



**UNIVERSIDAD TÉCNICA DEL NORTE**

**FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS**

**CARRERA DE INGENIERÍA EN ELECTRÓNICA Y REDES DE COMUNICACIÓN**

**TRABAJO DE GRADO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERÍA  
EN ELECTRÓNICA Y REDES DE COMUNICACIÓN**

**TEMA:**

**“COMPARACIÓN DE MÉTODOS DE REDUCCIÓN DE DIMENSIONALIDAD  
ENFOCADOS A ALGORITMOS DE CLASIFICACIÓN SUPERVISADOS APLICADO  
A DATOS DE REDES DE SENSORES”**

**AUTOR: MAYRA ALEXANDRA VALENCIA GUERRÓN**

**DIRECTOR: MSC. PAÚL DAVID ROSERO MONTALVO**

**Ibarra-Ecuador**

**2019**



**UNIVERSIDAD TÉCNICA DEL NORTE**  
**BIBLIOTECA UNIVERSITARIA**

**AUTORIZACIÓN DE USO Y PUBLICACIÓN**

**A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE.**

**1. IDENTIFICACIÓN DE LA OBRA**

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago la entrega del presente trabajo a la Universidad Técnica del Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a disposición la siguiente información:

<b>DATOS DE CONTACTO</b>			
<b>CÉDULA DE IDENTIDAD:</b>	1004183412		
<b>APELLIDOS Y NOMBRES:</b>	Valencia Guerrón Mayra Alexandra		
<b>DIRECCIÓN:</b>	Los Ceibos		
<b>EMAIL:</b>	mavalenciag@utn.edu.ec		
<b>TELÉFONO FIJO:</b>	2511545	<b>TELÉFONO MÓVIL:</b>	0989964059

<b>DATOS DE LA OBRA</b>	
<b>TÍTULO:</b>	“Comparación de métodos de reducción de dimensionalidad enfocados a algoritmos de clasificación supervisados aplicado a datos de redes de sensores”.
<b>AUTOR:</b>	Valencia Guerrón Mayra Alexandra
<b>FECHA:</b>	27 de noviembre de 2019
<b>SOLO PARA TRABAJOS DE GRADO</b>	
<b>PROGRAMA:</b>	<input checked="" type="checkbox"/> <b>PREGRADO</b> <input type="checkbox"/> <b>POSGRADO</b>
<b>TÍTULO POR EL QUE OPTA:</b>	Ingeniero en Electrónica y Redes de comunicación
<b>ASESOR /DIRECTOR:</b>	Ing. Paúl David Rosero Montalvo

## 2. CONSTANCIAS

El autor manifiesta que la obra objeto de la presente autorización es original y se la desarrolló, sin violar derechos de autor de terceros, por lo tanto, la obra es original y que es el titular de los derechos patrimoniales, por lo que asume la responsabilidad sobre el contenido de la misma y saldrá en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, a los 27 días del mes de noviembre de 2019.

**EL AUTOR:**



Mayra Alexandra Valencia Guerrón

CC: 1004183412



**UNIVERSIDAD TÉCNICA DEL NORTE**

**FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS**

**CERTIFICACIÓN.**

MAGISTER PAÚL ROSERO, DIRECTOR DEL PRESENTE TRABAJO DE TITULACIÓN CERTIFICA:

Que, el presente trabajo de Titulación “COMPARACIÓN DE MÉTODOS DE REDUCCIÓN DE DIMENSIONALIDAD ENFOCADOS A ALGORITMOS DE CLASIFICACIÓN SUPERVISADOS APLICADO A DATOS DE REDES DE SENSORES” Ha sido desarrollado por la señorita Valencia Guerrón Mayra Alexandra bajo mi supervisión.

Es todo en cuanto puedo certificar en honor de la verdad.

A handwritten signature in blue ink, appearing to read "Paúl Rosero", is written over a horizontal dotted line. The signature is enclosed within a circular scribble.

MSc. Paúl David Rosero Montalvo

1003378039

**DIRECTOR**

**DEDICATORIA**

*Fuiste sin dudar mi mayor motivación para culminar con éxito este trabajo de titulación, por eso quiero dedicarte este y cada esfuerzo realizado a lo largo de mi carrera, con todo mi amor, para ti hijo mío.*

## AGRADECIMIENTO

*Quiero agradecer a mis padres Javier y Mariela por su amor, dedicación y apoyo en cada uno de mis pasos, han sido un gran ejemplo y recordatorio diario de las recompensas que trae el esfuerzo. A mi hermana Johana quien ha sido una constante amiga y parte importante de lo que soy. Anderson por creer en mí y respaldarme siempre que lo necesité no solo para el desarrollo de mi tesis, sino también en mi vida.*

*A mis compañeros y amigos quienes estuvieron presentes durante toda o parte de la duración de mi carrera por su compañía y apoyo brindado.*

*A mis opositores MSc. Ana Umaquina y MSc. Luís Suarez, de manera especial agradezco a mi director de tesis el MSc. Paúl Rosero por cada enseñanza, detalle y momento dedicado para la realización de este proyecto.*

**ÍNDICE.**

AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE. ....	I
CERTIFICACIÓN. ....	III
DEDICATORIA .....	IV
AGRADECIMIENTO .....	V
ÍNDICE .....	VI
ÍNDICE DE FIGURAS.....	X
ÍNDICE DE TABLAS. ....	XVI
RESUMEN .....	XVII
ABSTRACT.....	XVIII
CAPÍTULO I. ANTECEDENTES .....	1
1.1 Tema.....	1
1.2 Problema .....	1
1.3 Objetivos .....	3
1.3.1 Objetivo general. ....	3
1.3.2 Objetivos específicos. ....	3
1.4 Alcance.....	3
1.5 Justificación.....	4
CAPÍTULO II. REVISIÓN BIBLIOGRÁFICA.....	6
2.1 Sistemas Embebidos .....	6
2.1.1 Características de Sistemas Embebidos .....	6

2.1.2 Componentes principales .....	7
2.1.3 Aplicaciones.....	8
2.2 Redes de Sensores Inalámbricos (WSN) .....	8
2.2.1 Componentes de un nodo WSN .....	9
2.2.2 Funcionamiento.....	10
2.2.3 Topologías de red.....	10
2.3 Big Data .....	11
2.3.1 Introducción del Big Data .....	12
2.3.2 Aplicaciones del Big Data.....	12
2.3.3 Minería de Datos .....	12
2.4 Reducción de Dimensionalidad.....	13
2.4.1 Análisis de Componentes Principales (PCA).....	15
2.4.2 Laplacian Eigenmaps (LE).....	16
2.4.3 Locally Linear Embedding (LLE).....	17
2.4.4 Stochastic Neighbor Embedding (SNE).....	18
2.4.5 T-distributed Stochastic Neighbor Embedding ( $t$ -SNE).....	20
2.5 Algoritmos de clasificación supervisada.....	22
2.5.1 k-nearest neighbors (k-NN).....	22
2.5.2 Support Vector Machines (SVM) .....	23
2.5.3 Clasificador Bayesiano.....	23
2.5.4 Árbol de decisión .....	24



CAPÍTULO III. DESARROLLO EXPERIMENTAL.....	25
3.1 Descripción General.....	25
3.2 Adquisición de Bases de Datos.....	26
3.2.1 Datos de posición del cuerpo con sensor acelerómetro. ....	27
3.2.2 Conjunto de datos sensores de silla de ruedas .....	28
3.2.3 Base de datos tipos de pisadas .....	29
3.2.4 Reconocimiento de patrones de datos en rosas.....	30
3.2.5 Detección de alcohol en conductores por sensores y visión por ordenador.....	32
3.3 Requerimientos .....	33
3.3.1 Elección de plataforma para gráficos de RD.....	34
3.3.2 Limpieza y acondicionamiento de datos .....	35
3.4 Reducción de dimensionalidad .....	37
3.4.1 PCA .....	38
3.4.2 LE.....	40
3.4.3 LLE .....	43
3.4.4 SNE .....	46
3.4.5 T-SNE .....	48
3.5 Clasificación.....	50
3.5.1 k-NN.....	51
3.5.2 SVM.....	52
3.5.3 Clasificador Bayesiano.....	54

3.5.4 Árbol de decisión .....	56
3.6 Matriz de Confusión.....	57
CAPÍTULO IV. PRUEBAS DE FUNCIONAMIENTO Y RESULTADOS .....	60
4.1 Pruebas con métodos de RD. ....	60
4.2 Pruebas con algoritmos de clasificación. ....	66
4.3 Resultados .....	92
CAPITULO V. CONCLUSIONES Y RECOMENDACIONES .....	96
5.1 CONCLUSIONES .....	96
5.2 RECOMENDACIONES .....	98
BIBLIOGRAFÍA .....	99
GLOSARIO DE TÉRMINOS Y ACRÓNIMOS.....	107
ANEXOS .....	108
ANEXO 1. Bases de datos originales .....	108
ANEXO 2. Bases de datos con RD.....	113

## ÍNDICE DE FIGURAS

Figura 1 Componentes de un nodo WSN.....	9
Figura 2 Topologías de una WSN.....	11
Figura 3 Estructura esférica artificial en 3D. ....	14
Figura 4 Visualización de método PCA.....	15
Figura 5 Visualización de método LE.....	17
Figura 6 Visualización de método LLE .....	18
Figura 7 Visualización de método SNE .....	20
Figura 8 Visualización de método <i>t</i> -SNE.....	21
Figura 9 Métodos de RD aplicados a estructura esférica de 3D. ....	22
Figura 10 Descripción general de funcionamiento .....	25
Figura 11 Esquema de ubicación del sistema .....	27
Figura 12 Esquema de ubicación de los sensores .....	29
Figura 13 Esquema de ubicación de sensores en las plantillas .....	30
Figura 14 Esquema de conexión del sistema .....	31
Figura 15 Esquema de conexión del sistema .....	32
Figura 16 Datos originales de BDD4 .....	36
Figura 17 Datos transformados de BDD4.....	36
Figura 18 Diagrama de flujo de proceso de RD.....	37
Figura 19 Eigen valores y vectores de BDD4.....	38
Figura 20 Varianza acumulativa de componentes principales de BDD4.....	39
Figura 21 PCA aplicado en BDD4.....	39
Figura 22 Cálculo de valor óptimo de k en BDD4 .....	40

Figura 23 Valores de $\rho$ para cada $k$ vecino .....	41
Figura 24 Valores de las dos dimensiones resultantes de LE .....	42
Figura 25 LE aplicado en BDD4.....	42
Figura 26 Cálculo de valor óptimo de $k$ en BDD4 .....	43
Figura 27 Valores de $\rho$ para cada $k$ vecino .....	44
Figura 28 Cálculo desarrollado por LLE .....	45
Figura 29 Valores de las dos dimensiones resultantes de LLE.....	45
Figura 30 LLE aplicado en BDD4 .....	46
Figura 31 Valores de las dos dimensiones resultantes de SNE.....	47
Figura 32 SNE aplicado en BDD4.....	47
Figura 33 Valores de las dos dimensiones resultantes de $t$ SNE.....	49
Figura 34 $t$ SNE aplicado en BDD4 .....	49
Figura 35 Diagrama de flujo de proceso de implementación de A. de Clasificación. ....	50
Figura 36 $k$ -NN aplicado en BDD4 .....	51
Figura 37 Clasificador SVM con kernel polynomial en BDD4.....	53
Figura 38 Clasificador SVM con kernel lineal en BDD4 .....	53
Figura 39 Clasificador SVM con kernel sigmoid en BDD4 .....	54
Figura 40 Valores de la variable independiente.....	55
Figura 41 Clasificador Bayesiano en BDD4.....	55
Figura 42 Esquema de clasificación del entrenamiento de datos.....	56
Figura 43 Resultado de la predicción.....	57
Figura 44 Clasificador Árbol de decisión en BDD4 .....	57
Figura 45 Tabla de conteos realizados con clasificador SVM en BDD4.....	58

Figura 46 Matriz de confusión con clasificador SVM en BDD4.....	59
Figura 47 Representación de los métodos de reducción de dimensionalidad en BDD1...	61
Figura 48 Representación de los métodos de reducción de dimensionalidad en BDD2...	62
Figura 49 Representación de los métodos de reducción de dimensionalidad en BDD3...	63
Figura 50 Representación de los métodos de reducción de dimensionalidad en BDD4...	64
Figura 51 Representación de los métodos de reducción de dimensionalidad en BDD4...	65
Figura 52 Clasificación con PCA en BDD1 .....	67
Figura 53 Clasificación con LE en BDD1 .....	68
Figura 54 Clasificación con LLE en BDD1 .....	69
Figura 55 Clasificación con SNE en BDD1.....	70
Figura 56 Clasificación con <i>t</i> -SNE en BDD1.....	71
Figura 57 Clasificación con PCA en BDD2 .....	72
Figura 58 Clasificación con LE en BDD2 .....	73
Figura 59 Clasificación con LLE en BDD2.....	74
Figura 60 Clasificación con SNE en BDD2.....	75
Figura 61 Clasificación con <i>t</i> -SNE en BDD1.....	76
Figura 62 Clasificación con PCA en BDD3 .....	77
Figura 63 Clasificación con LE en BDD3 .....	78
Figura 64 Clasificación con LLE en BDD3.....	79
Figura 65 Clasificación con SNE en BDD3.....	80
Figura 66 Clasificación con <i>t</i> -SNE en BDD3.....	81
Figura 67 Clasificación con PCA en BDD4 .....	82
Figura 68 Clasificación con LE en BDD4 .....	83

Figura 69 Clasificación con LLE en BDD4.....	84
Figura 70 Clasificación con SNE en BDD4.....	85
Figura 71 Clasificación con <i>t</i> -SNE en BDD4.....	86
Figura 72 Clasificación con PCA en BDD5 .....	87
Figura 73 Clasificación con LE en BDD5 .....	88
Figura 74 Clasificación con LLE en BDD5.....	89
Figura 75 Clasificación con SNE en BDD5.....	90
Figura 76 Clasificación con <i>t</i> -SNE en BDD5.....	91
Figura 77 Primeros 20 registros originales de la BDD1 .....	108
Figura 78 Últimos registros originales de la BDD1 .....	108
Figura 79 Primeros 30 registros originales de BDD2.....	109
Figura 80 Últimos registros originales de BDD2.....	109
Figura 81 Primeros registros originales de BDD3 .....	110
Figura 82 Últimos registros originales de BDD3.....	110
Figura 83 Primeros 20 registros originales de BDD4 .....	111
Figura 84 Últimos registros originales de BDD4.....	111
Figura 85 Primeros 20 registros originales de BDD5 .....	112
Figura 86 Últimos registros originales de BDD5.....	112
Figura 87 Primeros 20 registros de BDD1 aplicados PCA.....	113
Figura 88 Últimos registros de BDD1 aplicados PCA .....	113
Figura 89 Primeros registros de BDD1 aplicados LE.....	114
Figura 90 Últimos registros de BDD1 aplicados LE .....	114
Figura 91 Primeros registros de BDD1 aplicados LLE .....	115

Figura 92 Últimos registros de BDD1 aplicados LLE .....	115
Figura 93 Primeros registros de BDD1 aplicados SNE .....	116
Figura 94 Últimos registros de BDD1 aplicados SNE.....	116
Figura 95 Primeros registros de BDD1 aplicados tSNE .....	117
Figura 96 Últimos registros de BDD1 aplicados tSNE.....	117
Figura 97 Primeros registros de BDD2 aplicados PCA .....	118
Figura 98 Últimos registros de BDD2 aplicados PCA .....	118
Figura 99 Primeros registros de BDD2 aplicados LE.....	119
Figura 100 Últimos registros de BDD2 aplicados LE .....	119
Figura 101 Primeros registros de BDD2 aplicados LLE .....	120
Figura 102 Últimos registros de BDD2 aplicados LLE .....	120
Figura 103 Primeros registros de BDD2 aplicados SNE .....	121
Figura 104 Últimos registros de BDD2 aplicados SNE.....	121
Figura 105 Primeros registros de BDD2 aplicados tSNE .....	122
Figura 106 Últimos registros de BDD2 aplicados tSNE.....	122
Figura 107 Primeros registros de BDD3 aplicados PCA .....	123
Figura 108 Últimos registros de BDD3 aplicados PCA .....	123
Figura 109 Primeros registros de BDD3 aplicados LE.....	124
Figura 110 Últimos registros de BDD3 aplicados LE .....	124
Figura 111 Primeros registros de BDD3 aplicados LLE .....	125
Figura 112 Últimos registros de BDD3 aplicados LLE .....	125
Figura 113 Primeros registros de BDD3 con SNE .....	126
Figura 114 Últimos registros de BDD3 con SNE .....	126

Figura 115 Primeros registros de BDD3 con tSNE .....	127
Figura 116 Últimos registros de BDD3 con tSNE .....	127
Figura 117 Primeros registros de BDD4 aplicados PCA .....	128
Figura 118 Últimos registros de BDD4 aplicados PCA .....	128
Figura 119 Primeros registros de BDD4 aplicados LE .....	129
Figura 120 Últimos registros de BDD4 aplicados LE .....	129
Figura 121 Primeros registros de BDD4 aplicados LLE .....	130
Figura 122 Últimos registros de BDD4 aplicados LLE .....	130
Figura 123 Primeros registros de BDD4 con SNE .....	131
Figura 124 Últimos registros de BDD4 con SNE .....	131
Figura 125 Primeros registros de BDD4 aplicados tSNE .....	132
Figura 126 Últimos registros de BDD4 aplicados tSNE .....	132
Figura 127 Primeros registros de BDD5 aplicados PCA .....	133
Figura 128 Últimos registros de BDD5 aplicados PCA .....	133
Figura 129 Primeros registros de BDD5 aplicados LE .....	134
Figura 130 Últimos registros de BDD5 aplicados LE .....	134
Figura 131 Primeros registros de BDD5 aplicados LLE .....	135
Figura 132 Últimos registros de BDD5 aplicados LLE .....	135
Figura 133 Primeros registros de BDD5 aplicados SNE .....	136
Figura 134 Últimos registros de BDD5 aplicados SNE .....	136
Figura 135 Primeros registros de BDD5 aplicados tSNE .....	137
Figura 136 Últimos registros de BDD5 aplicados tSNE .....	137



**ÍNDICE DE TABLAS.**

Tabla 1 Descripción general de las bases de datos. ....	26
Tabla 2 Asignación de etiquetas para la posición del cuerpo .....	28
Tabla 3 Asignación de etiquetas respectivas para posibles problemas de salud.....	29
Tabla 4 Asignación de etiquetas para cada tipo de pisada.....	30
Tabla 5 Asignación de etiquetas para cada acción necesaria en el sistema .....	31
Tabla 6 Asignación de etiquetas para cada clase .....	32
Tabla 7 Requerimientos de Software .....	33
Tabla 8 Elección de plataforma.....	35
Tabla 9 Tabla resumen de resultados de los métodos de RD.....	66
Tabla 10 Precisión obtenida de clasificadores en PCA.....	92
Tabla 11 Precisión obtenida de clasificadores en LE.....	93
Tabla 12 Precisión obtenida de clasificadores en LLE. ....	93
Tabla 13 Precisión obtenida de clasificadores en SNE.....	94
Tabla 14 Precisión obtenida de clasificadores en $t$ SNE.....	94

## RESUMEN

El presente proyecto muestra la aplicación de métodos de reducción de dimensionalidad y algoritmos de clasificación con datos obtenidos de cinco distintos sistemas embebidos que hacen uso de redes de sensores inalámbricos con el objetivo de realizar una representación gráfica de los resultados y conjuntamente determinar el más adecuado mediante el cálculo de su eficiencia.

Para su desarrollo se realiza el análisis de características de los sets de datos y la preparación de estos, que implica su limpieza y acondicionamiento. Además, un breve estudio para determinar la plataforma correcta que permita trabajar con los datos analizados previamente. Se implementa cada método de reducción de dimensionalidad seleccionado para este trabajo a cada set de datos y consecutivamente se aplica los algoritmos de clasificación a los datos adquiridos.

En la parte final como resultados relevantes, se observan los métodos de RD de forma gráfica al ser aplicados a las bases de datos especificadas con anterioridad. Con ello, se realiza una comparación de los datos obtenidos del cálculo de la matriz de confusión para determinar la eficiencia lograda por cada algoritmo dentro de cada método de reducción de dimensionalidad.

## ABSTRACT

In this project dimensionality reduction methods and classification algorithms with data obtained from five different embedded systems that make use of wireless sensor networks are applied; in order to make a graphical representation of the results and determine the most appropriate by calculating its efficiency.

The analysis of the characteristics of the data sets and their preparation is carried out, which implies their cleaning and conditioning. In addition, a brief study to determine the correct platform to work with the previously analyzed data was performed. Each method of dimensionality reduction selected for this work is implemented to each data set and consecutively the classification algorithms are applied to the acquired data.

Finally, as relevant results, RD methods are observed graphically when applied to the databases specified above. With this, a comparison is made of the data obtained from the calculation of the confusion matrix to determine the efficiency achieved by each algorithm within each method of dimensionality reduction.

## **CAPÍTULO I. ANTECEDENTES**

En este capítulo se dará a conocer las razones que impulsaron el desarrollo de este proyecto de titulación, se describirá la problemática que se desea resolver, objetivos tanto general como específicos, el alcance y la justificación que sustenta este proyecto.

### **1.1 Tema**

“COMPARACIÓN DE MÉTODOS DE REDUCCIÓN DE DIMENSIONALIDAD ENFOCADOS A ALGORITMOS DE CLASIFICACIÓN SUPERVISADOS APLICADO A DATOS DE REDES DE SENSORES”

### **1.2 Problema**

En la actualidad con el rápido desarrollo y gran aplicación de la tecnología informática, apareció una gran cantidad de información de datos no estructurados de alta dimensión, gran variedad y velocidad más conocidos como Big Data (Lakshen, Guma Abdulkhader; Vraneš & Janev, 2016). Una de las técnicas que se ha desarrollado es la minería de datos la cual consiste en extraer datos de un conjunto de mayor tamaño que puede estar almacenada en una base de datos BBDD. De esta manera, permite recopilar, procesar, clasificar y seleccionar los datos más representativos que brinden información verdadera y relevante. Esto se basa en el criterio de que todos los datos no son útiles por motivos como la redundancia entre estos, pueden ser datos mal leídos, etc.

Cabe recalcar, que resulta de manera complicada el poder comprender y analizar para las personas por sí solas grandes cantidades de datos, por su complejidad puede ser de más de tres dimensiones o variables. Además, el tiempo que le llevaría realizar el proceso de análisis sería muy grande, y más aún si no se tiene experticia en el tema. Por lo cual, es importante encontrar un método que brinde una mejor percepción de estos (Guerrero, 2016).

