

TABLA DE CONTENIDO

1.- Introducción	1
2.- Descripción del estudio	3
2.1 Exposición del tema	3
2.2 Objetivos	5
3.- Marco teórico	7
3.1 Proceso KDD	7
3.2 Riesgo de créditos	8
3.2.3. Puntaje de comportamiento	8
3.3 Regresión Logística	12
3.4 Árboles de decisión	13
3.4.1 Método CHAID	14
3.5 Índice de estabilidad de la población	14
3.6 Conceptos Relevantes	15
3.6.1 Deuda	15
3.6.2 Probabilidad de Default	15
3.6.3 Clientes Won't pay	15
3.6.4 Clientes Can't Pay	15
4.- Metodología	16
4.1 Datos	16
4.2 Selección de los Datos	16
4.2.1 Verificación de los datos	17
4.2.2 Composición de los clientes	17
4.3 Pre-procesamiento de Datos	18
4.4. Transformación de datos	19
4.5 Análisis descriptivo	19
4.5.1 Clientes buenos	22
4.5.2 Clientes malos	27
5. Discusión	31
5.1 Riesgo según características de los clientes	31
5.1.1 Variable de la edad de los clientes	32
5.1.2 Variable del saldo de los clientes	34

5.1.3 Variable antigüedad de los clientes en el banco	37
5.1.4 Variable del sexo de los clientes	39
6. Resultados	42
6.1 Base de datos	42
6.2 Variables seleccionadas	43
6.2.1 Variable 1: Tendencia de los días de mora de un cliente	43
6.2.2 Variable 2: Grupo socioeconómico de los clientes	44
6.2.3 Variable 3: Edad	45
6.3 Resultados del modelo	47
7. Conclusiones y recomendaciones	56
Bibliografía	59
Anexos	61
Anexo A: Índice PSI de los clientes buenos	61
Anexo B: Descripción de categorías de variable producto más riesgoso de la cartera del cliente	65
Anexo C: Índice PSI de los clientes malos	66
Anexo D: Gráficas de riesgo para la variable subsegmento de clientes y producto más riesgoso de la cartera	70
Anexo E: Análisis de resultados de variables del modelo	74
Anexo F: Resultados del modelo, puntaje de todos los datos	85
Anexo G: Resultados finales	98