

TABLA DE CONTENIDO

1	ANTECEDENTES GENERALES O INTRODUCCIÓN	1
1.1	Características de la organización/empresa	1
1.2	Mercado y/o marco institucional	2
1.3	Desempeño organizacional	2
2	DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN	3
2.1	Información del área de la organización/empresa	3
2.2	Justificación	4
2.2.1	Estado del arte	4
2.2.2	Oportunidad	5
2.3	Hipótesis y posibles alternativas de solución	6
2.4	Descripción del proyecto	6
2.5	Propuesta de valor de las posibles soluciones	8
3	OBJETIVOS	9
3.1	Objetivo general:	9
3.2	Objetivos específicos:	9
4	MARCO CONCEPTUAL	10
4.1	Teoría de Grafos:	10
4.1.1	Propiedades de nodos	10
4.1.2	Eccentricity (Excentricidad)	10
4.1.3	Centrality (Centralidad)	10
4.1.4	Eigenvector Centrality (Centralidad vector propio)	10
4.1.5	Betweenness (Intermediación)	10
4.1.6	Nodos terminales	11
4.1.7	Sub grafo	11
4.1.8	Filtro de redes	11
4.2	KPI	11
4.3	Metodología RFM	11
4.4	Diseño experimental	12
4.4.1	Campaña piloto	12
4.4.2	Campaña de referidos	12
4.4.3	Tasa de lectura	12
4.4.4	Tasa de respuesta	12
4.4.5	Grupo de control	13
4.4.6	Test Z de diferencia en proporciones	13

4.4.7	Efecto incremental	14
4.5	Métodos de balanceo	14
4.5.1	Undersampling	14
4.5.2	Oversampling	14
4.5.3	Smote (Synthetic minority oversampling technique)	15
4.6	Network Effect (Efecto red)	15
4.7	Algoritmos de aprendizaje supervisado	16
4.7.1	Linear regression (Regresión lineal)	16
4.7.2	Logistic regression (Regresión logística)	16
4.7.2.1	Odds Ratio	16
4.7.3	Decision tree (Arbol de decisión)	16
4.7.4	Random forest	17
4.7.5	Gradient boosting	17
4.7.6	Ada boost	17
5	METODOLOGÍA	17
5.1	Entendimiento del negocio (variables de interés)	17
5.2	Entendimiento de los datos (Metodología KDD)	18
5.3	Diseño e implementación del piloto	18
5.4	Evaluación de resultados del piloto	19
5.5	Planteamiento de investigaciones futuras	19
6	FLUJO METODOLOGICO	19
7	ALCANCES	24
8	RESULTADOS ESPERADOS	25
9	DESARROLLO DE LA METODOLOGIA	25
9.1	Desarrollo metodológico 2018	26
9.1.1	Construcción base datos transferencias	26
9.1.2	Construcción métricas relaciones	28
9.1.2.1	Análisis exploratorio métricas	28
9.1.3	Limpieza de datos	31
9.1.4	Modelos machine learning	31
9.1.4.1	Campaña referidos 2018	31
9.1.4.2	Variable dependiente	32
9.1.4.3	Métodos de balanceo	33
9.1.5	Implementación modelos	34
9.1.6	Modelamiento redes	35
9.1.6.1	Filtro de red	35