La exuberancia de dimensionalidad puede desembocar en una inconsistencia y confusión al momento de implementarlas en diversas aplicaciones, además, se requiere un mayor tiempo de procesamiento para que los datos sean analizados. Por otra parte, al tener redundancia de datos nos enfrentamos a un espacio de almacenamiento mal ocupado. En consecuencia, un mal manejo de recursos. Como resultado, un aumento de complejidad computacional (Balari, 2005). La reducción de dimensionalidad (RD) permite el extraer las principales variables que capturen de manera robusta las características propias de los datos heterogéneos que conlleve a una información correcta se convierte en un requisito principal.

Estas técnicas tienen como principal objetivo disminuir el espacio de búsqueda, es decir, el número de características que componen este sistema, lo cual consecuentemente se refleja en un menor tiempo necesario para efectuar el proceso de clasificación de los datos y reducción de los mismos (Galarza Hernandez, 2017). Cabe recalcar, que las redes de sensores inalámbricos (WSN) están compuestas por varios nodos, los cuales tienen ciertas limitaciones, dentro de ellas están su limitada capacidad de memoria, su procesamiento y su alimentación energética la cual es a través de baterías. Por lo cual, es necesario reducir el número de datos para que tanto sus recursos computacionales como su consumo de batería sean óptimos.

## **1.3 Objetivos**

### **1.3.1 Objetivo general.**

Desarrollar una comparación entre los diferentes métodos de RD enfocados a algoritmos de clasificación supervisados, aplicados a datos de redes de sensores con el fin de determinar el más adecuado.

### **1.3.2 Objetivos específicos.**

- Realizar una revisión sistemática de literatura correspondiente a los métodos de RD.
- Analizar las características de la información contenida en bases de datos provenientes de WSN para realizar la RD a cada una de ellas.
- Determinar los algoritmos de clasificación supervisados de bajo consumo computacional para comprobar su eficiencia ante diferentes tipos de datos.
- Realizar una medida de cuantificación con el fin de seleccionar al RD y algoritmo de clasificación adecuado en relación con la naturaleza de los datos

## **1.4 Alcance**

Se realizará un estudio de cada uno de los de algoritmos de clasificación tanto supervisados como no supervisados ya existentes, con el fin de determinar los mejores métodos que puedan representar la RD aplicada a datos de redes de sensores.

Se normalizará ciertas bases de datos las cuales que se han recopilado con anterioridad de sistemas de redes de sensores, para poder trabajar con ellas como base dentro del proceso de programación.

Se procederá a la programación de algoritmos de clasificación para poder visualizar los resultados de RD al aplicar las bases de datos dentro de las mismas, y así proceder a la comparación de dichos algoritmos, para lo cual se propondrá una medida de cuantificación entre reducción de dimensionalidad y algoritmos de clasificación supervisados.

### **1.5 Justificación**

El aprovechar de una manera eficiente los grandes volúmenes de datos brinda varios beneficios, entre los cuales se puede destacar en el ámbito económico y científico (Van Rijsbergen, 1979). Y aunque en la actualidad los avances tecnológicos hacen posible el almacenamiento de esta gran cantidad de datos se hace más evidente la necesidad del uso de métodos que sean capaces de procesar solo los datos necesarios y además sean eficientes y eficaces (Combarro et al., 2009).

Cuando los parámetros dentro de una base de datos tienen un número de dimensionalidad alto, es conveniente aplicar la tecnología de RD con el objetivo de obtener los datos más relevantes los cuales son más pequeños de almacenar y más rápidos de procesar que toda la información obtenida sin afectar de manera incorrecta a los resultados de un sistema (Mao, Xue-min; Cai, Chuan-xi; Sun, 2012). Para lo cual existen diversos métodos o algoritmos que proponen como obtener el mínimo número de componentes que representen de igual manera al conjunto de datos.

Actualmente en el Ecuador el Gobierno Nacional a través del Plan Nacional del Buen Vivir y del Plan Nacional de Ciencia y Tecnología brinda apoyo a los proyectos tecnológicos con el fin

de mejorar la calidad de servicio y atención que brindan las empresas, haciendo uso eficiente de sus recursos en los cuales se hace una gran inversión. Además de realizar el cambio de la matriz productiva se intensifica la investigación en búsqueda el impulso hacia propias tecnologías las cuales pueden satisfacer necesidades como mejora de productos y servicios.

Cabe destacar que la mayoría de los sistemas de procesamiento en un computador o un nodo principal no hacen uso de una técnica de reducción de dimensionalidad, por lo cual lo convierte en un sistema ineficiente ya que no hace uso óptimo de sus recursos, y también lo hace costoso ya que se lo ha adquirido con características innecesarias de las que realmente se necesitará si se hace uso de RD.



## **CAPÍTULO II. REVISIÓN BIBLIOGRÁFICA**

En el presente capítulo se abarca aspectos como redes de sensores y sus limitaciones, además de los métodos utilizados para mejorar la eficiencia de estos. También, se describe acerca de los grandes volúmenes de datos y como se trabaja con ellos en la actualidad, los principios básicos y métodos de RD. En complemento, se menciona los algoritmos de clasificación supervisados los cuales son más usados en redes de sensores y bases de datos a utilizar.

### **2.1 Sistemas Embebidos**

La necesidad constante de recolectar información de cualquier tipo y poder almacenarla para su posterior análisis ha existido desde hace mucho tiempo atrás. Razón por la que se desarrollaron sistemas que permiten la recopilación de gran cantidad de datos mediante el uso de nodos sensores, los mismos que en la actualidad son aplicados en diferentes áreas, tales como la medicina, agricultura, prevención de desastres, entre otros (Dobrin, Stamatescu, Dragana, & Sgarciu, 2016). Esto se ha hecho posible a través del concepto de un sistema embebido que se interpreta como una plataforma que permite tener la potencia de una Unidad Central de Procesamiento (UCP) en un tamaño mucho menor, el cual se lo puede programar para cumplir con una función en específico. Este sistema puede contar con gran capacidad de memoria para almacenamiento y su cobertura dependerá de del tipo de red que se necesite dependiendo de la aplicación que se vaya a desarrollar (Sánchez Vítores, 2004).

#### **2.1.1 Características de Sistemas Embebidos**

Se caracterizan por ejecutar un programa específico de manera repetitiva, tener un bajo costo de implementación, tamaño reducido y consumo mínimo de potencia para lograr extender el

tiempo de vida de las baterías, además de contar con una arquitectura simplificada (Pérez, 2009). Estos sistemas deben responder a estímulos procedentes del entorno, procesando y respondiendo con información en tiempo real (tiempos de adquisición y procesamiento deben ser menores al período de actualización de dicha información), pueden controlar desde sistemas compuestos por pequeños sensores hasta sistemas distribuidos de telecomunicaciones (Ramesh, Sentilles, & Crnkovic, 2012).

### 2.1.2 Componentes principales

(Pérez, 2009) afirma que los componentes principales con los que cuenta un sistema embebido son:

- **Microcontrolador.**- Es un microprocesador que ha sido optimizado para aplicaciones de control embebidas.
- **Microprocesador.** - Es el núcleo computacional que realiza una gran cantidad de funciones en un solo circuito integrado.
- **Actuadores.** - Elementos electrónicos controlados por el sistema.
- **Memoria.** - En el caso de los microcontroladores está integrada en el propio chip. Dentro de la cual, una parte debe ser no volátil, es decir de tipo ROM, la cual está destinada a almacenar las instrucciones que realiza la aplicación. Y otra parte de memoria será tipo RAM, es decir volátil, y se reserva para guardar las variables y los datos.

- **Puertos de Entrada/Salida.-** Son aquellos puntos en los que los dispositivos periféricos se conectan y pueden intercambiar información tanto con la memoria como con el procesador central.

### **2.1.3 Aplicaciones**

Un sistema embebido se puede encontrar aplicado en una gran variedad de dispositivos electrónicos, desde teléfonos celulares, cámaras digitales, calculadoras, electrodomésticos, equipos de oficina, hasta en sistemas más complejos como automóviles o aplicaciones aeroespaciales (Parmesh, 2014).

## **2.2 Redes de Sensores Inalámbricos (WSN)**

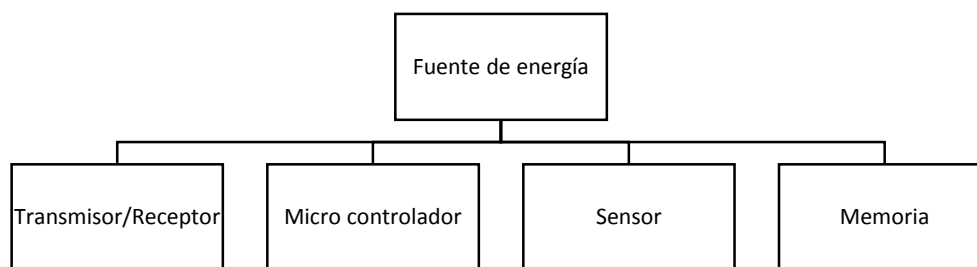
Son consideradas como un conjunto de nodos sensores que se encuentran distribuidos en un espacio determinado, además trabajan conjuntamente para controlar distintos parámetros, ya sea la temperatura, humedad, vibración o cualquier otro parámetro que deseemos dependiendo del tipo de sensor que usemos (Niels Aakvaag, 2006). Una red WSN es usada especialmente para la recopilación de datos en ambientes donde los sistemas de redes de sensores tradicionales no son eficientes y/o representan un gran costo de implementación (Rosero et al., 2018). En la actualidad es más común el uso de sensores inalámbricos, debido a su facilidad para trabajar con ellos. Ya que son mucho más cómodos y estéticos para aplicarlos en el cuerpo, máquinas, o áreas específicas tales como en edificaciones que se encuentren en un estado crítico (Dobrin et al., 2016).

A pesar de tener grandes beneficios, cabe recalcar que estos sensores tienen recursos limitados como la duración de su alimentación energética, ya que comúnmente es mediante

baterías. Debido a esto, es necesario la optimización de recursos para aprovechar al máximo el tiempo de trabajo de una WSN.

### 2.2.1 Componentes de un nodo WSN

Como cualquier sistema todo nodo WSN tiene componentes de los que se conforma básicamente, como se muestra en la Figura 1.



*Figura 1 Componentes de un nodo WSN  
Fuente: Adaptado de (Dobrin et al., 2016)*

Dobrin et al, (2016) asegura que los componentes con los que cuenta un nodo WSN son:

- **Sensor.** - Elemento encargado de medir parámetros del entorno de acuerdo con la aplicación a usarse.
- **Transmisor/Receptor inalámbrico.** – Realiza el proceso de comunicación con otros nodos.
- **Microcontrolador.** - Es el responsable de procesar la información recibida.
- **Fuente de energía.** - Brinda la alimentación energética para que todos los componentes del nodo funcionen.

- **Memoria.** – Almacena los datos recopilados por el sistema al ejecutar las funciones para las que está dedicado.

### 2.2.2 Funcionamiento

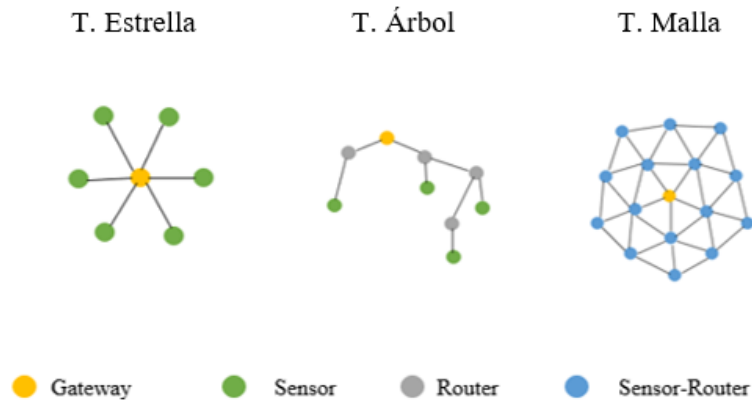
En una red WSN los nodos se encargan de sensar parámetros del entorno en el que están ubicados, procesan la información que deben transmitir ya sea directamente al Gateway el cual sirve como punto de unión con una red LAN o con Internet o hacia otro nodo que esté trabajando como enrutador de mensajes. Éste último es el encargado de transmitir los mensajes de nodos distantes hacia el nodo central el cual permite la interconexión hacia una red externa dependiendo de la topología implementada (Vela, 2016). La estación base se conecta con los nodos centrales, recopilando así la información de la red que permita finalmente presentar al usuario final.

### 2.2.3 Topologías de red

La topología de red para una WSN dependerá de la aplicación que se vaya a realizar, además de intentar obtener la mejor configuración de los componentes para un correcto funcionamiento de la red. Para lo cual, existen tres tipos de topologías con grandes ventajas para este tipo de red (San José Vieco, Pastor, García, & Zangroniz, 2011). Las cuales se detallan a continuación:

- **Estrella.** - Donde cada nodo tiene conexión directa con la puerta de enlace.
- **Árbol.** - Donde los nodos se conectan con un nodo-router hasta llegar a conectarse con la puerta de enlace.

- **Malla.** - Donde los nodos se pueden conectar con múltiples nodos tratando de enviar los datos por el camino que encuentre la red con mayor confiabilidad.



*Figura 2 Topologías de una WSN  
Fuente: Autoría propia*

### 2.3 Big Data

Actualmente existe gran cantidad de datos que son recopilados y transmitidos por diferentes dispositivos, como celulares, computadoras, electrodomésticos, etc. En las WSN sucede lo mismo, y el término Big Data hace referencia a grandes volúmenes de datos que pueden ser o no estructurados, tener diferente tamaño, tipo y velocidad, que identifica como característica a los datos con los que se trabaja en una WSN. Estos superan las capacidades de las herramientas computacionales tradicionales, por lo que requieren de herramientas más evolucionadas para que las bases de estos datos puedan ser almacenadas, analizadas, administradas, entre otros. (Sagioglu & Sinanc, 2013).

### **2.3.1 Introducción del Big Data**

La cantidad de información que es almacenada crece cuatro veces más rápido que la economía mundial, esto se debe al avance tecnológico que se ha dado en las últimas décadas(Ortuño, Barandiarán. José María, & Gil, 2014). La necesidad de almacenar información para su estudio se ha convertido en un factor importante. Cabe mencionar que un ejemplo de ello fue el año 2000 cuando inició en funcionamiento el telescopio Sloan Digital Sky Survey, que recopiló más información sobre astronomía de la que se había obtenido en la historia, y para la actualidad, el Gran Telescopio Sinóptico de la investigación de Chile recopila esa suma de datos cada cinco días aproximadamente (Mayer-Schönberger, 2013).

### **2.3.2 Aplicaciones del Big Data**

Existen diferentes áreas como por ejemplo en medicina, deporte, ciencias, investigación, marketing, entre otros, donde se puede aplicar técnicas de Big Data, siempre que sea necesario el extraer información dentro de grandes volúmenes de datos. Entre estos, se encuentra la detección de patrones de fraude mediante el análisis de gran cantidad de datos que se generan cada hora en entidades financieras, otro caso que es muy conocido, es el uso de Big Data para poder conocer lo que varios usuarios opinan sobre una empresa o producto y así realizar estrategias que mejoren parámetros deseados(Pérez Marquez, 2015).

### **2.3.3 Minería de Datos**

Se comprende como el análisis de todos los datos para adquirir los más relevantes de un conjunto grande de estos sin haber sido trabajados anteriormente (Ahmad Madni, Anwar, & Ali

Shah, 2017). También se lo puede considerar como un proceso de descubrimiento de nuevos patrones y tendencias dentro de un grupo almacenado de datos con el uso de tecnologías como reconocimiento de patrones, redes neuronales, algoritmos genéticos y otras técnicas avanzadas. La minería de Datos fue desarrollada para trabajar en conjunto con Big Data ya que este tipo de proceso es más útil trabajarlo dentro de grandes volúmenes de datos (López, 2007). Cabe recalcar que tanto la minería de datos como los métodos de RD se encuentran relacionados con el análisis estadístico de datos, el aprendizaje de máquina y la inteligencia artificial con el fin de comprender mejor grandes cantidades de datos (Aluja, 2001).

## **2.4 Reducción de Dimensionalidad**

Muchas veces se cree que mientras más información se tenga es mejor, pero en estos casos no se toma en cuenta que, procesar toda esta cantidad de datos requiere gran cantidad de recursos computacionales, además, al trabajar con datos sin ningún tratamiento significa trabajar con datos que pueden ser repetidos o irrelevantes, los mismos que pueden ser excluidos para un posterior análisis y representación gráfica (Griparis, Faur, & Datcu, 2015).

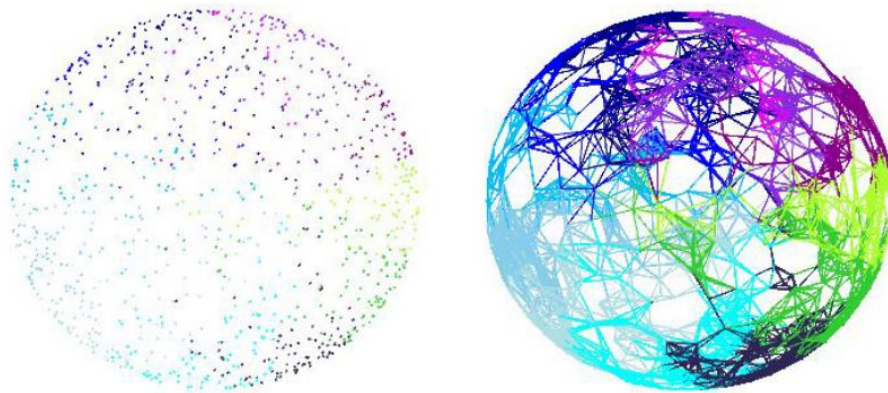
La RD es un proceso en el cual se realiza la transformación de datos que son de alta dimensión para reducir su complejidad y al realizar el procesamiento de estos se mejora el rendimiento computacional, lo que significa un ahorro tanto de coste como de tiempo. para representarlos en una dimensión reducida y de menor complejidad (Maaten, Postman, & Herik, 2008) (Griparis et al., 2015) .

Existen diferentes técnicas que permiten realizar este proceso de RD, los cuales se clasifican en métodos lineales, no lineales, espectrales basados en similitudes, disimilitudes,



métodos estocásticos basados en divergencias, entre otros. Según autores como: John A. Lee, Michel Verleysen, en su artículo *Nonlinear Dimensionality Reduction for Visualization* (Verleysen & Lee, 2013), así como también Jhon Gallego, Daniel Rengifo en su tesis con el tema *Comparación de técnicas de reducción de dimensionalidad para la clasificación de actividades físicas humanas utilizando métodos estadísticos* (Gallego & Rengifo, 2016) afirman que métodos espectrales basados en similitudes como Locally Linear Embedding, Laplacian Eigenmaps, métodos espectrales basados en disimilitudes como Análisis de componentes principales y métodos estocásticos basados en divergencias como SNE y *t*-SNE son los principales y la base de todos los métodos existentes de RD (Anaya, 2017). En base a esto, se trabajará con los ya mencionados métodos para el desarrollo de su comparación, los mismos que se detallan más adelante en este documento.

Para una mejor comprensión, es importante visualizar como se representan gráficamente datos en 3D, en la Figura 3 se muestra un ejemplo de un set de datos con 1500 puntos y alta dimensión.

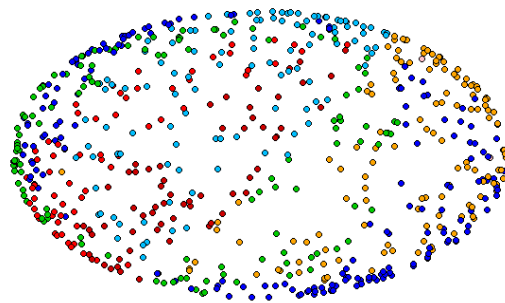


*Figura 3 Estructura esférica artificial en 3D.  
Fuente: (Anaya, 2017)*

### 2.4.1 Análisis de Componentes Principales (PCA)

Este método busca conseguir un nuevo conjunto de variables o componentes principales que están incorrelacionadas entre sí, a través de transformaciones ortogonales realizadas al conjunto de variables originales, todo esto con el objetivo de reducir la dimensionalidad de estos. Una característica de esta técnica es que logra trabajar fácilmente con gran cantidad de datos consiguiendo así evitar gran carga computacional (Fernández, 2011)(González García & Taborda Londoño, 2015).

PCA utiliza la varianza, que es una medida para expresar la cantidad de información que contiene una componente. Por esta razón, se prefiere como primer componente aquel que posee mayor varianza, mientras que por el contrario, el componente de menor varianza será el último. En la Figura 4 se muestra la transformación de la estructura esférica 3D en 2D mediante PCA.



*Figura 4 Visualización de método PCA  
Fuente: Autoría propia*

### 2.4.2 Laplacian Eigenmaps (LE)

Comúnmente los datos de alta dimensión muestran características de naturaleza no lineal, adaptativa y multifacética (Jiang, Lu, & Hong, 2011). Debido a esto, se hace necesario el uso de algoritmos de reducción de dimensionalidad no lineal como lo son LE y Locally Linear Embedding (LLE) donde los parámetros de vecindad deben seleccionarse previamente a su uso.

LE es un algoritmo que utiliza técnicas espectrales, es decir, se basa en la suposición de que los datos se encuentran en un colector de bajas dimensiones en un espacio de alta dimensión. Cada punto de datos se utiliza como un nodo en el gráfico y la conectividad entre los nodos se rige por la proximidad de los puntos vecinos (Qing Wu, Zongxian Qi, Zhicang Wang, & Yu zhang, 2018). Castro (2018) afirma que:

El fundamento de LE radica en que el espacio original contiene una cantidad suficiente de  $N$  puntos cercanos a un espacio  $d$  dimensional por lo que se puede representar dicho espacio subyacente con buenas precisiones mediante un grafo  $G = (VN, E)$ . De esta manera, los vértices  $v_i$  de dicho grafo se asocian con cada conjunto de puntos de referencia  $y_i$  y los bordes se obtienen mediante la conexión de  $v_i$  y  $v_j$  cuando existen puntos correspondientes que son vecinos. Para determinar la relación de la vecindad se pueden utilizar los  $k$  vecinos, los vecindarios esféricos o cualquier otro método basado en grafos (p. 28).

En la Figura 5 se muestra los resultados gráficos de la transformación realizada en la estructura esférica de 3D a 2D mediante el método LE.



