

INDICE DE CONTENIDOS

Capítulo 1 Introducción	1
1.1 Contextualización	2
1.2 Problemática.....	4
1.3 Objetivos.....	5
1.3.1 Objetivo general	5
1.3.2 Objetivos específicos.....	5
1.4 Alcance del trabajo	6
1.5 Metodología	6
1.6 Resultados esperados.....	7
1.7 Estructura de la tesis	8
Capítulo 2 Aprendizaje Automático	10
2.1 Descubrimiento de conocimiento en base de datos	11
2.2 Proceso KDD	12
2.2.1 Selección de los datos.....	12
2.2.2 Preprocesamiento de los datos	13
2.2.3 Transformación de los datos	14
2.2.4 Reducción.....	15
2.2.5 Aplicación de minería de datos	15
2.2.6 Interpretación y evaluación	16
Capítulo 3 Riesgo Crediticio y Loss Given Default	18
3.1 Conceptos básicos	19
3.1.1 Comité de Basilea para supervisión bancaria	19
3.1.2 Incumplimiento	21
3.1.3 Pérdida	22
3.2 Riesgo crediticio	22
3.3 Pérdida dado el default.....	23
3.4 Estudios realizados para la LGD	24
3.4.1 Estado del arte LGD	24
3.4.2 Estudios realizados sobre la medición de la LGD.....	26
3.4.3 Modelos a utilizar para comparación.....	27
Capítulo 4 Redes Neuronales, Support Vector Regression y Ensembles	29
4.1 Redes neuronales	30
4.2 Support vector regression	32
4.2.1 Concepto de aprendizaje estadístico	32

4.2.2	Hiperplanos separadores	34
4.2.3	Caso linealmente separable	35
4.2.4	Caso no linealmente separable	38
4.2.5	Support vector regression	41
4.3	Ensembles	43
4.3.1	Tipos de estrategias de combinación	45
Capítulo 5 Metodologías propuestas para la medición de la LGD		47
5.1	Aporte de la tesis	48
5.2	SVR	48
5.3	SVR con programación de kernel	49
5.3.1	Programación de kernel intervalar	49
5.3.2	Distribución LGD	50
5.3.3	Dataset	51
5.3.4	Pruebas de SVR	52
5.3.5	Resultados de estimaciones	55
5.3.6	Resolución de propuesta de modelamiento	58
5.4	Ensembles	58
Capítulo 6 Experimentación y Análisis de Resultados		61
6.1	Créditos de consumo	62
6.2	Metodología KDD	62
6.2.1	Selección y consolidación de datos	62
6.2.2	Preprocesamiento	64
6.2.3	Transformación	65
6.3	Consideraciones importantes	67
6.3.1	Créditos a largo plazo	67
6.3.2	Créditos a corto plazo	69
6.4	Características y configuración de las muestras	71
6.4.1	Base de datos corto plazo	71
6.4.2	Base de datos largo plazo	72
6.5	Resultados de modelos de la literatura	73
6.5.1	Regresión de respuesta fraccional	73
6.5.2	Árbol de regresión	75
6.5.3	Red neuronal	78
6.6	Resultados de modelos propuestos	82
6.6.1	Support vector regression (SVR)	82
6.6.2	Metodologías ensembles	84

6.7	Resultados modelos realizados	88
6.7.1	SVR con programación de kernel intervalar	88
6.7.2	Ensemble propuesto.....	89
6.8	Validaciones cruzadas.....	95
Capítulo 7 Conclusiones y trabajos futuros		98
7.1	Justificación de trabajo realizado	99
7.2	Modelos de minería de datos	99
7.3	Desarrollo de objetivos planteados	101
7.4	Discusión y trabajo a futuro	101
Bibliografía.....		103
Anexos		108

INDICE DE FIGURAS

Figura 2-1: Etapas del proceso KDD	12
Figura 3-1: Distribución de la LGD.....	24
Figura 4-1: Estructura de una red neuronal	31
Figura 4-2: Componentes de una red neuronal	31
Figura 4-3: Hiperplano separador $\{w,b\}$ para conjunto de entrenamiento	34
Figura 4-4: Caso linealmente separable	35
Figura 4-5: Hiperplano con vectores soportantes	37
Figura 4-6: Caso no linealmente separable	38
Figura 4-7: Ejemplificación funciones de kernel	39
Figura 4-8: Modelo SVR gráficamente.....	42
Figura 4-9: Motivo estadístico para la generación de ensembles	43
Figura 4-10: Motivo computacional para la generación de ensembles	44
Figura 4-11: Motivo representacional para la generación de ensembles	44
Figura 4-12: Representación regla fija de votación	45
Figura 5-1: Ejemplo de separación por intervalos	49
Figura 5-2: Gráfico dataset	52
Figura 5-3: Arquitectura de la metodología propuesta.....	59
Figura 5-4: Metodología propuesta	59
Figura 6-1: Variable de edad gráficamente.....	66
Figura 6-2: Variable de región gráficamente.....	66
Figura 6-3: LGD de corto plazo.....	72
Figura 6-4: LGD de largo plazo.....	72
Figura 6-5: Pruebas de árbol de regresión corto plazo.....	76
Figura 6-6: Pruebas de árbol de regresión largo plazo.....	77
Figura 6-7: Pruebas de red neuronal corto plazo	80
Figura 6-8: Pruebas de red neuronal largo plazo	81
Figura 6-9: Resultados de validación cruzada corto plazo	96
Figura 6-10: Resultados de validación cruzada largo plazo	97