9.1.7	Métricas redes	36
9.1.7.1	Análisis exploratorio métricas	37
9.1.8	Análisis estadístico	40
9.1.9	Análisis No Supervisado	42
9.1.9.1	Análisis no supervisado variables normalizadas	42
9.1.9.2	Análisis no supervisado variables percentiles	45
9.2	Desarrollo metodológico 2019	48
9.2.1	Construcción base datos relaciones	48
9.2.2	Construcción métricas relaciones	49
9.2.2.1	Análisis exploratorio métricas	49
9.2.3	Limpieza de datos	51
9.2.4	Modelamiento redes	51
9.2.4.1	Filtro de red	51
9.2.5	Métricas redes	52
9.2.6	Construcción dimensión enlace	53
9.2.6.1	Análisis e implementación modelos machine learning	53
9.3	Construcción relevancia cliente	55
9.4	Piloto campaña referidos	55
9.4.1	Diseño experimental	55
9.4.1.1	Clientes	55
9.4.1.2	Gancho	56
9.4.1.3	Contactos	56
9.4.1.4	Formulario	56
9.4.1.5	Graficas campaña	57
9.4.1.6	Funnel campaña	58
9.5	Análisis estadístico	60
9.5.1	Análisis correlación	61
9.5.2	Regresiones logísticas	61
9.6	Análisis no supervisado	63
10	RESULTADOS	67
10.1.1	Efecto incremental	67
10.1.2	Ajuste proporcional	68
10.2	Caracterización clientes	70
11	JUSTIFICACIÓN ECONÓMICA	70
12	CONCLUSIONES	72
13	INVESTIGACIONES FUTURAS	73

13.1	Relacionadas directamente al proyecto	73
13.2	Relacionadas al estudio de redes	74
14	BIBLIOGRAFÍA	75
15	ANEXOS	76
15.1	Anexo sección 9.1.5	76
15.2	Anexo sección 9.1.6.1	77
15.3	Anexos sección 9.1.9	77
15.4	Anexos sección 9.2.1	78
15.5	Anexos 9.2.2.1	79
15.6	Anexo sección 9.2.4.1	80
15.7	Anexo sección 9.2.5	81
15.8	Anexo sección 11	81

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Tarjetas de crédito activas 2018	1
Ilustración 2: Participación de mercado bancario en Chile año 2018	2
Ilustración 3: Tarjetas de crédito activas	3
Ilustración 4: Estructura Organizacional Gerencia Business Intelligence	4
Ilustración 5: Método de balanceo Undersampling	14
Ilustración 6: Método de balanceo Oversampling	15
Ilustración 7: Método de balanceo SMOTE	15
Ilustración 8: Macroproceso del proyecto	20
Ilustración 9: Ejemplo tabla transferencia	20
Ilustración 10: Ejemplo tabla de relaciones	21
Ilustración 11: Ejemplo tabla relaciones únicas con variable dependiente	21
Ilustración 12: Ejemplo entrenamiento modelo machine learning	22
Ilustración 13: Ejemplo clasificación modelo machine learning	22
Ilustración 14: Ejemplo resultado modelo machine learning	23
Ilustración 15: Distribución del monto de transferencias 2017	27
Ilustración 16: Distribución relaciones únicas de transferencias 2017	27
Ilustración 17: Distribución variable Delta	29
Ilustración 18: Boxplot variable Delta	29
Ilustración 19: Distribución variable Frequency	29
Ilustración 20: Boxplot variable Frequency	29
Ilustración 21: Distribución variable Monto	30
Ilustración 22: Boxplot variable Monto	30
Ilustración 23: Distribución variable Recency	30
Ilustración 24: Boxplot variable Recency	30
Ilustración 25: Distribución variable RF	30
Ilustración 26: Boxplot variable RF	30
Ilustración 27: Distribución variable dependiente modelo enlaces	33
Ilustración 28: Iteración filtro de redes 2018	36
Ilustración 29: Boxplot Betweenness	38
Ilustración 30: Boxplot Centrality	38
Ilustración 31: Boxplot Eccentricity	38
Ilustración 32: Boxplot Eigenvector	38
Ilustración 33: Curva ROC Regresión logística datos normalizados 2018	41
Ilustración 34: Curva ROC Regresión logística datos percentiles 2018	41
Ilustración 35: Elbow curve datos normalizados 2018	42
Ilustración 36: Calinski Harabasz datos normalizados 2018	42
Ilustración 37: Scatterplot Eigenvector-Betweenness	43
Ilustración 38: Scatterplot Centrality-Betweenness	43
Ilustración 39: Scatterplot Centrality-Eigenvector	43
Ilustración 40: Scatterplot Eccentricity-Eigenvector	44
Ilustración 41: Scatterplot Eccentricity-Betweenness	44
Ilustración 42: Scatterplot Eccentricity-Centrality	44
Ilustración 43: Scatterplot 3D variables (E-B-C)	44
Ilustración 44: Scatterplot 3D variables (E-B-C)	44
Ilustración 45: Scatterplot 3D variables (E-B-C) solo clientes refieren	44