*Figura 5 Visualización de método LE  
Fuente:(Anaya, 2017)*

### 2.4.3 Locally Linear Embedding (LLE)

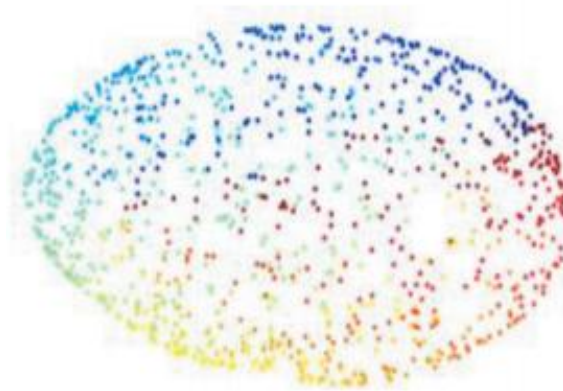
LLE busca hallar la variedad de baja dimensión dentro del conjunto de datos a trabajar los cuales son de alta dimensión, y además conservar la propiedad de que si dos puntos estaban próximos en el espacio origen lo seguirán estando en el nuevo espacio de baja dimensión realizando una búsqueda de vecinos más cercanos para construir la matriz de peso y descomponer parcialmente los valores propios (Quansheng Jiang, Minping Jia, & Jiayun Lv, 2008).

Como se menciona anteriormente, el algoritmo preserva las relaciones de vecindad, por esta razón cada punto de datos se puede reconstruir aproximadamente a partir de sus vecinos (Wen, Zhou, Wang, Zhang, & Wu, 2010). Este método necesita realizar una conexión manual de tres parámetros que son considerados como principales, los cuales son:

La dimensión de salida  $m$ , el número de vecinos más cercanos  $k$  y el parámetro de regularización  $\alpha$ . Como primer paso, se determina el vecindario para cada  $x$  identificando sus  $k$  vecinos más cercanos con base en la distancia Euclídea. Después

de determinar los vecindarios, cada punto  $x$  se representa como una combinación lineal ponderada de sus vecinos, encontrando los pesos de reconstrucción  $W$ . Para obtener la matriz con los datos mapeados de baja dimensión, se puede calcular los vectores propios asociados a los valores propios más pequeños (Valencia, Meza, Daza, Acosta, & Castellanos, 2010, p. 15).

(Roweis & Saul, 2000) destaca que LLE es capaz de retener la configuración local de los datos y es invariante ante rotaciones, rescalados, o traslaciones en los datos lo que lo convierte en un método especialmente versátil en el tratamiento de datos de carácter experimental. En la Figura 6 se visualiza la transformación de la estructura esférica al ser reducida su dimensionalidad mediante LLE.



*Figura 6 Visualización de método LLE  
Fuente:(Kumar, 2018)*

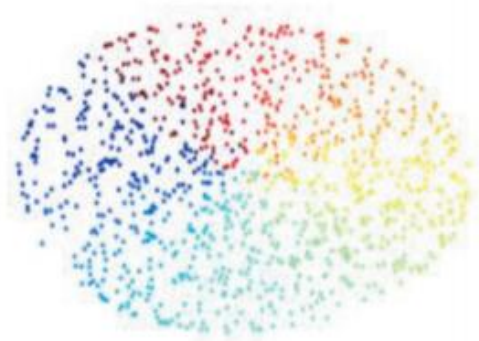
#### **2.4.4 Stochastic Neighbor Embedding (SNE)**

Es un método probabilístico de distribución estocástica de puntos más cercanos el cual se enfoca en conservar las identidades de vecinos de los datos en el mapa, se formuló como reducción

de una suma de divergencias de Kullback-Leibler, la cual es una medida que expresa la similitud o diferencia entre dos funciones de distribución (Popkes, 2019).

La idea principal es centrar una distribución gaussiana para cada valor de entrada de los datos de alta dimensión con el fin de usar su densidad para de esta manera establecer una distribución de probabilidad de todos los vecinos, para luego aproximar esta distribución de probabilidad tanto como sea posible repitiendo la estructura de parentesco en un espacio de menor dimensión. SNE es una mejora de algunos métodos como LLE o SOM en el cual los puntos de datos ampliamente separados pueden ser "colapsados" como vecinos cercanos en el espacio de baja dimensión. Un inconveniente que cabe destacar es su lenta convergencia en aplicaciones prácticas, el cual es el resultado de la implementación del descenso del gradiente, debido a esto se desarrolló t-SNE como corrección a este método (Kijoen, Hongmo, & Seungjin, 2004).

SNE como primer paso se calcula la probabilidad asimétrica para cada objeto y cada vecino potencial para obtener la distribución original, la diferencia la cual no es necesaria que sea simétrica se puede calcular utilizando la distancia euclídeana cuadrada a escala entre dos puntos de alta dimensión, el número efectivo de  $k$  vecinos se escoge manualmente. Luego se realiza el cálculo de la distribución inducida que, a diferencia de la original, se utiliza una varianza fija. El objetivo de la incrustación es hacer coincidir estas dos distribuciones lo mejor posible. Esto se logra minimizando una función de costo que es una suma de las divergencias de Kullback-Leibler entre las distribuciones original e inducida sobre vecinos para cada objeto (Hinton & Roweis, 2002). En la Figura 7 se muestra los resultados de la transformación de la estructura esférica con 3D a 2D mediante el método SNE.



*Figura 7 Visualización de método SNE  
Fuente: (Anaya, 2017)*

#### **2.4.5 T-distributed Stochastic Neighbor Embedding ( $t$ -SNE)**

(Gallego & Rengifo, 2016) Menciona que es una distribución estocástica de puntos más cercanos, este método hace uso de una distribución que mide la similitud entre pares de objetos de entradas y una medida de distribución entre parejas similares de los correspondientes puntos de análisis, representando así un conjunto con una menor dimensión.

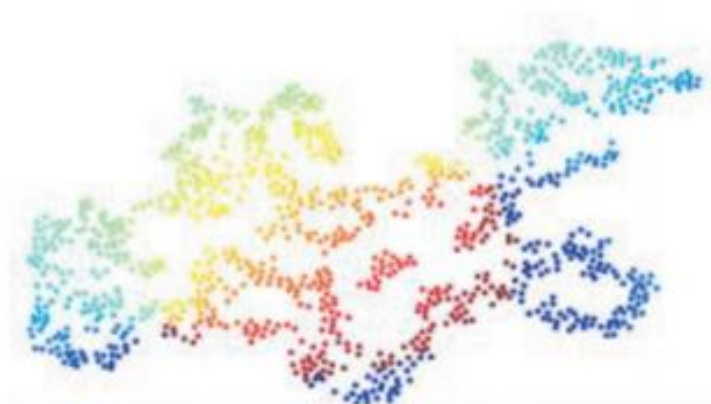
Este método comienza por encontrar patrones en los datos mediante la identificación de grupos que han sido observados según la similitud de los puntos de datos con varias características. Realiza la reducción de dimensionalidad debido a que asigna los datos multidimensionales a un espacio de menor dimensión. Por lo tanto, ya no es posible hacer algún tipo de deducción basándose solo en la salida obtenida de  $t$ -SNE.

Cabe mencionar que se puede utilizar en el proceso de clasificación y agrupación en clústeres utilizando su salida como la característica de entrada para otros algoritmos de clasificación. Se aplica ampliamente en el procesamiento de datos genómicos, archivos de voz e

imágenes como por ejemplo en el análisis del cerebro y exploraciones del corazón (Gallego & Rengifo, 2016).

Una aplicación importante de  $t$ -SNE es la identificación de subpoblaciones tumorales, ya que mediante una visualización no lineal de los datos es capaz de resolver mejor la heterogeneidad intratumoral biomolecular, ya que se puede descubrir subpoblaciones de tumores que están vinculadas estadísticamente a la supervivencia del paciente en el cáncer gástrico y al estado de metástasis en tumores primarios de cáncer de mama.

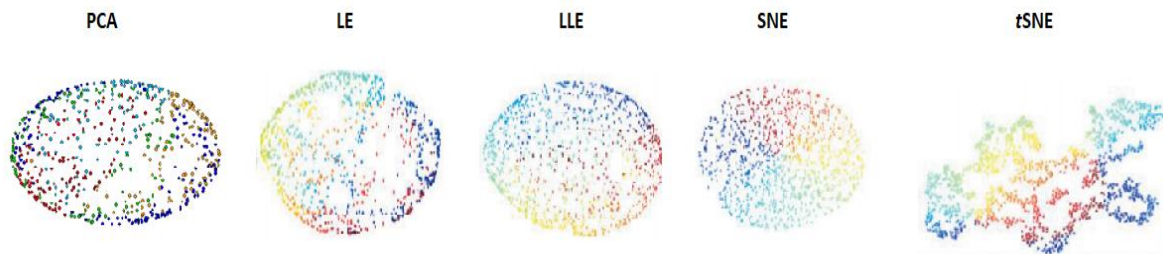
En la Figura 8 se muestra la transformación gráfica de la estructura esférica de 3D a 2D mediante  $t$ -SNE.



*Figura 8 Visualización de método  $t$ -SNE  
Fuente:(Anaya, 2017)*

En la Figura 9 se muestra todos gráficos resultantes de todos los métodos aplicados a la estructura esférica de alta dimensión, obteniendo una visión general y fácil de comparar visualmente.





*Figura 9 Métodos de RD aplicados a estructura esférica de 3D.  
Fuente: Adaptado de (Anaya, 2017)*

## 2.5 Algoritmos de clasificación supervisada

Estos algoritmos son capaces de encontrar la relación entre las características de cada punto de datos y su clase o etiqueta, y además establecer una línea de separación entre ellos, de manera que cuando se le facilitan nuevos datos, puede predecir la clase a la que pertenece según sus características por lo que son muy útiles. A continuación, se detallan los algoritmos que son considerados los más esenciales dentro la clasificación y predicción de datos (Iozu, 2017).

### 2.5.1 k-nearest neighbors (k-NN)

k-NN es un método no paramétrico para clasificar datos desconocidos, es decir, no hace suposiciones sobre la distribución de datos subyacente. Además, depende de varios factores clave diferentes, como una medida de distancia adecuada, una medida de similitud para votar y el parámetro k. Un conjunto de vectores y etiquetas de clase que están relacionados con cada vector constituyen cada uno de los datos de entrenamiento (Elmurngi & Gherbi, 2017).

Se utiliza con frecuencia para clasificar datos futuros debido a su simplicidad, facilidad de implementación y efectividad. Es uno de los diez principales algoritmos de minería de datos que

se ha aplicado ampliamente en varios campos de reconocimiento de patrones, diagnóstico de cáncer, clasificación de texto, etc. k-NN es un método de aprendizaje de alto coste computacional y basado en instancias (Taneja, Gupta, Goyal, & Gureja, 2014). El clasificador o modelo no se construye durante el período de entrenamiento, lo que lleva a un alto tiempo de computación y costo durante el período de clasificación.

### **2.5.2 Support Vector Machines (SVM)**

Es un algoritmo de clasificación robusto, el cual examina los datos e identifica los patrones utilizados para el análisis de regresión y clasificación. Fue principalmente diseñado para dar solución a problemas de clasificación de grandes dimensiones como una implementación de la teoría del aprendizaje estadístico (Yang, Wan, & Yang, 2012).

Establece un hiperplano de separación y un margen máximo libre de datos de entrenamiento al escoger un subconjunto llamado vectores de soporte, es decir, SVM hace una búsqueda para encontrar una curva que sea idónea para separar y clasificar los datos de entrenamiento, y así garantizar que la separación entre ésta y ciertas observaciones del conjunto de datos de entrenamiento, conocidos como vectores de soporte sea la más grande posible (Braun, Weidner, & Hinz, 2011).

### **2.5.3 Clasificador Bayesiano**

El modelo de clasificador bayesiano es una clase de clasificador de probabilidad que puede realizar un razonamiento probabilístico bajo incertidumbre usando el teorema de Bayes (Kok-Chin Khor, Choo-Yee Ting, & Somnuk-Phon Amnuaisuk, 2010). Comparado con algoritmos de

clasificación más sofisticados, como el árbol de decisión y la red neuronal, el clasificador bayesiano puede ofrecer una muy buena precisión de clasificación en muchas aplicaciones prácticas.

Este clasificador sigue siendo popular por su rendimiento competitivo en muchos dominios de investigación y su simplicidad en la computación que permite a los investigadores ahorrar una gran cantidad de costos computacionales.

#### **2.5.4 Árbol de decisión**

Es una herramienta robusta de clasificación de datos que se caracteriza por ser fácil de interpretar. Su objetivo principal es predecir el valor de una variable mediante el aprendizaje de reglas de decisión simples derivadas de las características que presentan los datos. Es comúnmente usada por no requerir mucha preparación de los datos a trabajar a diferencia de otras técnicas (Bouza & Santiago, 2012). Por lo mismo tiene una tasa de error mayor que otros algoritmos como lo es SVM.

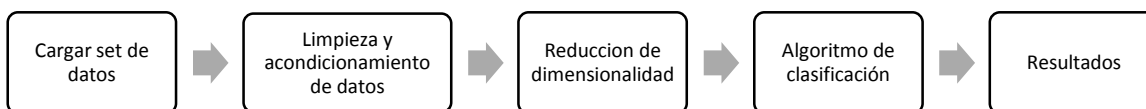
En este algoritmo cada nodo rama representa una elección o respuesta entre distintas alternativas, mientras que cada nodo de hoja representa una decisión. Un cambio en los datos, aunque sea el más pequeño, puede representar un gran cambio en la estructura del árbol.

## CAPÍTULO III. DESARROLLO EXPERIMENTAL

En este capítulo se realiza un análisis de las características de las bases de datos con las cuales se van a trabajar, además se presenta todo el proceso necesario para la comparación de los cinco métodos de RD y determinación de los algoritmos de clasificación supervisados de bajo consumo computacional presentados en el capítulo anterior.

### 3.1 Descripción General

En la Figura 10 se describe de manera general y secuencial las etapas a realizarse para la correcta comparación de los resultados.



*Figura 10 Descripción general de funcionamiento  
Fuente: Autoría propia*

El proceso para realizarse en cada una de las cinco bases de datos es primero cargar el set de datos los cuales se encuentra en formato csv, luego se procede a la limpieza y acondicionamiento de estos para que de esta manera se pueda obtener resultados con mejor precisión al aplicar el método de reducción de dimensionalidad. A continuación, se procede a emplear un algoritmo de clasificación y mediante la matriz de confusión obtener la exactitud con

la que los datos fueron clasificados al aplicar el método de RD con cada uno de los algoritmos de clasificación.

### 3.2 Adquisición de Bases de Datos

Para el desarrollo de la RD en este trabajo se hace uso de cinco diferentes bases de datos recopiladas por sistemas embebidos aplicados en diferentes áreas. La mayoría de estas bases de datos han sido seleccionadas de una plataforma de búsqueda de datos científica como lo es IEEE Dataport. Esto se ha realizado debido a que son datos que han sido verificados y útiles para poder ser usados en este proyecto, los mismos que se enlistan y describen en la Tabla 1.

Tabla 1  
Descripción general de las bases de datos.

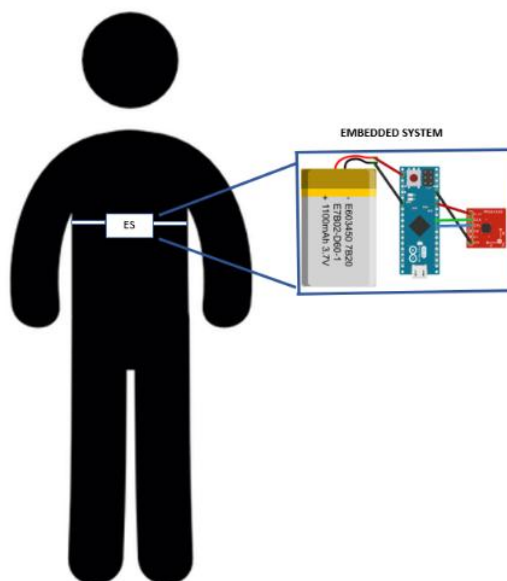
<b>Nombre</b>	<b>Repositorio</b>	<b>Link</b>	<b>Variab obtenidas</b>
-BDD1	IEEEdataport	<a href="https://iee-dataport.org/open-access/body-position-data-accelerometer-sensor">https://iee-dataport.org/open-access/body-position-data-accelerometer-sensor</a>	Posición del cuerpo
-BDD2	IEEEdataport	<a href="https://iee-dataport.org/open-access/data-set-wheelchair-sensors">https://iee-dataport.org/open-access/data-set-wheelchair-sensors</a>	Problemas de salud por uso de silla de ruedas
-BDD3	IEEEdataport	<a href="https://iee-dataport.org/open-access/footstep-analysis-using-pressure-sensors">https://iee-dataport.org/open-access/footstep-analysis-using-pressure-sensors</a>	Tipo de pisada
-BDD4	IEEEdataport	<a href="https://iee-dataport.org/open-access/roses-greenhouse-cultivation-database-repository-rosesgreenhdb">https://iee-dataport.org/open-access/roses-greenhouse-cultivation-database-repository-rosesgreenhdb</a>	Acciones necesarias en cultivos de rosas
-BDD5	IEEEdataport	<a href="https://iee-dataport.org/open-access/alcohol-detection-drivers-sensors-and-computer-vision">https://iee-dataport.org/open-access/alcohol-detection-drivers-sensors-and-computer-vision</a>	Presencia de alcohol

*Fuente: Autoría propia*

### 3.2.1 Datos de posición del cuerpo con sensor acelerómetro.

Los datos son obtenidos a través de realizar pruebas en diez personas diferentes mediante un sistema electrónico que permite conocer la posición del cuerpo humano mediante el uso de un acelerómetro de tres ejes para detectar cinco posiciones comunes, las cuales son decúbito ventral, decúbito lateral derecho, decúbito lateral izquierdo, decúbito supino y sentado (Rosero, 2018). El acelerómetro usado obtiene datos de tres ejes posibles (X, Y, Z) .

Como se muestra en la Figura 11 el sensor se lo ubica en el pecho, mientras que el eje Y apunta hacia la barbilla de la persona que lo está usando.



*Figura 11 Esquema de ubicación del sistema  
Fuente: (Rosero, 2018)*

En la Tabla 2, se muestra la asignación de etiquetas realizadas para cada uno de los casos.

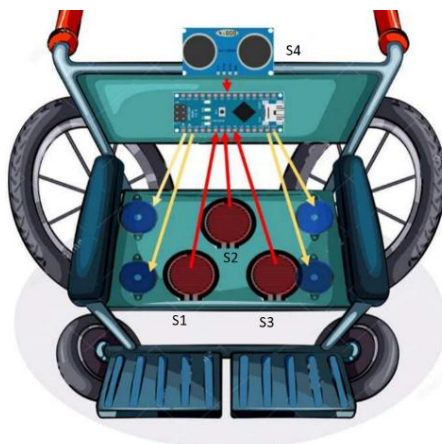
Tabla 2  
Asignación de etiquetas para la posición del cuerpo

<b>Posición</b>	<b>Etiqueta</b>
-Decúbito supino	1
-Decúbito lateral izquierdo	2
-Decúbito lateral derecho	3
-Sentados	4
-Decúbito ventral	5

*Fuente: Autoría propia*

### 3.2.2 Conjunto de datos sensores de silla de ruedas

En el sistema se pretende detectar los posibles problemas de salud en personas que utilicen silla de ruedas mediante la implementación de sensores de posición de presión en las mismas. En la Figura 12 se representa los sensores de posición de presión y los actuadores con círculos rojos grandes y azules en el asiento. Los círculos negros muestran los sensores de ultrasonido en el respaldo, las flechas amarillas muestran la salida periférica y las flechas rojas indican la entrada periférica en el convertidor analógico-digital (P. Rosero-Montalvo, Peluffo-Ordonez, López, Serrano, & Rosero, 2019) (Rosero, 2018).



*Figura 12 Esquema de ubicación de los sensores*  
*Fuente:(P. Rosero-Montalvo et al., 2019) (Rosero, 2018)*

En la Tabla 3, se muestra la asignación de etiquetas realizadas para cada uno de los casos.

Tabla 3

Asignación de etiquetas respectivas para posibles problemas de salud

<b>Posibles problemas de salud</b>	<b>Etiqueta</b>
-Posición correcta sin daño	1
-Presión más alta o lado derecho Problemas respiratorios, desequilibrio muscular, estrés en el hígado, el estómago y el riñón derecho	2
-Presión más alta o lado izquierdo Problemas respiratorios, estrés por desequilibrio muscular en el hígado, el estómago y el riñón izquierdo	3
-Presión hacia adelante más alta Problemas con las rodillas, dolor de espalda y estrés en hombros de abod.	4

*Fuente: Autoría propia*

### 3.2.3 Base de datos tipos de pisadas

Los datos en el sistema son recopilados por sensores de presión y son procesados por un Arduino mega 2560. Como se muestra en la Figura 13, los sensores de presión son representados por círculos negros, los cuales se encuentran distribuidos en las tres zonas de cada pie.



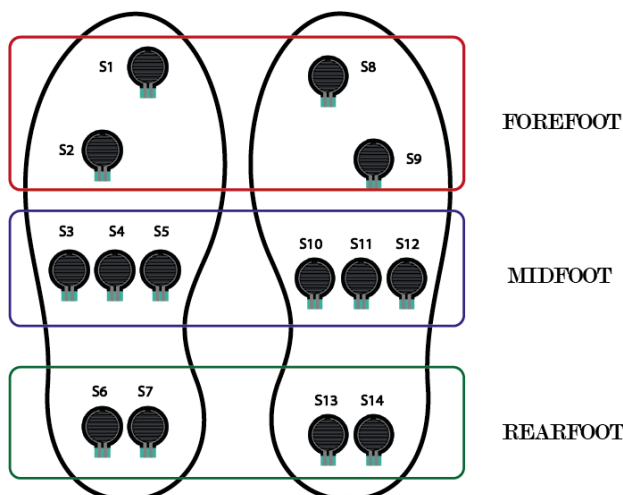


Figura 13 Esquema de ubicación de sensores en las plantillas  
Fuente: (Fuentes, Rosero, & Peluffo, 2019)

Para el pie izquierdo: S1 y S2 abarcan el área del antepié. S3, S4 y S5 la zona del medio pie; S6 y S7 la zona del retropié o talón; para el pie derecho: S8 y S9 representan la zona del antepié. S10, S11, S12 la zona del medio pie; S13 y S14 la zona del talón (Fuentes et al., 2019). En la Tabla 4, se muestra la asignación de etiquetas realizadas para cada tipo de pisada.

