

TABLA DE CONTENIDOS

	página
Dedicatoria	I
Agradecimientos	II
Tabla de Contenidos	III
Índice de Figuras	VI
Índice de Tablas	VII
Resumen	VIII
1. Introducción	10
1.1. Introducción	10
1.1.1. Objetivo General	10
1.1.2. Objetivos Específicos	10
1.1.3. Alcances	10
1.2. Estructura de la Memoria	11
2. Antecedentes	13
2.1. Introducción	13
2.2. Reconocimiento de Patrones	13
2.3. <i>Machine Learning</i>	14
2.3.1. Aprendizaje supervisado	14
2.3.2. Clasificación	15
2.3.3. Algoritmo <i>k-nn</i>	15
2.3.4. Aprendizaje No Supervisado	16
2.3.5. <i>Clustering</i>	16
2.3.6. Aprendizaje Semisupervisado	17
2.4. Medidas de Similitud	17
2.4.1. Distancia Euclídeana	17
2.4.2. Distancia de Hamming	18

2.4.3.	Distancia Manhattan	18
2.5.	Medidas de Desempeño	18
2.5.1.	<i>Accuracy</i>	19
2.5.2.	<i>True Positive Rate</i>	20
2.5.3.	<i>False Positive Rate</i>	20
2.5.4.	<i>Specificity</i>	20
2.5.5.	Precisión	20
2.6.	Modelos de Validación	21
2.6.1.	Validación <i>Holdout</i>	21
2.6.2.	Validación Cruzada de k pasos	21
2.6.3.	Validación Cruzada leave-one-out	22
2.7.	Diseño Experimental	22
2.7.1.	Best Guest	22
2.7.2.	One Factor At a Time	22
2.7.3.	Diseño Factorial	23
3.	Metodología de Desarrollo de Software	24
3.1.	Desarrollo guiado por pruebas	24
3.2.	Java	25
3.3.	Eclipse	25
3.4.	Weka	26
3.5.	JUnit	26
4.	Diseño e Implementación	27
4.1.	Componentes usados en la Construcción del Algoritmo	27
4.1.1.	Trazado de Simetrales	27
4.1.2.	Entropía de Shannon	30
4.1.3.	Firmas	33
4.1.4.	Aplicación de la Distancia de Hamming a las Firmas	33
4.1.5.	Firma Activa	33
4.2.	RHC	33
4.2.1.	BuildClassifier	34
4.2.2.	Método <code>classifyInstance</code>	36

5. Resultados Experimentales	39
5.1. Introducción	39
5.2. Metodología Experimental	39
5.2.1. Medidas de Desempeño	39
5.2.2. Diseño Factorial	39
5.3. Conjuntos de Datos	40
5.4. Resultados	41
6. Conclusiones y trabajo futuro	47
6.1. Conclusiones	47
6.2. Trabajo Futuro	47
Bibliografía	49

ÍNDICE DE FIGURAS

	página
2.1. Ilustración de k -nn para $k= 1$ y $k = 5$	16
2.2. Matriz de Confusión	19
4.1. Representación vectorial de Iris para dos de sus cuatro atributos . . .	28
4.2. Dos maneras distintas de generar un hiperplano	29
4.3. Segmentación del espacio vectorial de para Iris	31
4.5. Entropía(b)	32

ÍNDICE DE TABLAS

	página
5.1. Conjunto de datos utilizados para evaluar clasificadores.	40
5.2. Desempeño de RHC para el Conjunto de datos Iris	42
5.3. Desempeños de RHC para el conjunto de datos Glass	43
5.4. Desempeños de RHC para el conjunto de datos Wine	44
5.5. Desempeños de RHC para el conjunto de datos Wine_Quality (red) .	44
5.6. Desempeños de RHC para el conjunto de datos WBDC)	45
5.7. Desempeños de RHC para el conjunto de datos Yeast)	46
5.8. Accuracy para RHC y otros algoritmos de clasificación	46