Ilustración 46: Elbow curve datos percentiles 2018.....	45
Ilustración 47: Calinski Harabasz datos percentiles 2018	45
Ilustración 48: Scatterplot Betweenness-Eigenvector	46
Ilustración 49: Scatterplot Centrality-Betweenness	46
Ilustración 50: Scatterplot Centrality-Eigenvector.....	46
Ilustración 51: Scatterplot Eccentricity-Eigenvector.....	46
Ilustración 52: Scatterplot Eccentricity-Betweenness	46
Ilustración 53: Scatterplot Eccentricity-Centrality	46
Ilustración 54: Scatterplot 3D variables E-B-C	47
Ilustración 55: Scatterplot 3D variables E-B-C (Solo clientes refieren).....	47
Ilustración 56: Caracterización relaciones 2018/2019	48
Ilustración 57: Distribución variable Frequency	50
Ilustración 58: Boxplot variable Frequency.....	50
Ilustración 59: Distribución variable Monto	50
Ilustración 60: Boxplot variable Monto.....	50
Ilustración 61: Distribución variable RF	50
Ilustración 62: Boxplot variable RF.....	50
Ilustración 63: Iteración filtro de redes 2019.....	52
Ilustración 64: Distribución Betweenness redes 2019	52
Ilustración 65: Distribución Centrality redes 2019	52
Ilustración 66: Distribución Eccentricity redes 2019	53
Ilustración 67: Distribución Eigenvector redes 2019.....	53
Ilustración 68: Comparación distribución normal con distribución Logistic Regression..	54
Ilustración 69: Comparación distribución normal con distribución Logistic Regression Balanced	54
Ilustración 70: Comparación distribución normal con distribución Gradient Boosting	54
Ilustración 71: Comparación distribución normal con distribución Gradient Boosting Balanced	54
Ilustración 72: Formulario referidos campaña 2019	57
Ilustración 73: Graficas email campaña 2019.....	57
Ilustración 74: Grafica email campaña 2019	58
Ilustración 75: Curva ROC Regresión logística campaña 2019.....	62
Ilustración 76: Curva ROC Regresión Logística campaña 2019 (sin variables Enganchado).....	63
Ilustración 77: Elbow curve datos campaña 2019	64
Ilustración 78: Calinski Harabasz curve datos campaña 2019	64
Ilustración 79: Agrupación clúster campaña 2019.....	64
Ilustración 80: Análisis de dispersión variables de relevancia clientes campaña 2019 ..	65
Ilustración 81: Curva ROC Modelo Logistic Regression.....	76
Ilustración 82: Curva ROC Modelo Logistic regression balanced.....	76
Ilustración 83: Curva ROC Modelo Random Forest	76
Ilustración 84: Curva ROC Modelo Random Forest balanced.....	76
Ilustración 85: Curva ROC Modelo Gradient Boosting	76
Ilustración 86: Curva ROC Modelo Gradient Boosting Balanced	76
Ilustración 87: Curva ROC Modelo Ada Boost	77
Ilustración 88: Curva ROC Modelo Ada Boost Balanced	77
Ilustración 89: Distribución monto transferencias 2018/2019	78
Ilustración 90: Distribución relaciones únicas transferencias 2018/2019	78
Ilustración 91: Distribución variable Delta.....	79