Tabla 4

Asignación de etiquetas para cada tipo de pisada

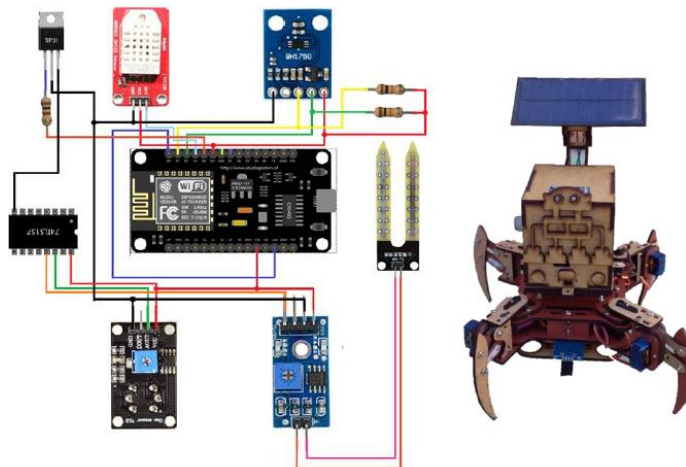
Tipo pisada	Etiqueta
-Normal	1
-Plano	2
-Cavo	3

Fuente: Autoría propia

### 3.2.4 Reconocimiento de patrones de datos en rosas

Es un sistema electrónico el cual consta de distintos sensores como se puede visualizar en la Figura 14 el cual cuenta con elementos como sensor de humedad del suelo, un sensor de humedad relativa, un sensor de temperatura, un sensor de luminosidad y un sensor de CO<sub>2</sub>, los

cuales se encargan de recopilar datos del ambiente. Todos estos sensores se conectan al sistema embebido mediante un módulo inalámbrico (Champutiz, Rosero, Fuentes, & Peluffo, 2019).



*Figura 14 Esquema de conexión del sistema*  
Fuente: (Champutiz et al., 2019)

Este sistema tiene como objetivo medir las variables que influyen en el desarrollo de las rosas e identificar el patrón existente entre estos valores como se representa en la Tabla 5.

Tabla 5

Asignación de etiquetas para cada acción necesaria en el sistema

<b>Tipo Ambiente</b>	<b>Etiqueta</b>
-Muy seco	1
-Ambiente correcto	2
-Demasiado calor	3
-Muy frío	4

*Fuente: Autoría propia*

### 3.2.5 Detección de alcohol en conductores por sensores y visión por ordenador

Como se muestra en la Figura 15, este sistema cuenta con un sensor MQ-3 para medir la concentración de alcohol en el medio ambiente, un sensor de temperatura MLX90621 que mide la temperatura de los puntos definidos en la cara del conductor y una cámara Raspberry Pi para obtener imágenes y de esta manera identificar y reconocer el grosor de la pupila (P. D. Rosero-Montalvo, López-Batista, Peluffo-Ordóñez, Erazo-Chamorro, & Arciniega-Rocha, 2019) (Rosero & Peluffo, 2019).

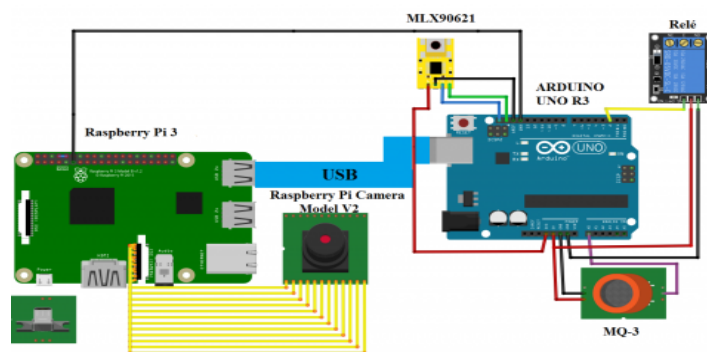


Figura 15 Esquema de conexión del sistema  
Fuente: (P. D. Rosero-Montalvo et al., 2019) (Rosero & Peluffo, 2019)

Al obtener ya los datos se detecta mediante la verificación de 5 variables si existe o no presencia de alcohol en el conductor, las cuales se identifican por etiquetas. En la Tabla 6 se enlistan los mismos.

Tabla 6

Asignación de etiquetas para cada clase

Clase	Etiqueta
-No presencia de alcohol	1
-Presencia de alcohol	2

Fuente: Autoría propia

### 3.3 Requerimientos

Para el desarrollo de este trabajo de comparación de métodos de RD los cuales deben ser desarrollados mediante programación, se debe establecer los requerimientos de Software que están descritos con el acrónimo SRSH y se les asigna una prioridad con la que debe cumplir dentro de este trabajo (Ver Tabla 7).

Tabla 7

Requerimientos de Software

		SRSH		
		Requerimientos	Prioridad	
			Alta	Media
Requerimientos de software				
SRSH1	La plataforma debe ser orientada al análisis de datos.	X		
SRSH2	La plataforma debe realizar una buena representación gráfica de datos.	X		
SRSH3	Costo bajo o gratuito de licencia.		X	
SRSH4	Debe consumir menor cantidad de memoria y RAM de la PC.	X		
SRSH5	Permite cargar set de datos en cualquier formato.		X	
SRSH6	Permite la limpieza y acondicionamiento de datos.	X		
SRSH7	Debe poseer librerías orientadas a los métodos de RD.	X		
SRSH8	Debe poseer librerías orientadas a algoritmos de clasificación.	X		
SRSH9	La plataforma debe ser de fácil programación.		X	

*Fuente: Autoría propia*

### 3.3.1 Elección de plataforma para gráficos de RD

La visualización de los resultados al aplicar los diferentes métodos de RD es importante en la comparación de éstos y para la interpretación de los datos. Para este proyecto es importante una plataforma orientada al análisis de datos que permita cumplir con la metodología planteada para este proyecto, es decir, que permita cargar un set de datos específico, la limpieza y acondicionamiento de datos, el implementar métodos de RD, algoritmos de clasificación y la representación gráfica de datos. Todo esto con un consumo bajo tanto de memoria como de RAM de la PC, además, es importante considerar que sea fácil para programar y que no sea de un alto costo de adquisición. Debido a esto se a tomado en cuenta cuatro plataformas que se destacan las cuales son RStudio, Python, Matlab y SAS, las mismas que serán utilizadas para determinar la mejor plataforma para el desarrollo de este trabajo.

Para la comparación y elección de plataforma se realiza la asignación de valor 1 para determinar los requerimientos con los que cumple y con un valor 0 para los requerimientos con los que no cumple o solo cumple parcialmente cada una de las plataformas y finalmente obtener un puntaje de valoración total (Ver Tabla 8).

Tabla 8  
Elección de plataforma

Plataforma	Requerimientos									Valoración total
	SRSH1	SRSH2	SRSH3	SRSH4	SRSH5	SRSH6	SRSH7	SRSH8	SRSH9	
-RSTUDIO	1	1	1	1	1	1	1	1	1	9
-PYTHON	1	1	1	1	1	1	1	1	0	8
-MATLAB	1	1	0	0	1	1	1	1	0	6
-SAS	1	1	0	0	1	1	1	1	1	7
Cumple totalmente: 1										
Cumple parcialmente: 0										

*Fuente: Autoría propia*

Se escoge como plataforma para la realización de este trabajo a RStudio, debido a que en su versión gratuita principalmente permite cumplir con la metodología planteada, es decir, cuenta con variedad de librerías orientadas tanto para la implementación de métodos de RD como la de algoritmos de clasificación y el desarrollo de gráficos los cuales permitan visualizar los resultados de aplicar RD. Además, RStudio cuenta con un lenguaje de programación de mayor facilidad y comprensión en comparación con las demás plataformas para poder desarrollar el presente trabajo.

### 3.3.2 Limpieza y acondicionamiento de datos

Es necesario antes de implementar cualquier método de RD, el realizar un proceso de limpieza y acondicionamiento de los datos que se obtiene de las distintas bases de datos. Debido a que cada base de datos hace uso de varios tipos de sensores los cuales miden diferentes parámetros, razón por la cual se obtienen distintos rangos de valores en cada una de las variables. Este proceso trata de escalar los datos para que ninguna de las variables de las que esté compuesta una base de

datos sea más importante que otras y afecte directamente a los resultados. En la Figura 16 se muestra los datos originales del dataset.

	↑ i.HS..Analog. ⇅	L..Lux. ⇅	T..Â.. ⇅	CO2..Analog. ⇅	HR... ⇅	clase ⇅
1	713	0.83	16.2	76	96.8	1
2	717	2.50	16.2	75	96.5	1
3	717	561.67	16.1	78	96.7	1
4	716	640.00	15.9	79	96.5	1
5	718	568.33	16.1	79	96.6	1
6	712	935.83	16.4	80	96.1	1
7	717	971.67	16.2	80	96.5	1
8	716	675.00	16.2	81	96.9	1
9	712	835.00	16.2	84	96.9	1
10	703	835.00	16.8	85	94.7	1
11	713	335.83	16.3	90	95.3	1

Figura 16 Datos originales de BDD4

Fuente: Autoría propia

Mediante la función de normalización en cada una de las variables dependientes para poder escalar todos los valores en rangos de 0 a 1 como se muestra a continuación en la Figura 17:

	↑ HS ⇅	LUX ⇅	Tem ⇅	CO2 ⇅	HR ⇅
1	0.5966667	0.08666667	0.010000000	0.070000000	0.9933333
2	0.6133333	0.09666667	0.010000000	0.06666667	0.9733333
3	0.6133333	0.23666667	0.003333333	0.100000000	0.9900000
4	0.6033333	0.25000000	0.000000000	0.11666667	0.9733333
5	0.6266667	0.24000000	0.003333333	0.11666667	0.9833333
6	0.5900000	0.30666667	0.033333333	0.13333333	0.9533333
7	0.6133333	0.31333333	0.010000000	0.13333333	0.9733333
8	0.6033333	0.25666667	0.010000000	0.14000000	0.9966667
9	0.5900000	0.28333333	0.010000000	0.16000000	0.9966667
10	0.5866667	0.28333333	0.040000000	0.16666667	0.8666667
11	0.5966667	0.21000000	0.026666667	0.22333333	0.9133333

Figura 17 Datos transformados de BDD4

Fuente: Autoría propia

### 3.4 Reducción de dimensionalidad

Cuando ya están los datos limpios y representados en una misma escala, se procede a realizar la implementación del método de RD respectivo, en la Figura 18 se presenta un diagrama de flujo del proceso en general de la implementación de cada método de RD:



Figura 18 Diagrama de flujo de proceso de RD

Fuente: Autoría propia



### 3.4.1 PCA

Para el análisis de componentes principales se debe realizar el cálculo de eigenvalores y de eigenvectores de las variables dependientes, para lo cual se usa la función *cov* la cual permite obtener la matriz de covarianza para luego aplicar la función *eigen* a la misma y así obtener los valores y vectores de los datos (Ver Figura 19).

```
> dat.eig = eigen(data.cov)
> dat.eig
eigen() decomposition
$values
[1] 0.236119224 0.110829237 0.048988676 0.017804717 0.006451424

$vectors
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.2245129 0.64602680 0.6700496 0.2691349 -0.10413115
[2,] -0.5480273 0.06936163 -0.2161280 0.8021274 -0.06881425
[3,] -0.5637510 0.10727087 0.2986240 -0.2522408 0.71963608
[4,] -0.2130787 0.72499027 -0.5115346 -0.3623779 -0.18973991
[5,] 0.5348292 0.20179298 -0.3917632 0.2986887 0.65615850
```

Figura 19 Eigen valores y vectores de BDD4  
Fuente: Autoría propia

Se procede a obtener los componentes principales de las variables dependientes que componen a la BDD, en este caso se obtiene cinco componentes principales y mediante el uso de la función *summary* se puede visualizar la desviación estándar, proporción de varianza y la proporción acumulativa la cual es de interés para determinar el número de componentes principales a utilizar para la reducción de dimensionalidad (Ver Figura 20).

```
> summary(pca.dataset)
Importance of components:
                PC1    PC2    PC3    PC4    PC5
Standard deviation  0.5224 0.3219 0.2200 0.12027 0.08163
Proportion of Variance 0.6119 0.2323 0.1085 0.03243 0.01494
Cumulative Proportion 0.6119 0.8442 0.9526 0.98506 1.00000
```

Figura 20 Varianza acumulativa de componentes principales de BDD4

Fuente: Autoría propia

Como se muestra en la Figura 20 el componente PC1 tiene una porción de varianza del 61.19%, lo cual significa que este componente representa un poco más de la mitad del total de la información. Al acumular este valor de varianza con el del componente PC2 se puede ver que se obtiene una representación del 84.42% de información total, que es un porcentaje muy significativo con el cual se puede realizar una RD. Y visualizar con relación a los dos componentes principales la aplicación de PCA como se muestra en la Figura 21:

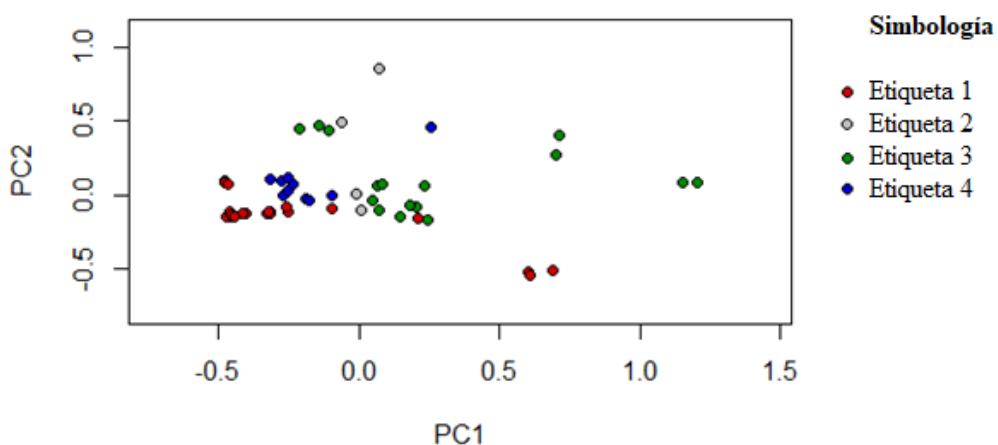


Figura 21 PCA aplicado en BDD4

Fuente: Autoría propia

### 3.4.2 LE

Para el desarrollo del método LE en RStudio es necesario cargar la librería *Rdimtools*, ya que contiene varios métodos utilizados para la reducción de dimensionalidad. Esta librería permite el uso de la función *do.lapeig*, la cual permite realizar los calculos necesarios para la RD mediante el método LE.

Antes de realizar la RD se debe hacer el cálculo de el valor  $k$  óptimo de vecinos mediante la función *calc\_k* donde se especifica el dataset, la dimensión intrínseca, el valor mínimo de  $k$  y el valor máximo de  $k$ . Los resultados se pueden expresar gráficamente como se muestra en la Figura 22:

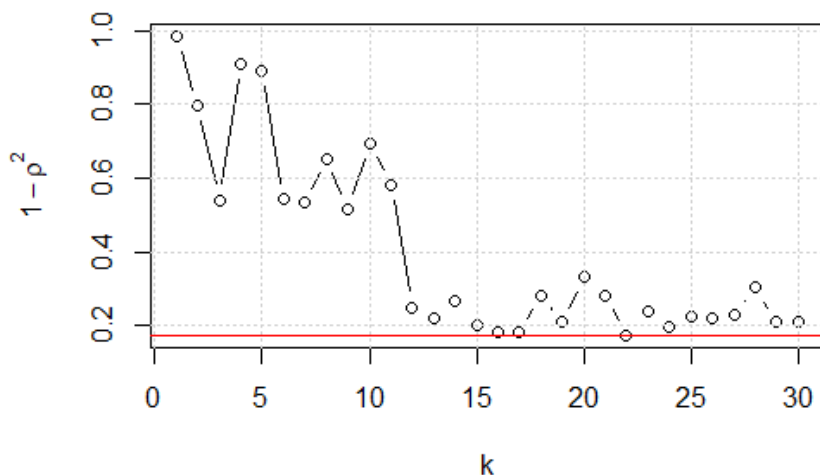


Figura 22 Cálculo de valor óptimo de  $k$  en BDD4  
Fuente: Autoría propia

También se lo puede expresar mediante una tabla con los valores numéricos para cada  $k$  vecino, donde se escoge el número de vecinos con el valor más pequeño de  $\rho$ , que en este caso el número de  $k$  sería 16 como se muestra a continuación en las Figura 23:

	k	rho
1	1	0.9845426
2	2	0.7976554
3	3	0.5417747
4	4	0.9119127
5	5	0.8919747
6	6	0.5421528
7	7	0.5324345
8	8	0.6538042
9	9	0.5169321
10	10	0.6958041
11	11	0.5828936
12	12	0.2505057
13	13	0.2193897
14	14	0.2672286
15	15	0.2019025
16	16	0.1809177
17	17	0.1819593
18	18	0.2834095
19	19	0.2090331
20	20	0.3332867
21	21	0.2828475
22	22	0.1750188
23	23	0.2375021
24	24	0.1986897
25	25	0.2258006
26	26	0.2198122
27	27	0.2315376
28	28	0.3027562
29	29	0.2099915
30	30	0.2122896

Figura 23 Valores de  $\rho$  para cada  $k$  vecino  
Fuente: Autoría propia

La función `do.lapeig` necesita como parametros el dataset, el número de dimensiones a las cuales se va a reducir el conjunto de datos originales, el tipo de construcción del gráfico de vecinos que en este caso será mediante knn con un número de 16 vecinos calculados anteriormente, y el valor de grafico de ponderación laplaciano verdadero.

```
le.test = do.lapeig(test_set,ndim=2,type = c("knn",16),weighted = TRUE)
```

El objeto que se tiene como respuesta almacena los valores de las dos dimensiones en el elemento Y cómo se puede verificar a en la Figura 24:

```

> le.test$Y
      [,1]      [,2]
[1,] -0.16812334 -5.023487e-03
[2,] -0.16817355 -5.030342e-03
[3,] -0.16829923 -5.040001e-03
[4,] -0.17044265 -4.268660e-03
[5,] -0.15773557  8.620293e-04
[6,] -0.15145199 -6.363892e-04
[7,] -0.14503311 -1.077677e-04
[8,] -0.14497014 -1.077300e-04
[9,] -0.14119422 -2.673571e-05
[10,] -0.17699356 -4.567581e-03
[11,] -0.17878596 -5.070750e-05
[12,] -0.19613152 -4.805705e-03
[13,] -0.19679786 -5.828571e-03
[14,] -0.19691675 -5.829532e-03

```

Figura 24 Valores de las dos dimensiones resultantes de LE  
Fuente: Autoría propia

Con estos valores obtenidos se pueden graficar los resultados para poder visualizar la RD realizada por este método como se muestra en la Figura 25:

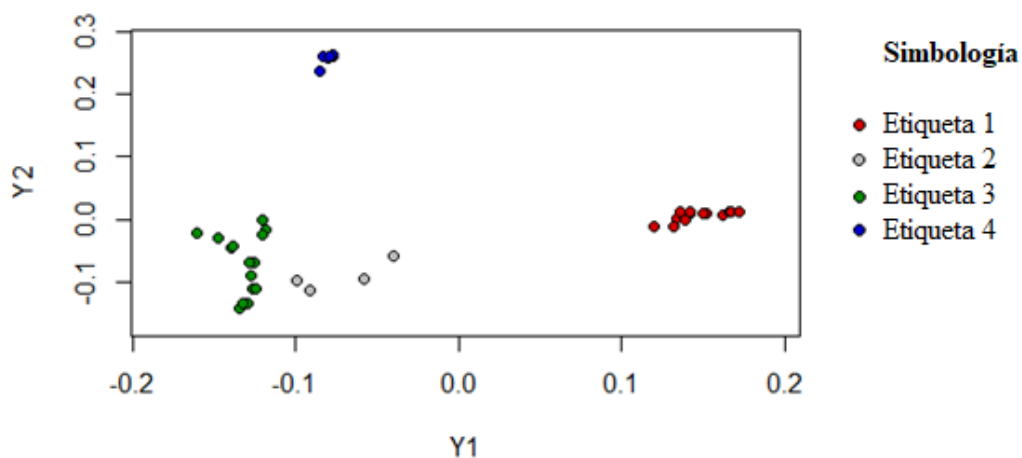


Figura 25 LE aplicado en BDD4  
Fuente: Autoría propia

### 3.4.3 LLE

Para el desarrollo de LLE en RStudio es necesario instalar la librería *lle*, este realiza todos los pasos del algoritmo LLE llamando a las funciones del paquete. Como primer paso se debe calcular el valor óptimo de los  $k$  vecinos más cercanos, para esto se hace uso del método *calc\_k*, donde se especifica el dataset, la dimensión intrínseca, el valor mínimo de  $k$  y el valor máximo de  $k$ . Los resultados se pueden expresar gráficamente como se muestra en la Figura 26:

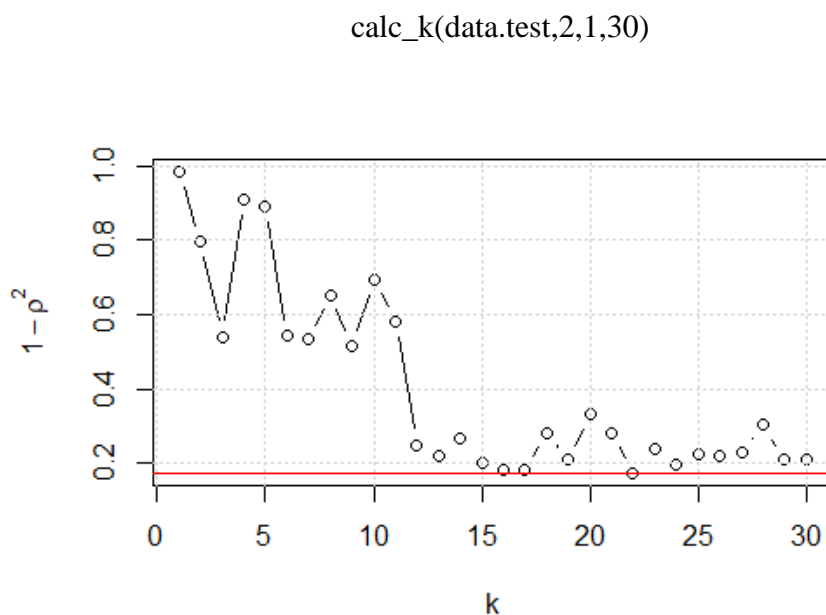


Figura 26 Cálculo de valor óptimo de  $k$  en BDD4  
Fuente: Autoría propia

También se lo puede expresar mediante una tabla con los valores numéricos para cada  $k$  vecino, donde se escoge el número de vecinos con el valor más pequeño de  $\rho$ , que en este caso el número de  $k$  sería 16 como se muestra a continuación en las Figura 27:

	k	rho
1	1	0.9845426
2	2	0.7976554
3	3	0.5417747
4	4	0.9119127
5	5	0.8919747
6	6	0.5421528
7	7	0.5324345
8	8	0.6538042
9	9	0.5169321
10	10	0.6958041
11	11	0.5828936
12	12	0.2505057
13	13	0.2193897
14	14	0.2672286
15	15	0.2019025
16	16	0.1809177
17	17	0.1819593
18	18	0.2834095
19	19	0.2090331
20	20	0.3332867
21	21	0.2828475
22	22	0.1750188
23	23	0.2375021
24	24	0.1986897
25	25	0.2258006
26	26	0.2198122
27	27	0.2315376
28	28	0.3027562
29	29	0.2099915
30	30	0.2122896