INDICE DE TABLAS

Tabla 3-1: Principales modelos avanzados de riesgo crediticio	23
Tabla 3-2: Modelos realizados para la LGD	25
Tabla 5-1: Modelos con límites LGD	51
Tabla 5-2: Parámetros SVR kernel lineal	55
Tabla 5-3: Resultados kernel lineal	55
Tabla 5-4: Resultados kernel polinomial	56
Tabla 5-5: Parámetros SVR kernel rbf	56
Tabla 5-6: Resultados kernel rbf	56
Tabla 5-7: Kernels intervalares	57
Tabla 5-8: Parámetros SVR kernel intervalar	57
Tabla 5-9: Resultados kernel intervalar	57
Tabla 5-10: Validaciones cruzadas kernels	58
Tabla 5-11: Ejemplo de agrupación predicciones	60
Tabla 6-1: Características Generales de la Base de Datos	63
Tabla 6-2: Agrupación de Variables de la Base de Datos	63
Tabla 6-3: Resumen transformaciones realizadas	67
Tabla 6-4: Preprocesamiento largo plazo	68
Tabla 6-5: Variables macroeconómicas largo plazo	69
Tabla 6-6: Preprocesamiento corto plazo	70
Tabla 6-7: Variables Macroeconómicas Largo Plazo	71
Tabla 6-8: Regresión de respuesta fraccional corto plazo	73
Tabla 6-9: Regresión de respuesta fraccional largo plazo	74
Tabla 6-10: Parámetros árbol de regresión	75
Tabla 6-11: Resumen resultados árbol de regresión corto plazo	76
Tabla 6-12: Resumen resultados árbol de regresión largo plazo	77
Tabla 6-13: Parámetros red neuronal	78
Tabla 6-14: Resumen resultados entrenamiento red neuronal corto plazo	79
Tabla 6-15: Resumen resultados testeo corto plazo red neuronal	79
Tabla 6-16: Resumen resultados entrenamiento red neuronal largo plazo	80
Tabla 6-17: Resumen resultados testeo red neuronal largo plazo	81
Tabla 6-18: Parámetros SVR	82
Tabla 6-19: Resumen resultados SVR corto plazo	83
Tabla 6-20: Resumen resultados SVR largo plazo	84
Tabla 6-21: Parámetros para ensemble votación	85
Tabla 6-22: Resumen resultados búsqueda en grilla votación corto plazo	85
Tabla 6-23: Resumen resultados búsqueda en grilla votación largo plazo	86
Tabla 6-24: Resumen resultados random forest corto plazo	87
Tabla 6-25: Resumen resultados random forest largo plazo	87
Tabla 6-26: Parámetros SVR con programación de kernel intervalar	88
Tabla 6-27: Resumen resultados SVR kernel intervalar corto plazo	89
Tabla 6-28: Resumen Resultados SVR kernel intervalar largo plazo	89
Tabla 6-29: Parámetros ensemble propuesto red neuronal	90
Tabla 6-30: Parámetros ensemble propuesto SVR	90
Tabla 6-31: Resultados entrenamiento ensemble propuesto corto plazo	91
Tabla 6-32: Parámetros redes neuronales ensemble propuesto corto plazo	91
Tabla 6-33: Parámetros SVR ensemble propuesto corto plazo	91
Tabla 6-34: Parámetros SVR ensemble propuesto corto plazo	92

Tabla 6-35: Resultados ensemble propuesto corto plazo.....	92
Tabla 6-36: Resultados entrenamiento ensemble propuesto largo plazo	93
Tabla 6-37: Parámetros redes neuronales ensemble propuesto largo plazo	93
Tabla 6-38: Parámetros SVR ensemble largo plazo.....	93
Tabla 6-39: Parámetros SVR ensemble propuesto largo plazo	94
Tabla 6-40: Resultados ensemble propuesto largo plazo.....	94
Tabla 6-41: Validaciones cruzadas corto plazo	95
Tabla 6-42: Validaciones cruzadas largo plazo	96