tabla 26:</b> Promedios de las variables transaccionales de los clientes de cada cuadrante .....	60
<b>Tabla 27:</b> Categorías a potenciar en orden de prioridad.....	65
<b>Tabla 28:</b> Modelo (I) – Efecto cliente: Resultados regresión lineal 1 (parámetros betas) .....	86
<b>Tabla 29:</b> Modelo (I) – Efecto cliente: Resultados regresión lineal 1 (estadísticos) .....	86
<b>Tabla 30:</b> Modelo (I) – Efecto cliente: Resultados regresión lineal 2 (parámetros betas) .....	86
<b>Tabla 31:</b> Modelo (I) – Efecto cliente: Resultados regresión lineal 2 (estadísticos) .....	87
<b>Tabla 32:</b> Modelo (I) – Efecto contexto: Resultados regresión lineal 1 (parámetros betas) .....	88
<b>Tabla 33:</b> Modelo (I) – Efecto contexto: Resultados regresión lineal 1 (estadísticos)....	89
<b>Tabla 34:</b> Modelo (I) – Efecto contexto: Resultados regresión lineal 2 (parámetros betas) .....	89
<b>Tabla 35:</b> Modelo (I) – Efecto contexto: Resultados regresión lineal 2 (estadísticos)....	89
<b>Tabla 36:</b> Modelo (I) – Efecto tipo de vendedor: Resultados regresión lineal 1 y 2 .....	90
<b>Tabla 37:</b> Modelo (I) – Efecto tipo de vendedor: Resultados regresión lineal 1 y 2 (estadísticos) .....	90
<b>Tabla 38:</b> Resultados regresión lineal categoría aceites .....	91
<b>Tabla 39:</b> Resultados regresión lineal categoría detergentes .....	91
<b>Tabla 40:</b> Resultados regresión lineal categoría galletas .....	92
<b>Tabla 41:</b> Resultados regresión lineal categoría pastas .....	92
<b>Tabla 42:</b> Resultados regresión lineal categoría salsas.....	93
<b>Tabla 43:</b> Residuos de las regresiones lineales de cada categoría .....	93
<b>Tabla 44:</b> Modelo (I): 3 efectos – Resultados regresión de frontera .....	94
<b>Tabla 45:</b> Categoría: Detergentes – Variables incluidas en los modelos de regresión de fronteras .....	95
<b>Tabla 46:</b> Categoría: Galletas – Variables incluidas en los modelos de regresión de fronteras .....	96
<b>Tabla 47:</b> Categoría: Pastas – Variables incluidas en los modelos de regresión de fronteras .....	97
<b>Tabla 48:</b> Ineficiencia estocástica en cada Modelo (II) de fronteras .....	98

<b>Tabla 49:</b> Análisis descriptivo de las principales variables transaccionales, incluyendo la superficie de la bodega .....	99
<b>Tabla 50:</b> Resultados modelo de regresión lineal 1 con variables de la encuesta.....	100
<b>Tabla 51:</b> Resultados modelo de regresión lineal 2 con variables de la encuesta.....	101
<b>Tabla 52:</b> Porcentaje de clientes por zona y cuadrantes .....	102
<b>Tabla 53:</b> Resultados Test Chi-Cuadrado de clientes por zona y cuadrantes .....	102
<b>Tabla 54:</b> Porcentaje de vendedores por tipo y cuadrantes.....	103
<b>Tabla 55:</b> Resultados Test Chi-Cuadrado de clientes por zona y cuadrantes .....	103

## ÍNDICE DE ANEXOS

Anexo 1: Marco conceptual – KDD .....	73
Anexo 2: Marco conceptual – Regresión Lineal .....	74
Anexo 3: Marco conceptual – Evaluación de modelos .....	75
Anexo 4: Marco conceptual – Modelo de Regresión de Fronteras.....	78
Anexo 5: Marco conceptual – Modelo de frontera de producción determinística .....	80
Anexo 6: Marco conceptual – Comparación de modelos de fronteras de producción....	81
Anexo 7: Marco conceptual – Customer Lifetime Value .....	82
Anexo 8: Mapa político de Perú .....	84
Anexo 9: Matriz de correlación de Pearson entre ventas 2013 de cada categoría.....	85
Anexo 10: Modelo (I) - Regresión lineal efecto cliente .....	86
Anexo 11: Modelo (I) - Regresión lineal efecto contexto .....	88
Anexo 12: Modelo (I) - Regresión lineal efecto tipo de vendedor .....	90
Anexo 13: Resultados Modelo (II): Regresión lineal categorías principales .....	91
Anexo 14: Modelo (I) - Resultados regresión de fronteras estocásticas .....	94
Anexo 15: Selección de modelos de regresión de frontera para detergentes, galletas y pastas.....	95
Anexo 16: Ineficiencia estocástica en cada categoría del modelo de fronteras .....	98
Anexo 17: Resultados complementarios .....	99
Anexo 18: Test Chi-Cuadrado de clientes por zona y cuadrantes.....	102
Anexo 19: Test Chi-Cuadrado de vendedores por tipo y cuadrantes.....	103