Figura 27 Valores de  $\rho$  para cada  $k$  vecino  
Fuente: Autoría propia

Se procede a aplicar la función *lle* en la cual se especifica el dataset, la dimensión intrínseca, el número de vecinos, el método de regularización, los valores lógicos que indican el tiempo para realizar la selección de subconjuntos, los valores lógicos que indican el tiempo para calcular la dimensión intrínseca y el parámetro umbral para la dimensión intrínseca.

```
lle.test=lle(training_set, m=2, k=16, reg=1, ss=FALSE, id=TRUE, v=0.9 )
```

Al ejecutar esta línea de código inmediatamente la función realiza el cálculo y descubrimiento de los vecinos más cercanos y calcula su peso, es decir sigue con el procedimiento del algoritmo LLE como se muestra en la Figura 28.

```
> lle.test=lle(test_set[,1:5], m=2, k=22, reg=1, ss=FALSE, id=TRUE, v=0.9 )
finding neighbours
calculating weights
intrinsic dim: mean=2.409836, mode=2
computing coordinates
```

*Figura 28 Cálculo desarrollado por LLE  
Fuente: Autoría propia*

Entre los resultados obtenidos de la función *lle* se puede observar en la variable *Y* las dos dimensiones en las cuales queda reducida el dataset test de la BDD4 (Ver Figura 29).

```
> lle.test$Y
      [,1]      [,2]
[1,] -0.40118254 -1.69052502
[2,] -0.39800161 -1.68626384
[3,] -0.39746214 -1.68915927
[4,] -0.36293850 -1.63949220
[5,]  0.29407768 -0.40829545
[6,] -0.56499452  1.22349617
[7,] -1.04913586  1.26991237
[8,] -0.94277235  1.28852901
[9,] -1.18124578  1.25989497
[10,]  0.57582239  0.29943252
[11,]  0.13965113 -0.84262615
[12,]  0.09094074 -0.93116913
[13,]  0.04810821 -1.03101642
[14,]  0.05801659 -1.03551395
[15,]  0.03320116 -1.07983760
[16,]  0.05274827 -1.04570486
[17,] -0.06559751 -1.28871316
[18,] -0.07913839 -1.31425208
[19,] -0.08724657 -1.32617139
[20,] -0.17318481 -1.48546352
```

*Figura 29 Valores de las dos dimensiones resultantes de LLE  
Fuente: Autoría propia*

Finalmente se hace la representación visual de los resultados obtenidos por el método de RD como se muestra en la Figura 30:



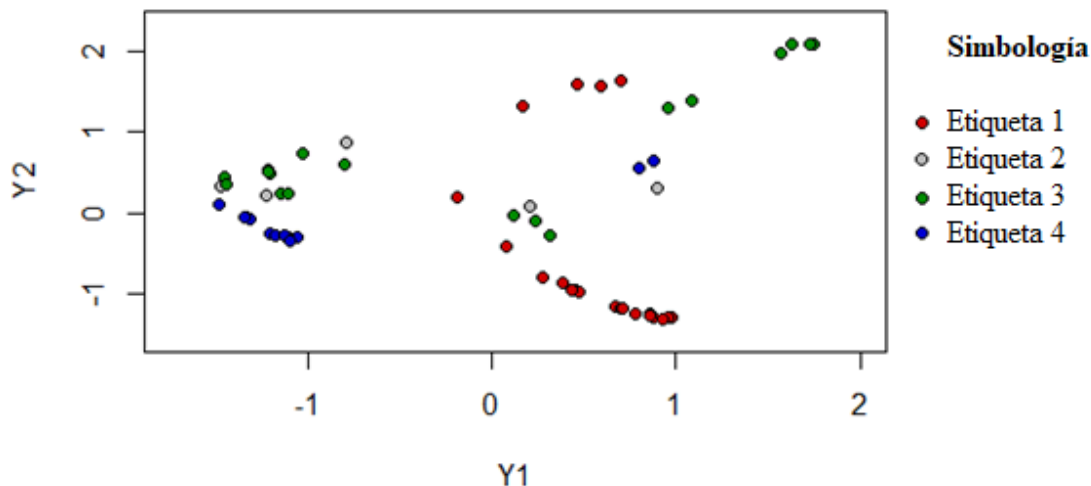


Figura 30 LLE aplicado en BDD4  
Fuente: Autoría propia

### 3.4.4 SNE

Para el desarrollo de SNE en RStudio, es importante el uso de la librería *Rdimtools* que contiene varios métodos utilizados para la reducción de dimensionalidad, entre ellos se encuentra SNE. Luego de haber cargado la librería correspondiente, se hace un llamado a la función *do.sne* para poder establecer los parámetros necesarios para la correcta reducción de dimensión del dataset ingresado. Como primer parámetro se debe establecer el dataset con el que se va a trabajar, luego el número de dimensiones a las cuales se va a reducir el conjunto original, el nivel deseado de perplejidad cuyo valor recomendable es entre un rango de 5-50, para este caso se ha aplicado un valor de 30, el número máximo de iteraciones, y finalmente si hace uso de *pca* como paso preliminar para el cálculo de SNE.

```
sne.test = do.sne(test_set, ndim=2, perplexity = 30, maxiter = 100, pca = TRUE)
```

El objeto que se tiene como respuesta almacena los valores de las dos dimensiones en el elemento *Y* cómo se puede verificar a en la Figura 31:

```

> sne.test$Y
      [,1]      [,2]
[1,] -0.424318028 0.76806875
[2,] -0.212154746 1.38770038
[3,]  0.132346742 0.89110697
[4,]  0.322999935 1.08408448
[5,]  0.023987431 0.62222780
[6,] -0.583938262 0.55371113
[7,]  0.452831377 0.85533024
[8,]  0.188659859 0.55697532
[9,]  0.304409323 1.14344355
[10,] 0.353715669 0.79685239
[11,] -0.340381979 0.85455580
[12,] -0.145796481 0.96597195
[13,] -0.095970492 0.64207451
[14,] -0.163422223 1.10303288

```

Figura 31 Valores de las dos dimensiones resultantes de SNE.

Fuente: Autoría propia

Con estos valores obtenidos se pueden graficar los resultados para poder visualizar la RD realizada por este método como se muestra en la Figura 32:

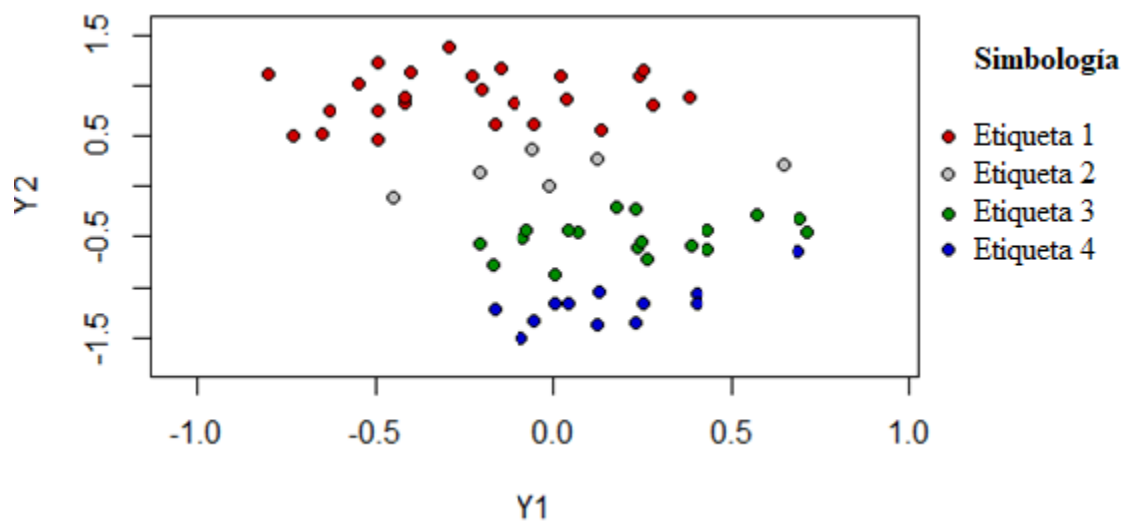


Figura 32 SNE aplicado en BDD4

Fuente: Autoría propia

### 3.4.5 T-SNE

En *t*-SNE a diferencia de PCA en donde se debe primero calcular los componentes principales y luego analizar el porcentaje de información que contiene cada uno de ellos, se debe especificar de antemano el número de dimensiones a las cuales se tiene que reducir el espacio original. En RStudio se hace uso de la librería *Rtsne* el cual contiene una implementación del algoritmo *t*-SNE mediante la aproximación Barnes-Hut el cual lo hace más veloz y se lo puede aplicar a datos con miles de observaciones.

Al emplear la función *Rtsne* se debe establecer los parámetros necesarios que este necesita, primero la matriz de datos, luego el número de dimensiones a las cuales se quiere reducir el espacio original, la perplejidad que debe ser menor que el número de observaciones totales que tiene el set de datos, cuyo valor recomendable es entre un rango de 5-50, para este caso se ha aplicado un valor de 30, y finalmente theta el cual tiene un valor predeterminado de 0.5 para *t*-SNE.

El objeto que se tiene como respuesta almacena los valores de las dos dimensiones en el elemento *Y* cómo se puede verificar a en la Figura 33:

```

> tsne.test$Y
      [,1]      [,2]
[1,] -1.7918859 -19.814065
[2,] -1.6350870 -19.810396
[3,] -1.8555740 -19.752745
[4,] -1.6509669 -19.608996
[5,] -1.8615368 -15.390979
[6,] -2.5225651 -13.441694
[7,] -2.5077613 -13.465470
[8,] -2.5088213 -13.463725
[9,] -2.5126630 -13.455657
[10,] -1.8810925 -14.958583
[11,] -1.8353415 -16.150332
[12,] -1.3887898 -16.301329
[13,] -1.2224315 -16.652659
[14,] -1.4767351 -16.706351
[15,] -2.0781063 -16.683635
[16,] -2.0360963 -16.689171
[17,] -1.9975106 -17.415093
[18,] -1.3985460 -17.448247
[19,] -1.2274075 -17.520364
[20,] -2.1176689 -18.056490

```

Figura 33 Valores de las dos dimensiones resultantes de tSNE  
Fuente: Autoría propia

Con estos valores obtenidos se pueden graficar los resultados para poder visualizar la RD realizada por este método como se muestra en la Figura 34:

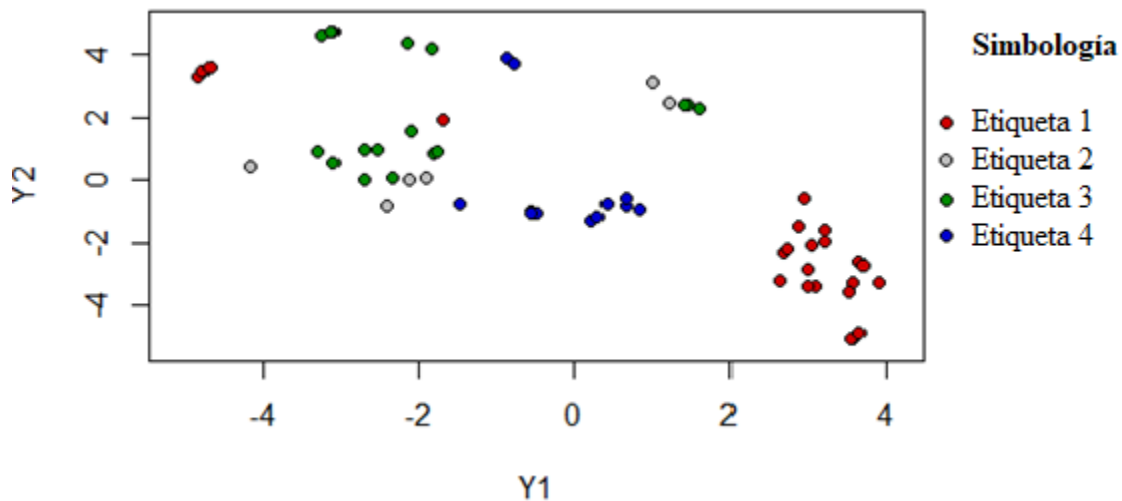
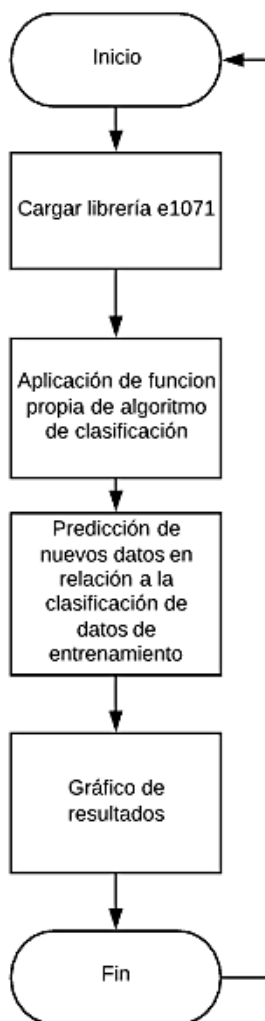


Figura 34 tSNE aplicado en BDD4  
Fuente: Autoría propia

### 3.5 Clasificación

Aquí se realiza la implementación de cada uno de los cuatro algoritmos de clasificación con sus respectivos gráficos representativos de los resultados obtenidos (Ver figura 35).



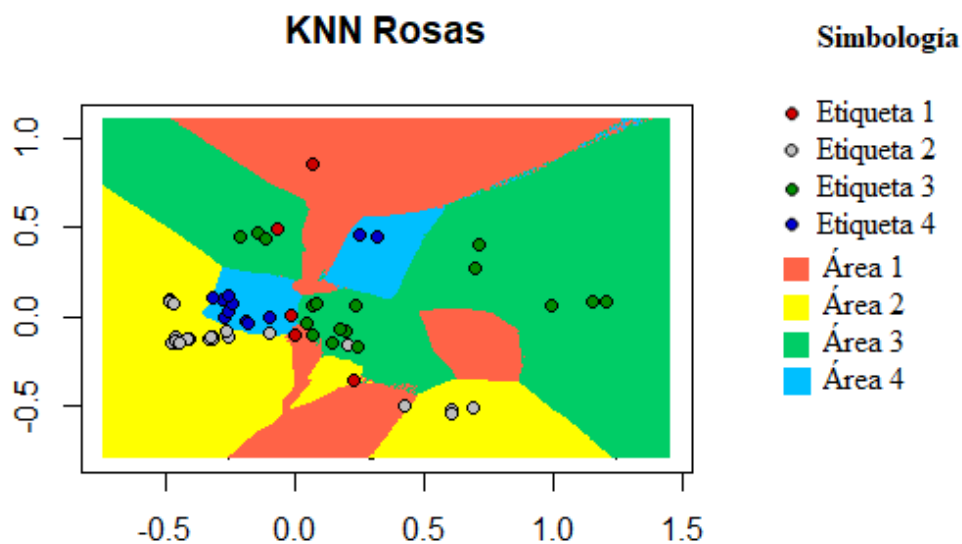
*Figura 35 Diagrama de flujo de proceso de implementación de A. de Clasificación.  
Fuente: Autoría propia*

### 3.5.1 k-NN

Para poder construir el clasificador k-NN se hace uso de la función *knn* del paquete *class*, y como parámetros se incluyen los datos de entrenamiento, los datos de prueba, la variable dependiente del set de entrenamiento y el número de vecinos.

```
Classknn = knn(train = data.entrenamiento[,-3], test = data.test[,-3],
               cl=data.entrenamiento[,3], k = 3)
```

Al obtener esta respuesta se procede finalmente a realizar el gráfico respectivo al haber aplicado este algoritmo de clasificación, como en la Figura 36:



*Figura 36 k-NN aplicado en BDD4*  
*Fuente: Autoría propia*

### 3.5.2 SVM

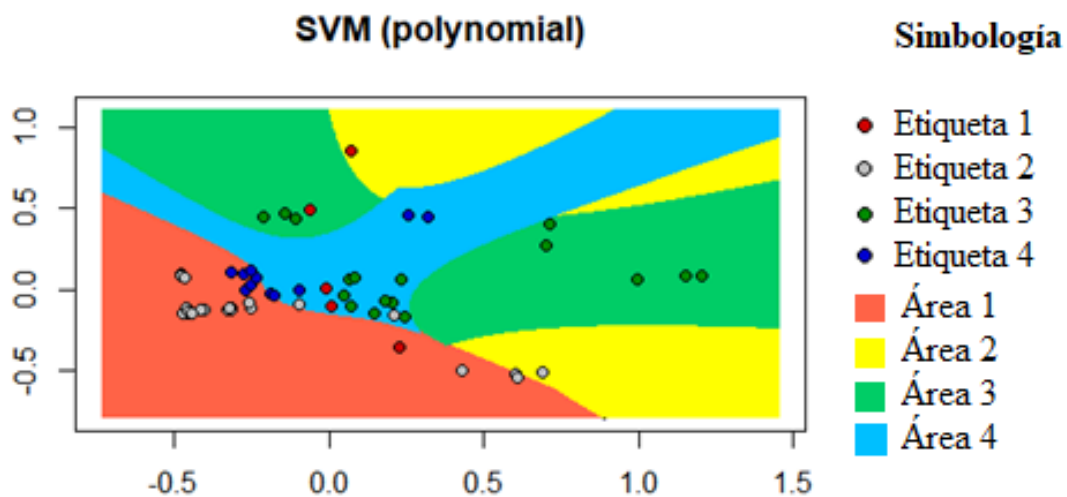
Como primer paso, se carga la librería `e1071` la cual contiene algoritmos necesarios para obtener modelos de clasificación, entre ellos SVM. Para poder obtener la clasificación de nuestros datos de entrenamiento, se hace el llamado a la función `svm`, la cual necesita como parámetros la variable independiente de la base de datos, los datos de entrenamiento, el tipo de función que va a realizar la cual es como clasificador y por último el kernel que se va a aplicar, es decir, el tipo de separación de los datos.

```
classifier = svm(formula = Clase ~ ., data = data.entrenamiento, type = 'C-classification',  
                kernel = 'polynomial')
```

Luego de trabajar con los datos de entrenamiento se procede a realizar la predicción sobre los valores del conjunto de prueba, para esto, se emplea el método `predict` donde se introduce como parámetros la respuesta obtenida por la clasificación de los datos de entrenamiento, y los datos de prueba.

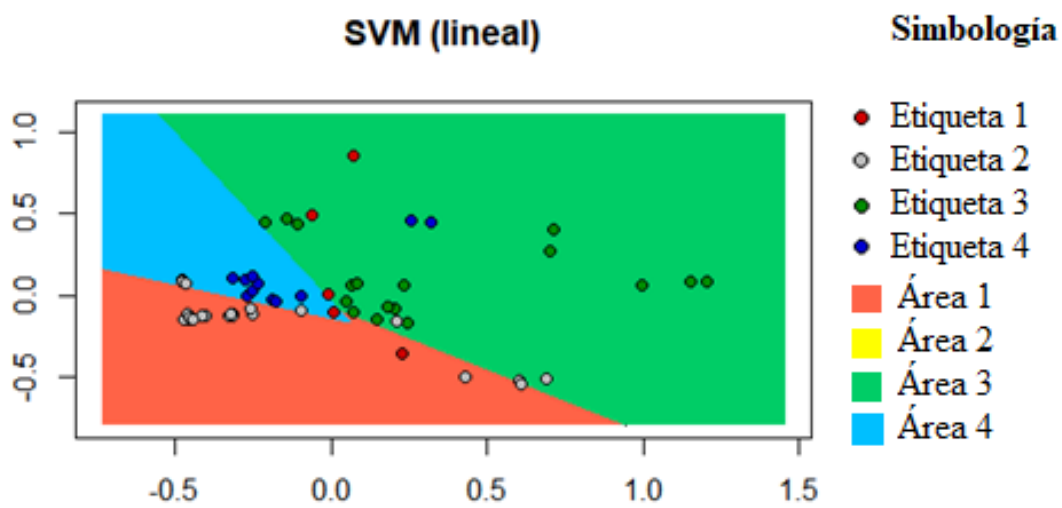
```
y_pred = predict(classifier, newdata = data.test[-3])
```

Al obtener todas estas respuestas se procede finalmente a realizar el gráfico respectivo al haber aplicado este algoritmo de clasificación, a continuación, se muestra los resultados de este algoritmo con tres diferentes tipos de kernel. Como se puede ver el algoritmo SVM con kernel polinomial en las Figuras 37:



*Figura 37 Clasificador SVM con kernel polynomial en BDD4  
Fuente: Autoría propia*

En la figura 38 se observa el clasificador SVM con kernel lineal:



*Figura 38 Clasificador SVM con kernel lineal en BDD4  
Fuente: Autoría propia*

En la figura 39 se observa el clasificador SVM con kernel sigmoide:



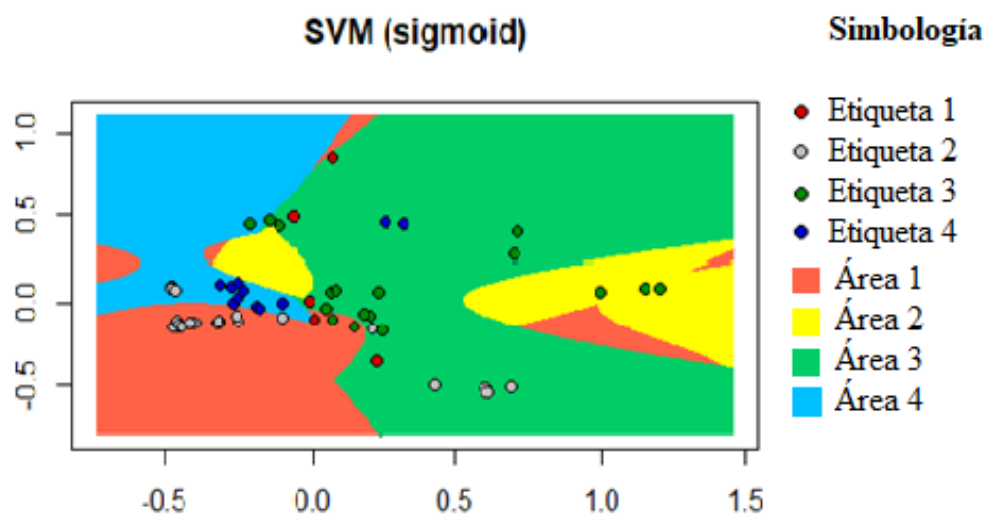


Figura 39 Clasificador SVM con kernel sigmoid en BDD4  
Fuente: Autoría propia

### 3.5.3 Clasificador Bayesiano

Para el uso de este clasificador se utiliza la función *naive\_bayes* con el set de datos de entrenamiento. Esta función pide como parámetros la variable independiente para clasificar los datos a usar, la cual se la determina de la siguiente manera: *clase* ~ .