Ilustración 92: Boxplot variable Delta	79
Ilustración 93: Distribución variable Frequency	79
Ilustración 94: Boxplot variable Frequency	79
Ilustración 95: Distribución variable Monto	79
Ilustración 96: Boxplot variable Monto.....	79
Ilustración 97: Distribución variable Recency	80
Ilustración 98: Boxplot variable Recency	80
Ilustración 99: Distribución variable RF	80
Ilustración 100: Boxplot variable RF	80

ÍNDICE DE TABLAS

Tabla 1: Caracterización relaciones 2017/2018	26
Tabla 2: Descripción monto transferencias 2017	27
Tabla 3: Descripción relaciones únicas 2017	27
Tabla 4: Descriptivo tabla relaciones 2017/2018.....	29
Tabla 5: Filtro outliers base relaciones 2017-2018.....	31
Tabla 6: Funnel campaña referidos 2018.....	32
Tabla 7: Métricas de desempeño regresión logística para distintos métodos de balanceo	33
Tabla 8: Métricas de desempeño modelos Machine Learning datos 2017/2018.....	34
Tabla 9: Descriptivo métricas red 2017/2018	37
Tabla 10: Correlación métricas redes 2017/2018.....	37
Tabla 11: Filtro outliers métricas red 2017/2018	37
Tabla 12: Correlación transformaciones de variables	39
Tabla 13: Correlación transformación percentiles	40
Tabla 14: Regresión logística datos normalizados 2018	41
Tabla 15: Regresión logística datos percentiles 2018	41
Tabla 16: Agrupación clúster análisis no supervisado datos normalizados 2018.....	42
Tabla 17: Test de diferencia en proporciones datos normalizados 2018	43
Tabla 18: Agrupación clúster análisis no supervisado datos percentiles 2018.....	45
Tabla 19: Test de diferencia en proporciones datos percentiles 2018	46
Tabla 20: Descripción monto transferencias 2018/2019	49
Tabla 21: Descripción relaciones únicas transferencias 2018/2019.....	49
Tabla 22: Descriptivo tabla relaciones 2018/2019.....	49
Tabla 23: Filtro outliers base relaciones 2018/2019	51
Tabla 24: Caracterización clientes analizados piloto.....	56
Tabla 25: Funnel campaña referidos 2019.....	58
Tabla 26: Clientes que contactados por categoría	59
Tabla 27: Clientes que se analizan por categoría	59
Tabla 28: Clientes que refieren por categoría	59
Tabla 29: Tasa de respuesta por categoría.....	60
Tabla 30: Correlación variable dependiente campaña referidos 2019	61
Tabla 31: Regresión logística campaña 2019	62

Tabla 32: Regresión logística campaña 2019 (Sin variables Enganchado)	63
Tabla 33: Test diferencia en proporciones clúster campaña 2019	65
Tabla 34: Mapa calor tasa respuesta	66
Tabla 35: Mapa calor número clientes	66
Tabla 36: Tasa de respuesta segmentos campaña.....	67
Tabla 37: Test de diferencia en proporciones clúster y grupo de control	67
Tabla 38: Distribución productos contratados clientes segmentados por origen.....	68
Tabla 39: Distribución ajustada de productos contratados segmentados por origen	69
Tabla 40: Tasa de respuesta clúster ajustados	69
Tabla 41: Test de diferencia en proporciones con segmentos ajustados	69
Tabla 42: Descripción monto transferencias 2018/2019	78
Tabla 43: Descripción relaciones únicas transferencias 2018/2019.....	78

ÍNDICE DE FÓRMULAS

Ecuación 1: Tasa de lectura correos	12
Ecuación 2: Tasa de respuesta correos	12
Ecuación 3: Hipótesis test de diferencia en proporciones	13
Ecuación 4: Valor p ecuación diferencia en proporciones	13
Ecuación 5: Valor z ecuación de diferencia en proporciones	14
Ecuación 6: Definición de variable dependiente Refiere	21
Ecuación 7: Definición de variable dependiente Refiere 2018	32
Ecuación 8: Definición de variable dependiente Refiere 2019	60