Además, se debe especificar los datos de las variables dependientes del set de entrenamiento, y finalmente el factor donde se encuentran los valores de las clases como se muestra en la Figura 40.



### 3.5.4 Árbol de decisión

Se usa la función *rpart* para la implementación de este clasificador. Esta función requiere una fórmula para especificar la variable independiente de la clasificación. La fórmula que se usa es tipo `~ .`, la cual expresa que se realizará la clasificación de la variable *clase*, y definiendo a las demás variables como predictoras. Además, necesita que se defina el set de entrenamiento como se expresa a continuación:

```
Classarbol = rpart(formula = class~., data=training_set)
```

La variable obtiene como resultado el esquema de clasificación realizada al dataset de entrenamiento como se ve en la Figura 42.

```
1) root 240 140 1 (0.41666667 0.09166667 0.30000000 0.19166667)
2) pc1< 0.4555174 175 84 1 (0.52000000 0.09714286 0.12000000 0.26285714)
4) pc2< 0.3068764 137 46 1 (0.66423358 0.03649635 0.00729927 0.29197080)
8) pc1< -0.1926997 67 2 1 (0.97014925 0.00000000 0.00000000 0.02985075) *
9) pc1>=-0.1926997 70 32 4 (0.37142857 0.07142857 0.01428571 0.54285714)
18) pc2< -0.3188672 14 4 1 (0.71428571 0.28571429 0.00000000 0.00000000) *
19) pc2>=-0.3188672 56 18 4 (0.28571429 0.01785714 0.01785714 0.67857143)
38) pc2>=-0.2225608 33 16 4 (0.42424242 0.03030303 0.03030303 0.51515152)
76) pc2< -0.03666352 22 8 1 (0.63636364 0.04545455 0.04545455 0.27272727) *
77) pc2>=-0.03666352 11 0 4 (0.00000000 0.00000000 0.00000000 1.00000000) *
39) pc2< -0.2225608 23 2 4 (0.08695652 0.00000000 0.00000000 0.91304348) *
5) pc2>=0.3068764 38 18 3 (0.00000000 0.31578947 0.52631579 0.15789474)
10) pc2>=0.5863052 10 1 2 (0.00000000 0.90000000 0.00000000 0.10000000) *
11) pc2< 0.5863052 28 8 3 (0.00000000 0.10714286 0.71428571 0.17857143)
22) pc1< 0.02819864 21 1 3 (0.00000000 0.04761905 0.95238095 0.00000000) *
23) pc1>=0.02819864 7 2 4 (0.00000000 0.28571429 0.00000000 0.71428571) *
3) pc1>=0.4555174 65 14 3 (0.13846154 0.07692308 0.78461538 0.00000000)
6) pc2< -0.3691982 12 3 1 (0.75000000 0.16666667 0.08333333 0.00000000) *
7) pc2>=-0.3691982 53 3 3 (0.00000000 0.05660377 0.94339623 0.00000000) *
```

Figura 42 Esquema de clasificación del entrenamiento de datos

Fuente: Autoría propia

Se hace uso de la función *predict* con el set de prueba para generar un vector que contenga los valores predichos por el modelo que se ha entrenado anteriormente, además se especifica el parámetro *type = "class"*. Y como resultado se puede observar en la Figura 43 los valores de predicción obtenidos y almacenados en la variable *y\_predarbol*.

```
y_predarbol = predict(classarbol, newdata = data, test[-3],type = 'class')
```

```
> y_predarbol
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
1 1 1 1 4 1 4 3 1 1 1 4 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 2 1 3 1
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61
3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 1 4 4 2 4 4 4 1 4 1 4 1 1 1
Levels: 1 2 3 4
```

Figura 43 Resultado de la predicción

Fuente: Autoría propia

Finalmente se realiza el gráfico correspondiente de los resultados obtenidos por la clasificación mediante el algoritmo árbol de decisión en el dataset de BDD4 reflejado en la Figura 44.

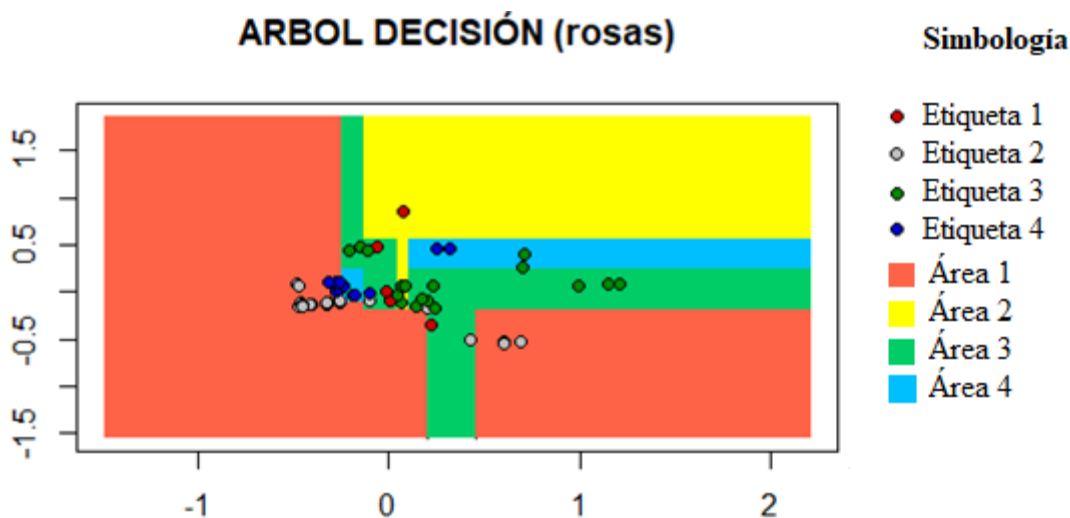


Figura 44 Clasificador Árbol de decisión en BDD4

Fuente: Autoría propia

### 3.6 Matriz de Confusión

Es importante realizar la evaluación del rendimiento de cada uno de los algoritmos. Debido a esto, se hace uso de la matriz de confusión que es una herramienta muy útil que permite la visualización del desempeño realizado por el algoritmo de aprendizaje supervisado. Esta matriz brinda el conteo de aciertos y errores de las clases que se han clasificado, estos conteos se dividen

entre verdaderos positivos (TP), falsos negativos (FN), falsos positivos (FP) y verdaderos negativos (TN). Con estos valores también se pueden calcular distintas métricas de las cuales nos interesa el valor que nos entrega *Acurracy*, que es el porcentaje total de aciertos del modelo que está siendo evaluado, la misma que se calcula de la siguiente manera con la Ecuación 1:

$$Acurracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

En RStudio para obtener la matriz de confusión se realiza un cálculo usando los factores de clasificación para construir una tabla con los conteos de aciertos y errores.

```
cm = table(data.test[, 3], y_pred)
```

Los resultados de estos cálculos se pueden visualizar en la Figura 45.

	y_pred			
	1	2	3	4
1	25	0	0	0
2	3	1	2	0
3	5	0	13	0
4	10	2	0	0

Figura 45 Tabla de conteos realizados con clasificador SVM en BDD4  
Fuente: Autoría propia

Para luego mediante la función *confusionMatrix* realiza el cálculo de las distintas métricas en base a la tabla calculada anteriormente.

```
cmat = confusionMatrix( cm )
```

Los resultados de estos cálculos se pueden visualizar en la Figura 46.

```

Overall Statistics

      Accuracy : 0.6393
    95% CI : (0.5063, 0.7584)
  No Information Rate : 0.7049
  P-Value [Acc > NIR] : 0.8951

      kappa : 0.4309

  McNemar's Test P-value : NA

Statistics by Class:

                Class: 1 Class: 2 Class: 3 Class: 4
Sensitivity      0.5814  0.33333  0.8667   NA
Specificity      1.0000  0.91379  0.8913  0.8033
Pos Pred Value   1.0000  0.16667  0.7222   NA
Neg Pred Value   0.5000  0.96364  0.9535   NA
Prevalence        0.7049  0.04918  0.2459  0.0000
Detection Rate   0.4098  0.01639  0.2131  0.0000
Detection Prevalence 0.4098  0.09836  0.2951  0.1967
Balanced Accuracy 0.7907  0.62356  0.8790   NA

```

*Figura 46 Matriz de confusión con clasificador SVM en BDD4  
Fuente: Autoría propia*

Finalmente, con la Matriz de Confusión se puede expresar que la precisión obtenida al realizar la implementación del algoritmo de clasificación SVM en el set de datos BDD4 es de un 63.93%. Al repetir este proceso en todos los algoritmos podemos realizar una comparación de la precisión de cada uno de ellos con cada set de datos.

## **CAPÍTULO IV. PRUEBAS DE FUNCIONAMIENTO Y RESULTADOS**

En este capítulo se realizan las pruebas finales de todos los métodos de RD y algoritmos de clasificación con los sets de datos especificados, en cada literal de las figuras de este capítulo está representado un método de RD, en PCA las variables que representan las dos dimensiones resultantes de este proceso se las denomina PC1 y PC2, mientras que, en el resto de los métodos se los denomina Y1 y Y2. Además, se presenta la tabla resumen en la que se encuentra el total de porcentajes de precisión obtenidos de las pruebas realizadas. El capítulo finaliza con las conclusiones y recomendaciones del presente trabajo de titulación.

Es de importancia el poder evidenciar de forma visual la comparación de cada método de RD realizado en RStudio, evaluar la calidad de los aspectos destacados de cada uno de ellos en representaciones resultantes de espacios de menor dimensión, además, analizar la precisión obtenida de los algoritmos de clasificación.

### **4.1 Pruebas con métodos de RD.**

En la Figura 47 se puede ver los resultados obtenidos en el set de datos BDD1 con los cinco tipos de representaciones en baja dimensión planteados en el capítulo anterior. Métodos espectrales como LLE, LE donde gráficamente no se puede ver una muy buena agrupación entre las distintas etiquetas que conforman el set de datos, métodos lineales como PCA, métodos estocásticos basados en divergencias como SNE y *t*-SNE PCA donde se puede ver una mejor agrupación y donde se diferencian claramente cada grupo de etiquetas.

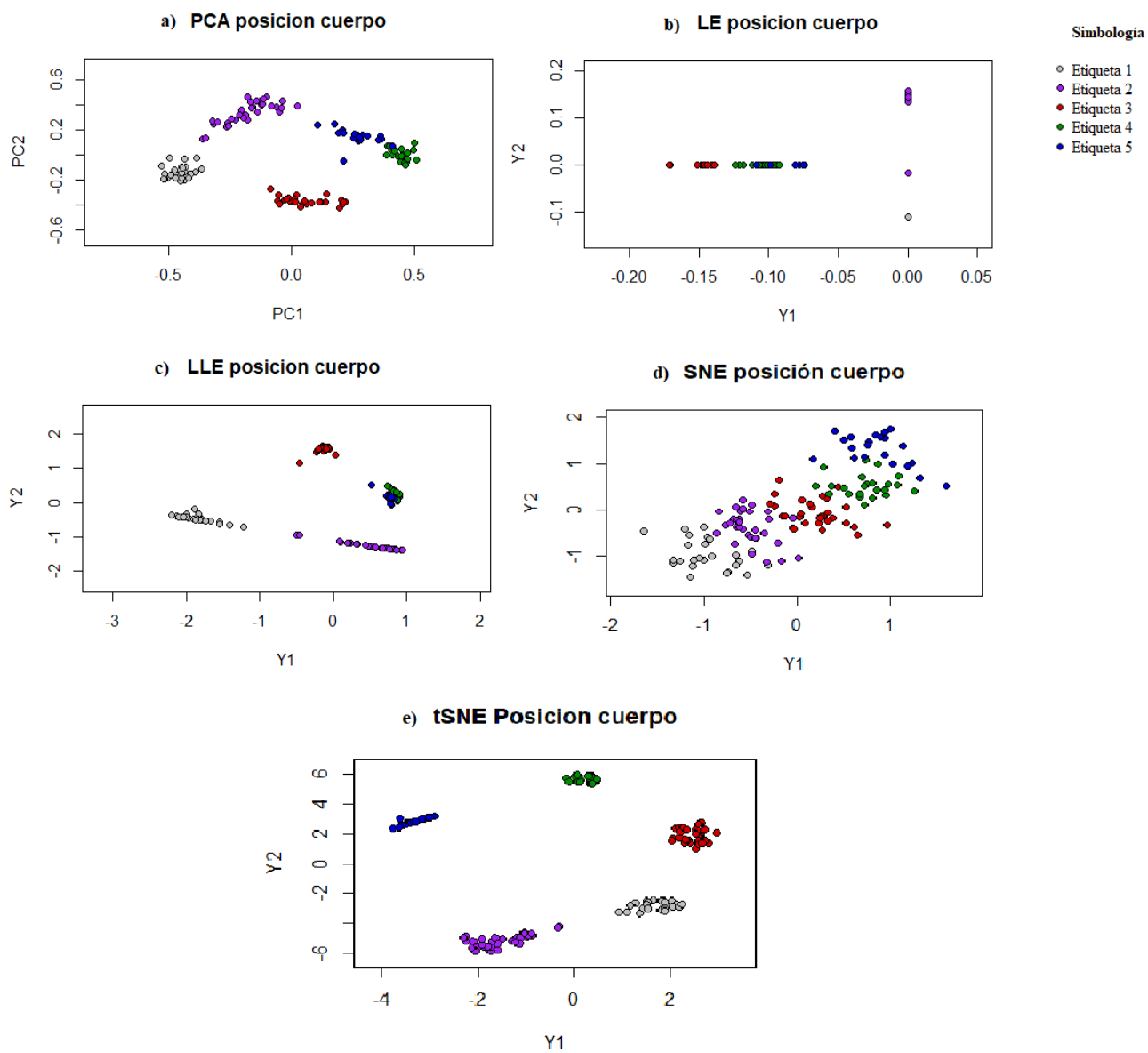


Figura 47 Representación de los métodos de reducción de dimensionalidad en BDD1  
Fuente: Autoría propia



En la Figura 48 se presenta los resultados obtenidos en el set de datos BDD2 con los cinco tipos de representaciones en baja dimensión planteados en el capítulo anterior. En cada literal de este gráfico se puede apreciar la representación de los datos en baja dimensión, además de una excelente agrupación en los literales b), d) y e) de las distintas etiquetas correspondientes al set de datos.

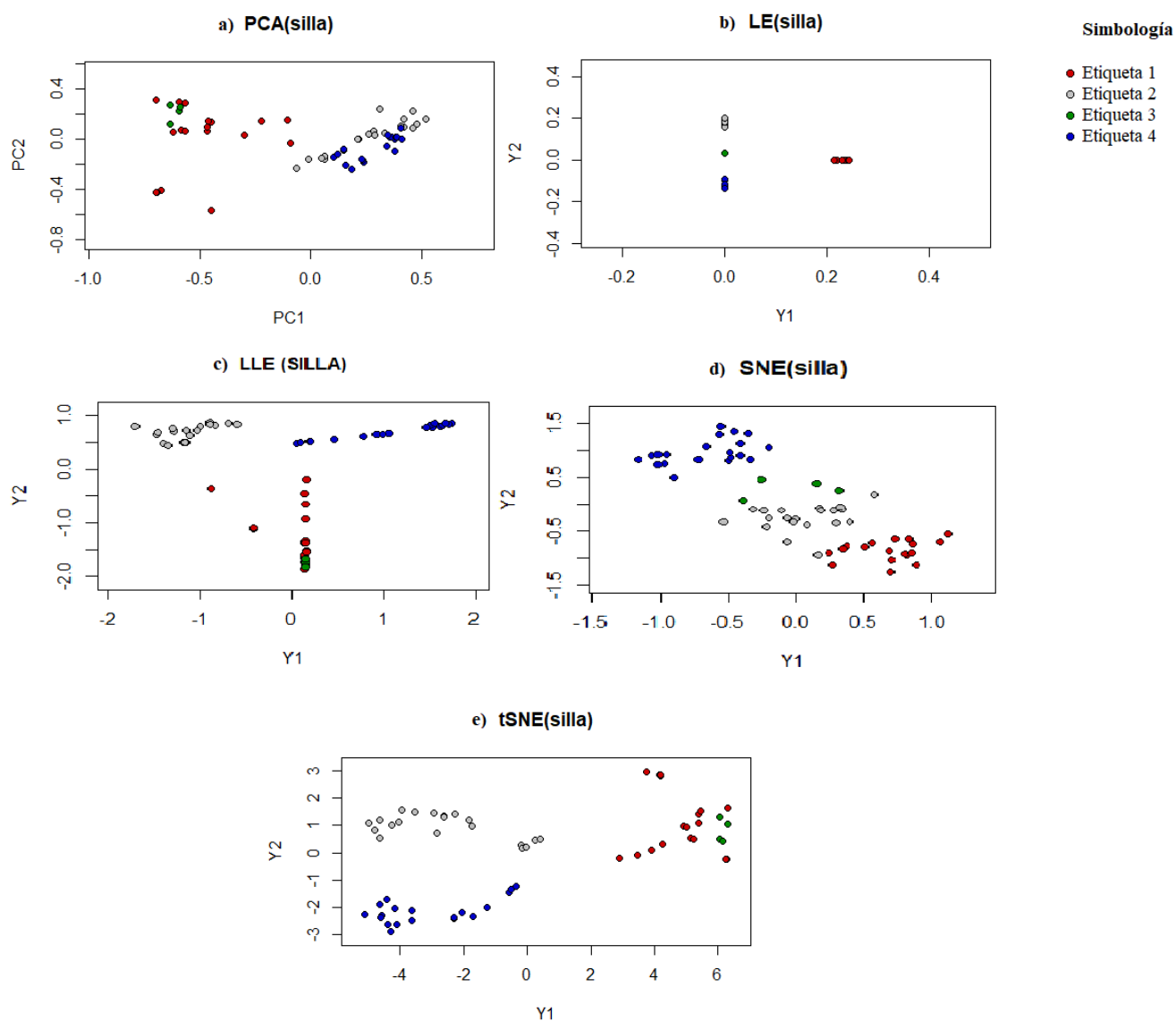


Figura 48 Representación de los métodos de reducción de dimensionalidad en BDD2  
Fuente: Autoría propia

En la Figura 49 se reflejan los resultados obtenidos en el set de datos BDD3 con los cinco tipos de representaciones en baja dimensión planteados en el capítulo anterior. Métodos espectrales como LLE, LE, lineales como PCA, estocásticos basados en divergencias como SNE y  $t$ -SNE. En los literales a), b) y e) se puede apreciar claramente las mejores agrupaciones diferenciando fácilmente los tres tipos de etiquetas del set de datos, identificados cada uno por su color.

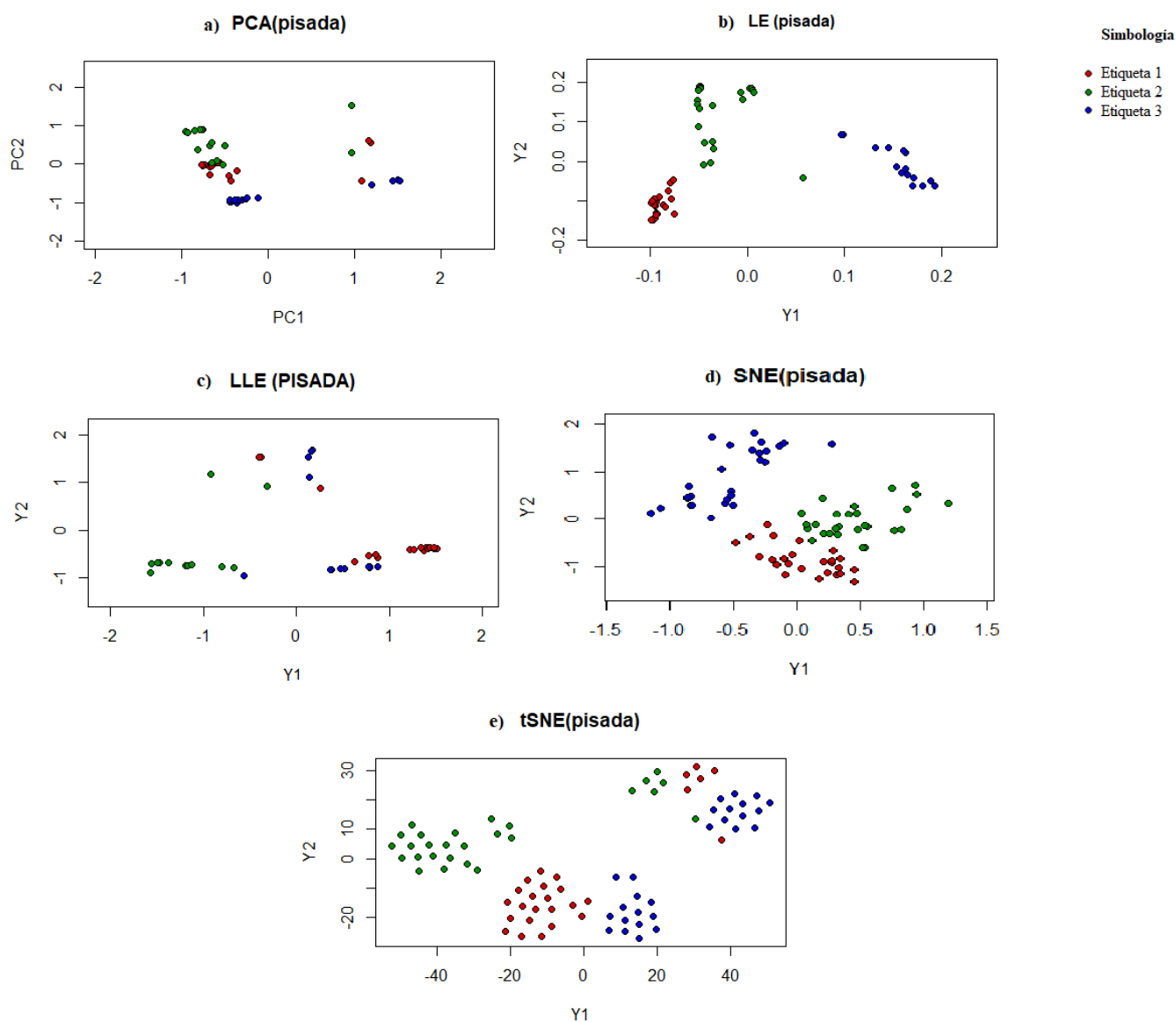


Figura 49 Representación de los métodos de reducción de dimensionalidad en BDD3  
Fuente: Autoría propia

En la Figura 50 se observan los resultados obtenidos en el set de datos BDD4 con los cinco tipos de representaciones en baja dimensión planteados en el capítulo anterior. Se puede comprobar que en los literales b) y d) existe una mejor agrupación de los cuatro distintos grupos de etiquetas a diferencia de los resultados con los otros métodos de RD.

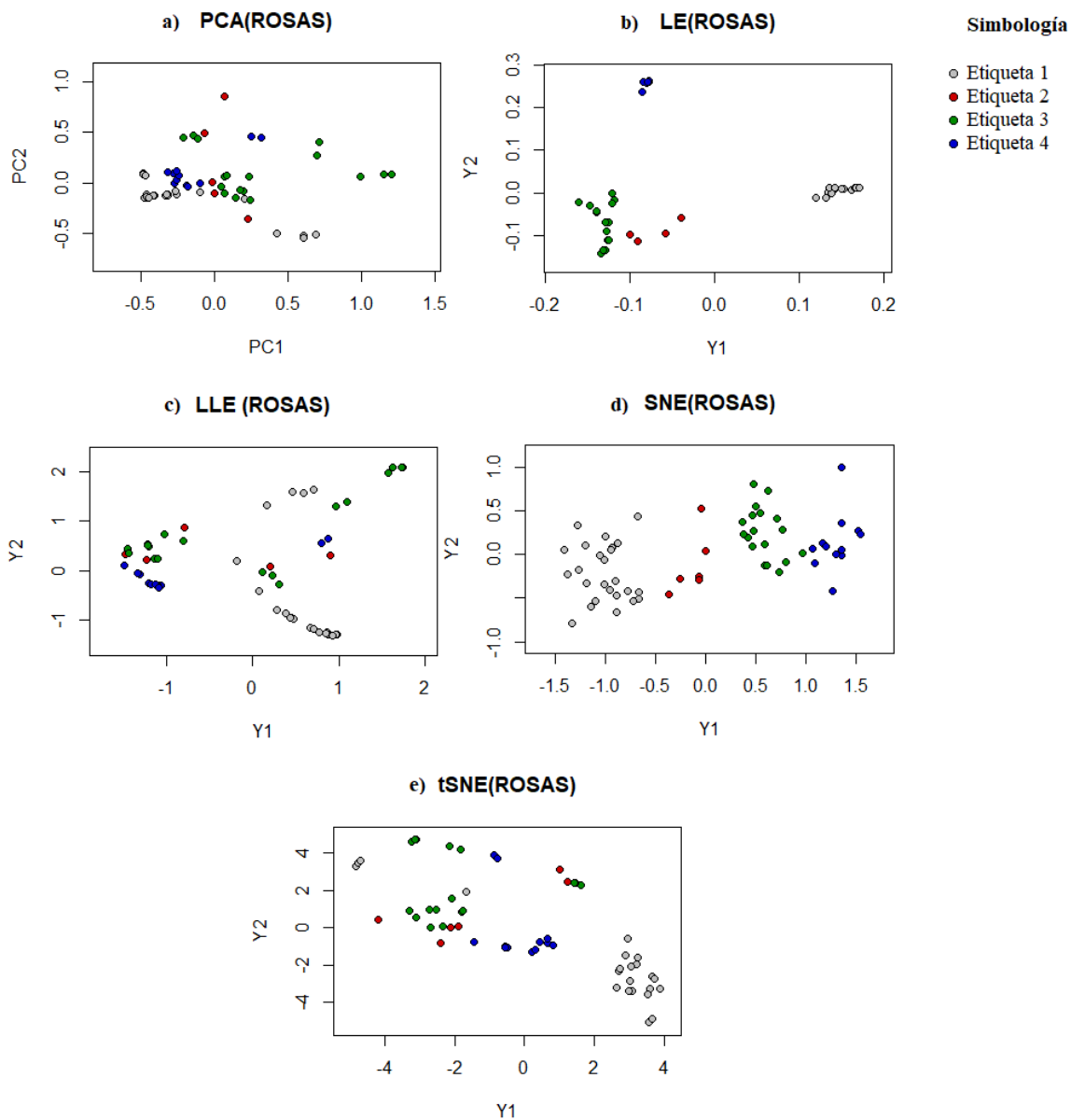


Figura 50 Representación de los métodos de reducción de dimensionalidad en BDD4  
Fuente: Autoría propia

En la Figura 51 se expresan los resultados obtenidos en el set de datos BDD5 con los cinco tipos de representaciones en baja dimensión planteados en el capítulo anterior. Se puede apreciar en este caso que, al existir solo dos tipos de etiquetas, es mucho más evidente el distinguir una separación entre las mismas, manteniéndose como en algunos de los ejemplos anteriores los literales a), c) y e) con la mejor representación del set de datos en baja dimensión.

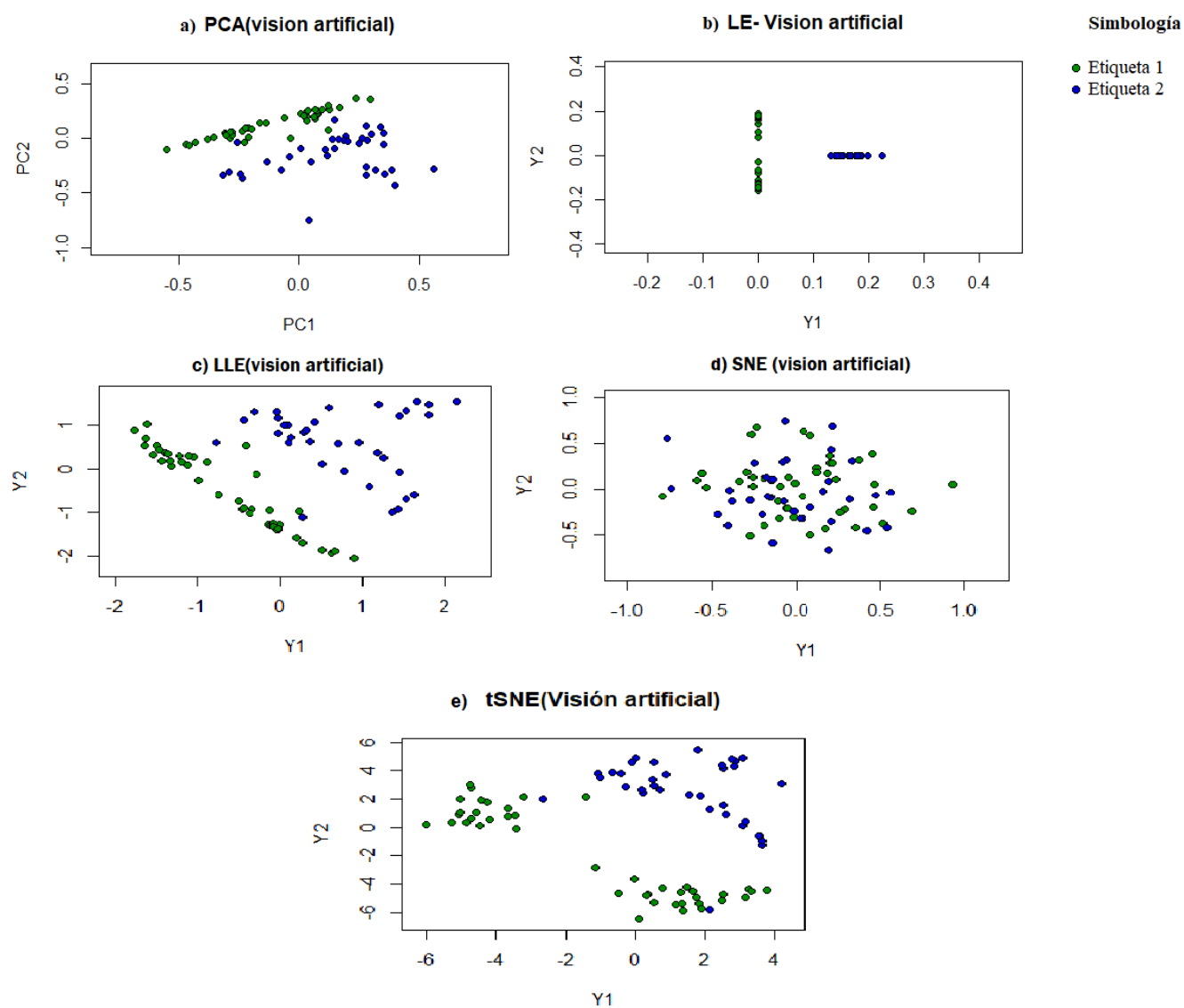


Figura 51 Representación de los métodos de reducción de dimensionalidad en BDD4  
Fuente: Autoría propia

Se observa que los métodos estocásticos como SNE y *t*SNE tienen buenos resultados ya que se puede observar claramente la separación de cada una de las clases o etiquetas, de la misma manera con el método basado en disimilitudes PCA al utilizar la varianza.

Los métodos basados en similitudes como LLE y LE al igual que los anteriores métodos realizan una representación de baja dimensión. Pero, cabe destacar que LE y LLE, presentan cierta transformación horizontal en su representación visual, lo que puede indicar que este método tiene una pequeña predisposición hacia el concepto de la preservación de los datos locales. En la Tabla 9 se presenta los resultados de los métodos de RD relacionando los niveles de agrupación y representación obtenidos de los datos en dos dimensiones, los cuales están considerados como alto, medio y bajo.

Tabla 9

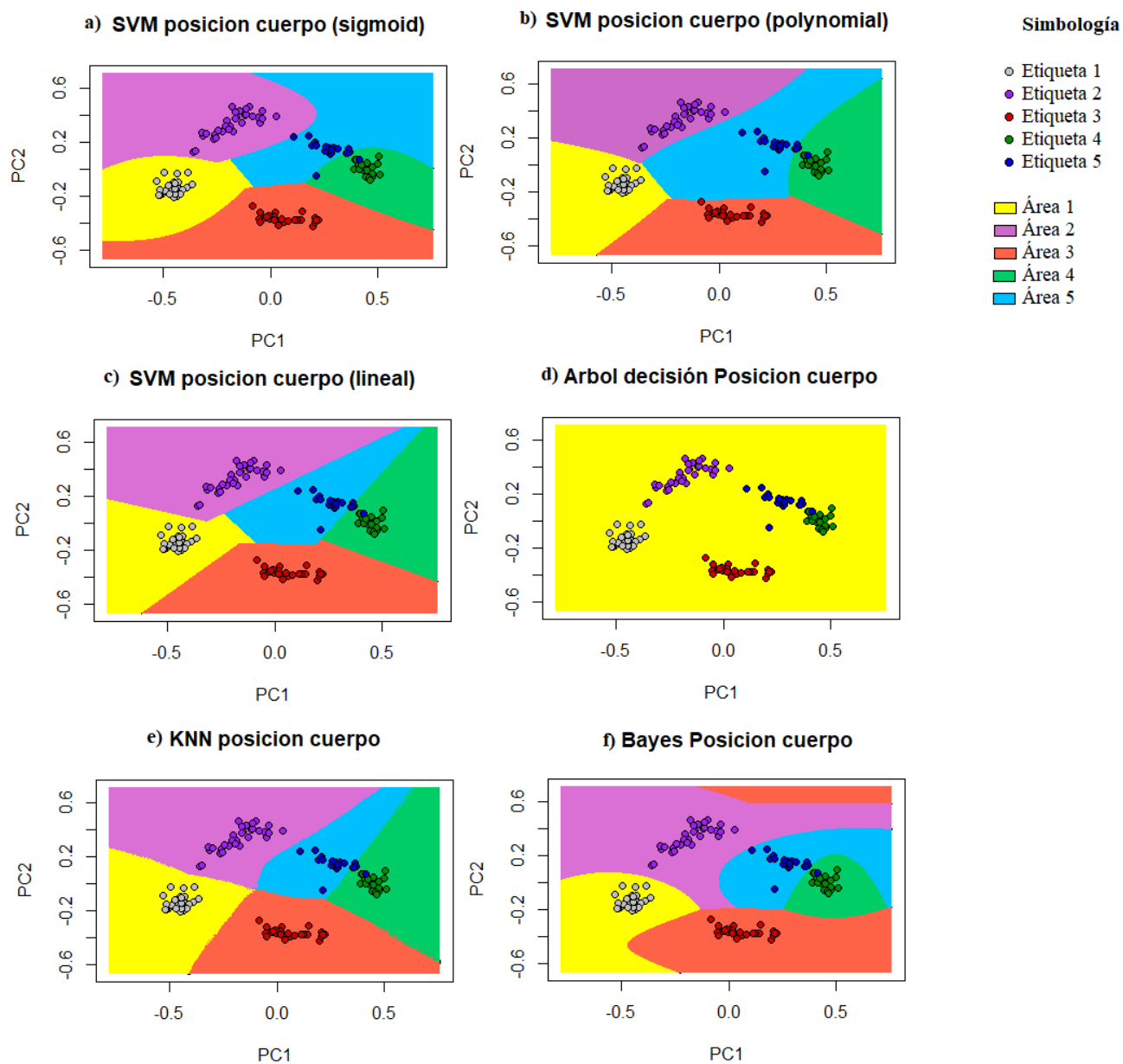
Tabla resumen de resultados de los métodos de RD.

<b>BDD</b>	<b>PCA</b>	<b>LE</b>	<b>LLE</b>	<b>SNE</b>	<b><i>t</i>SNE</b>
<b>ROSAS</b>	Bajo	Alto	Bajo	Alto	Bajo
<b>POSICIÓN</b>	Alto	Bajo	Alto	Alto	Alto
<b>SILLA</b>	Medio	Bajo	Medio	Alto	Alto
<b>PISADA</b>	Medio	Alto	Bajo	Medio	Alto
<b>VISIÓN</b>	Alto	Bajo	Alto	Bajo	Alto

*Fuente: Autoría propia*

#### **4.2 Pruebas con algoritmos de clasificación.**

De la misma manera, es importante la representación visual de los resultados obtenidos al implementar los algoritmos de clasificación con cada método de RD y BDD. En la Figura 52 se puede ver la clasificación realizada en BDD1 con PCA. Donde se evidencia claramente una clasificación muy precisa realizada para cada una de las etiquetas del set de datos, a excepción en el literal d) donde el área 1 domina todo el espacio de clasificación.



*Figura 52 Clasificación con PCA en BDD1*  
*Fuente: Autoría propia*

En la Figura 53 se presenta la clasificación realizada en BDD1 con LE, donde es muy notable la diferencia de los resultados, ya que en casi todos los literales, es decir a), b) c) e) y f) se observa además de no coincidir el área con los grupos correspondientes de las etiquetas, solo se

muestran 3 de las 5 áreas que existen para la clasificación. Al igual que en la Figura 51 se observa que el literal d) sigue predominando el área 1.

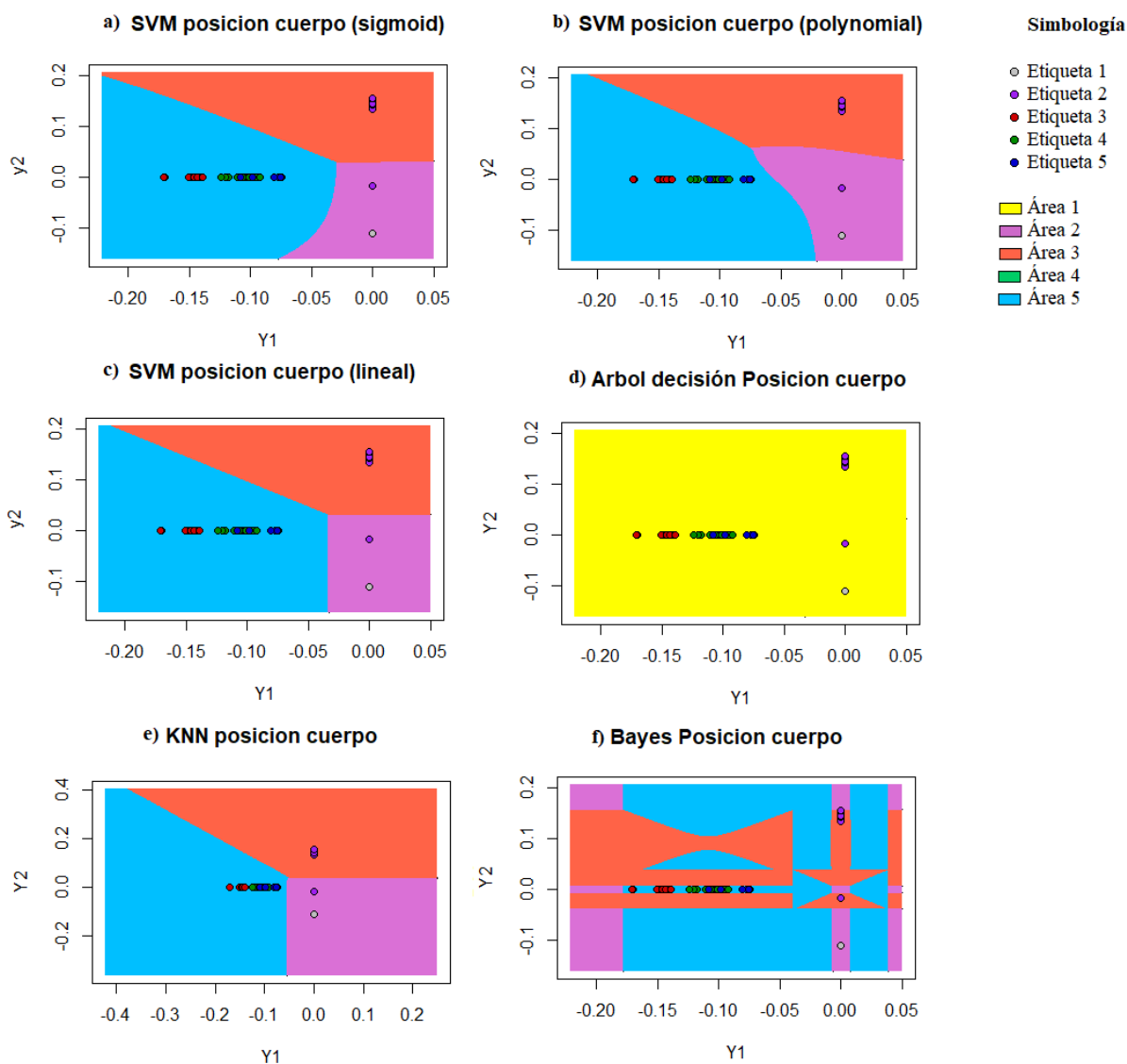


Figura 53 Clasificación con LE en BDD1

Fuente: Autoría propia

En la Figura 54 se evidencia la clasificación realizada en BDD1 con LLE, donde los resultados gráficos muestran de igual manera que existe una baja precisión de la clasificación

realizada, debido a que solo se muestran en los literales a), b), c), e) y f) de 2 a 3 áreas de las 5 que tiene este set de datos, y de igual manera el literal d) sigue siendo dominado por el área 1.

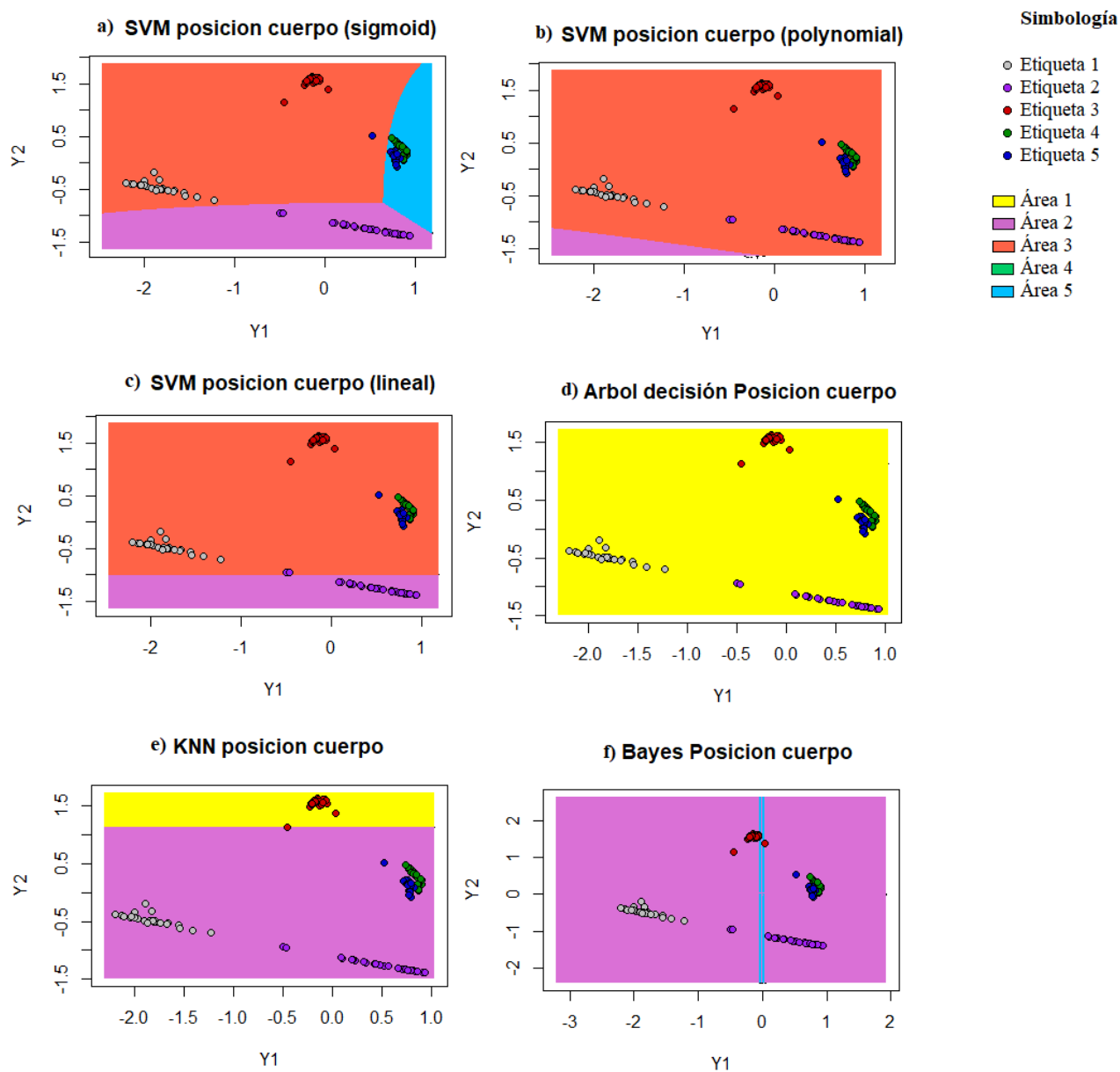


Figura 54 Clasificación con LLE en BDD1  
 Fuente: Autoría propia



En la Figura 55 se refleja la clasificación realizada en BDD1 con SNE, donde se visualiza una buena precisión de clasificación realizada en los literales a), b), c) y f). Por el contrario, en los literales d) y e) existe la dominación de una sola área representando así un bajo nivel de clasificación.

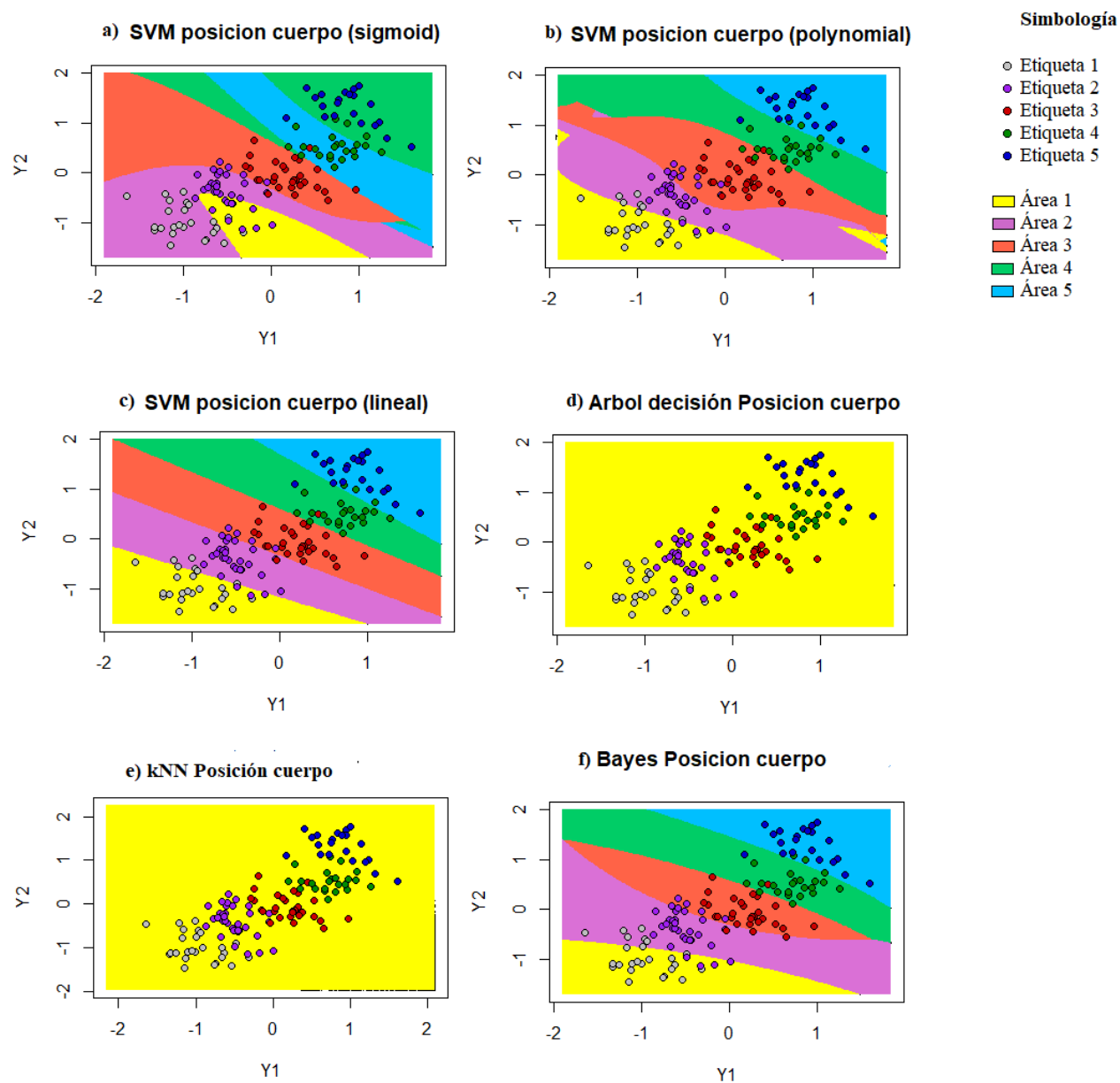


Figura 55 Clasificación con SNE en BDD1  
Fuente: Autoría propia

En la Figura 56 se puede observar la clasificación realizada en BDD1 con  $t$ -SNE, donde al igual que en las Figuras 52 y 53 correspondientes a los métodos LE y LLE solo muestran de 2 a 3 áreas de clasificación. Mientras se sigue manteniendo en el literal d) la dominación del área 1.

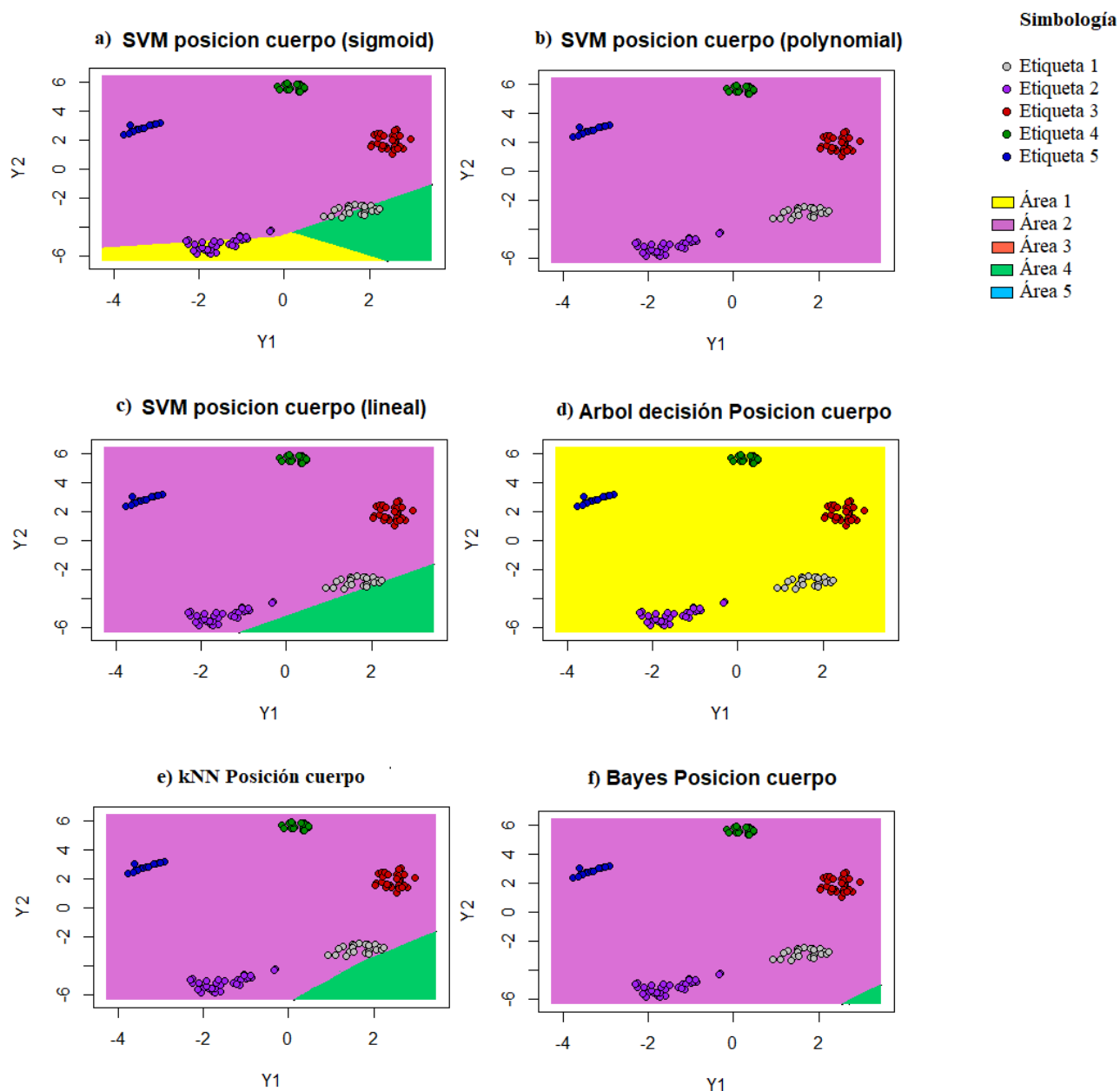


Figura 56 Clasificación con  $t$ -SNE en BDD1  
 Fuente: Autoría propia

En la Figura 57 se presenta la clasificación realizada en BDD2 con PCA, donde se puede comprobar que existe una buena clasificación realizada a excepción con el clasificador bayesiano y árbol de decisión, también es evidente que algunos puntos de las distintas etiquetas se encuentran fuera de su área correspondiente. Asimismo, cabe mencionar que se mantiene la dominación del área 1 en el gráfico del literal d) al igual que en figuras anteriores.

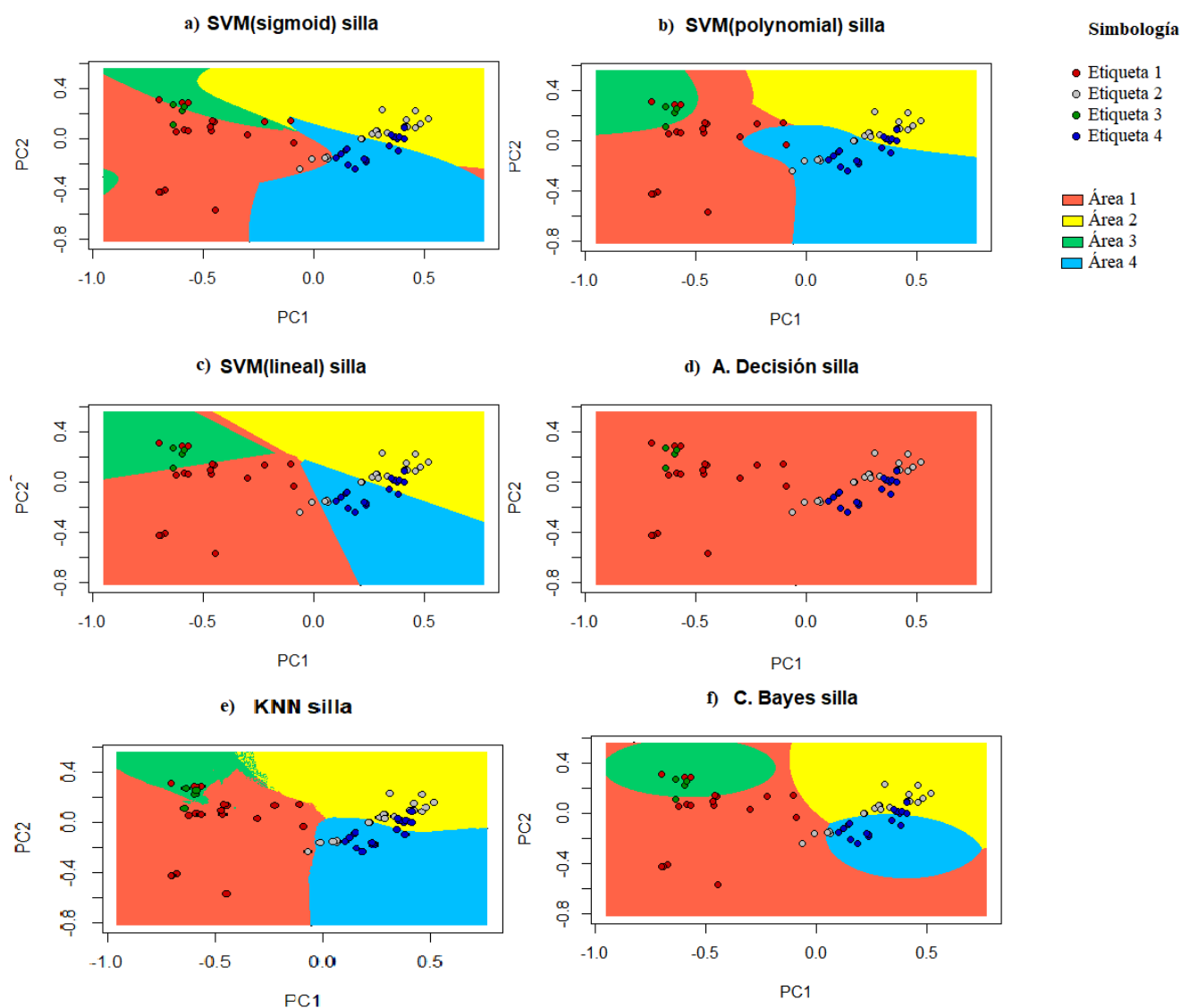


Figura 57 Clasificación con PCA en BDD2  
Fuente: Autoría propia

En la Figura 58 se muestra la clasificación realizada en BDD2 con LE, en el que existe la representación solo de 3 de las 4 áreas que corresponden al set de datos, por lo que se puede decir que existe un bajo nivel de precisión de la clasificación realizada.

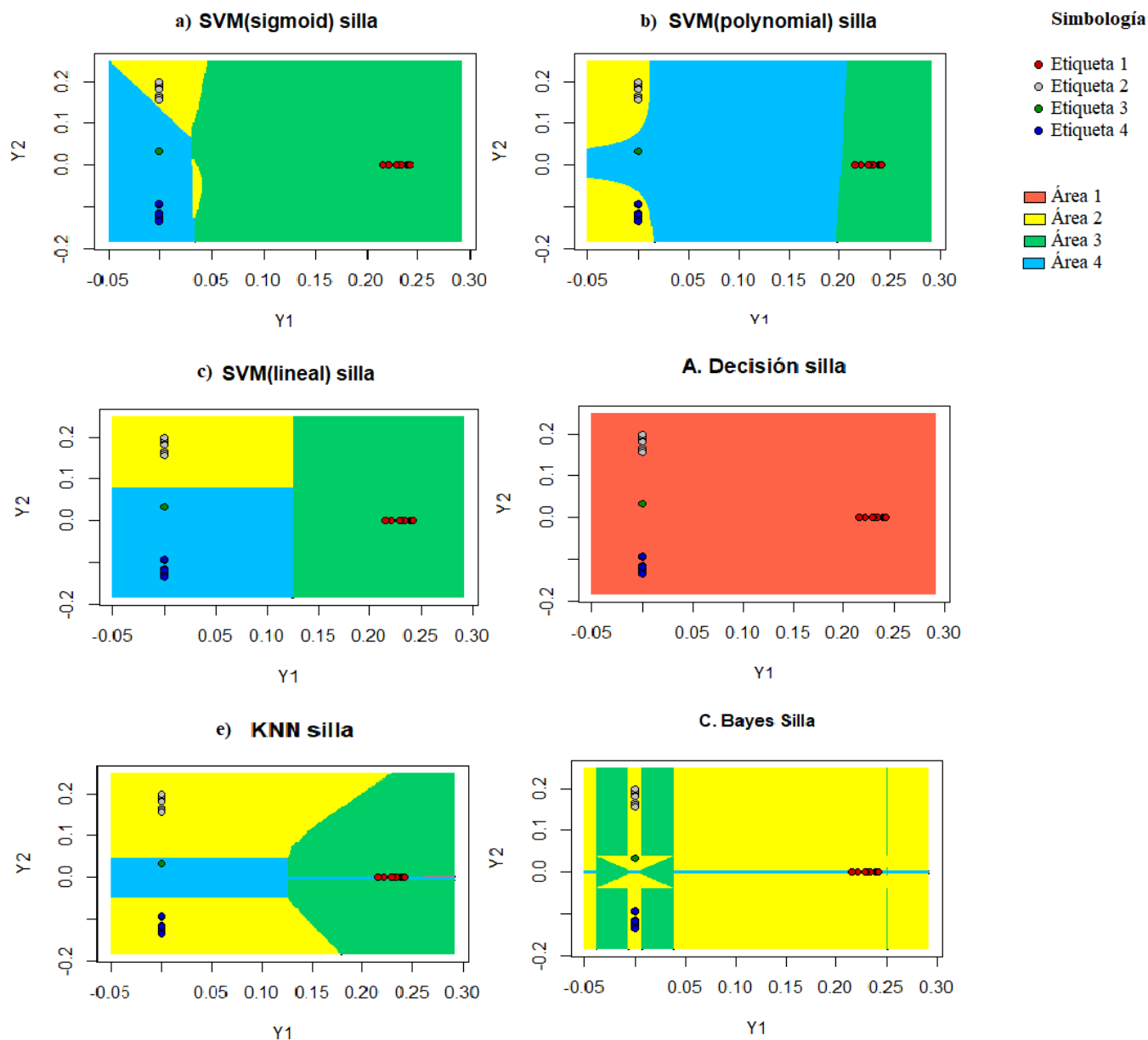


Figura 58 Clasificación con LE en BDD2  
Fuente: Autoría propia

En la Figura 59 se presenta la clasificación realizada en BDD2 con LLE, donde se observa fácilmente que solo en el literal f) correspondiente al clasificador Bayesiano se encuentran representadas todas las áreas de clasificación, aun así, se verifica que muchos de los puntos de etiquetas no están dentro de su área respectiva. En el literal d) se mantiene dominante el área de clasificación 1.

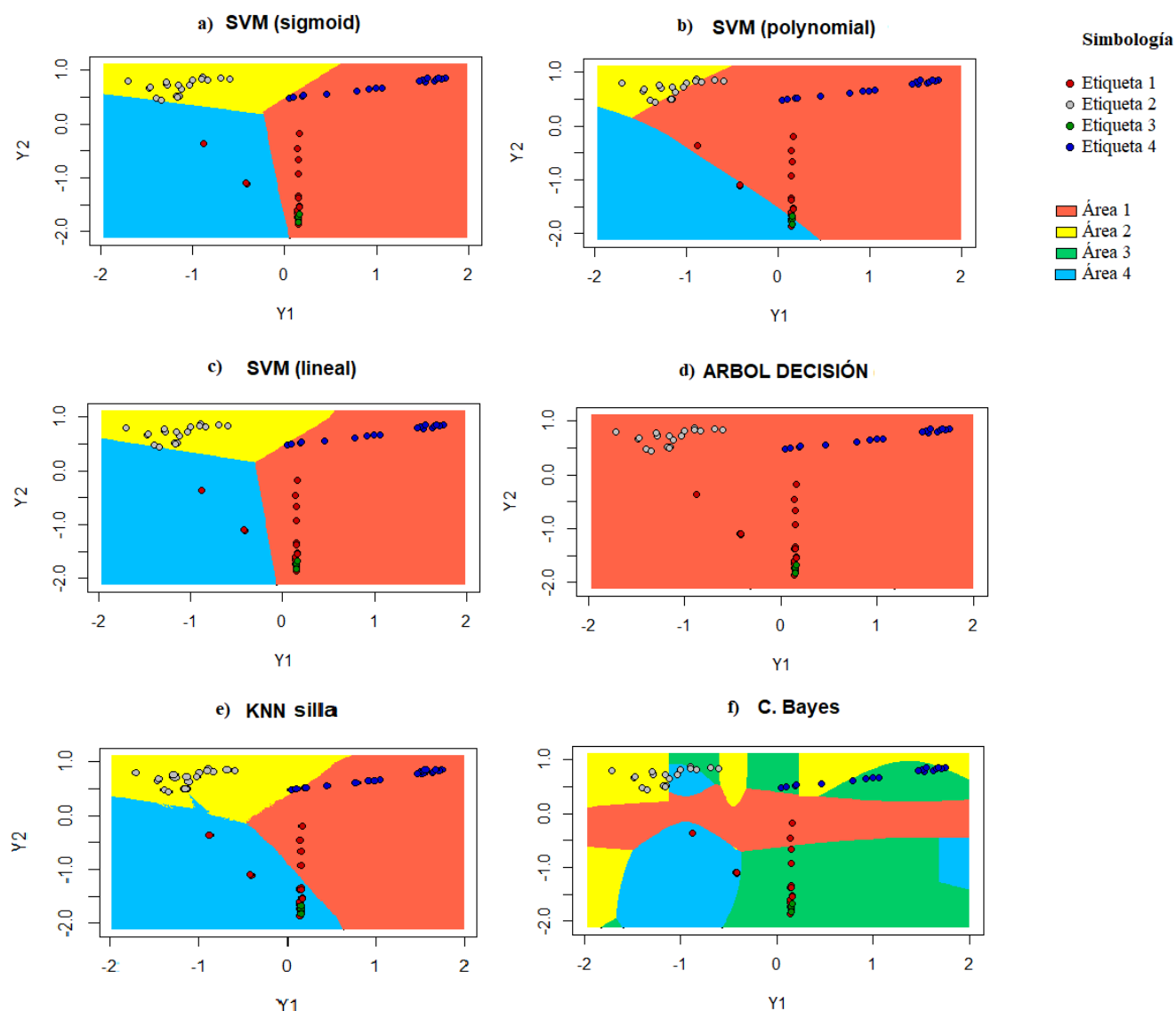


Figura 59 Clasificación con LLE en BDD2

Fuente: Autoría propia

En la Figura 60 se refleja la clasificación realizada en BDD2 con SNE, donde se visualiza que en la mayoría de literales dentro del gráfico se encuentran representadas todas las áreas de clasificación, a excepción del literal d) al igual que en figuras anteriores. También se observa como en los literales e) y f) existe una mejor la clasificación de cada etiqueta con respecto a su área.

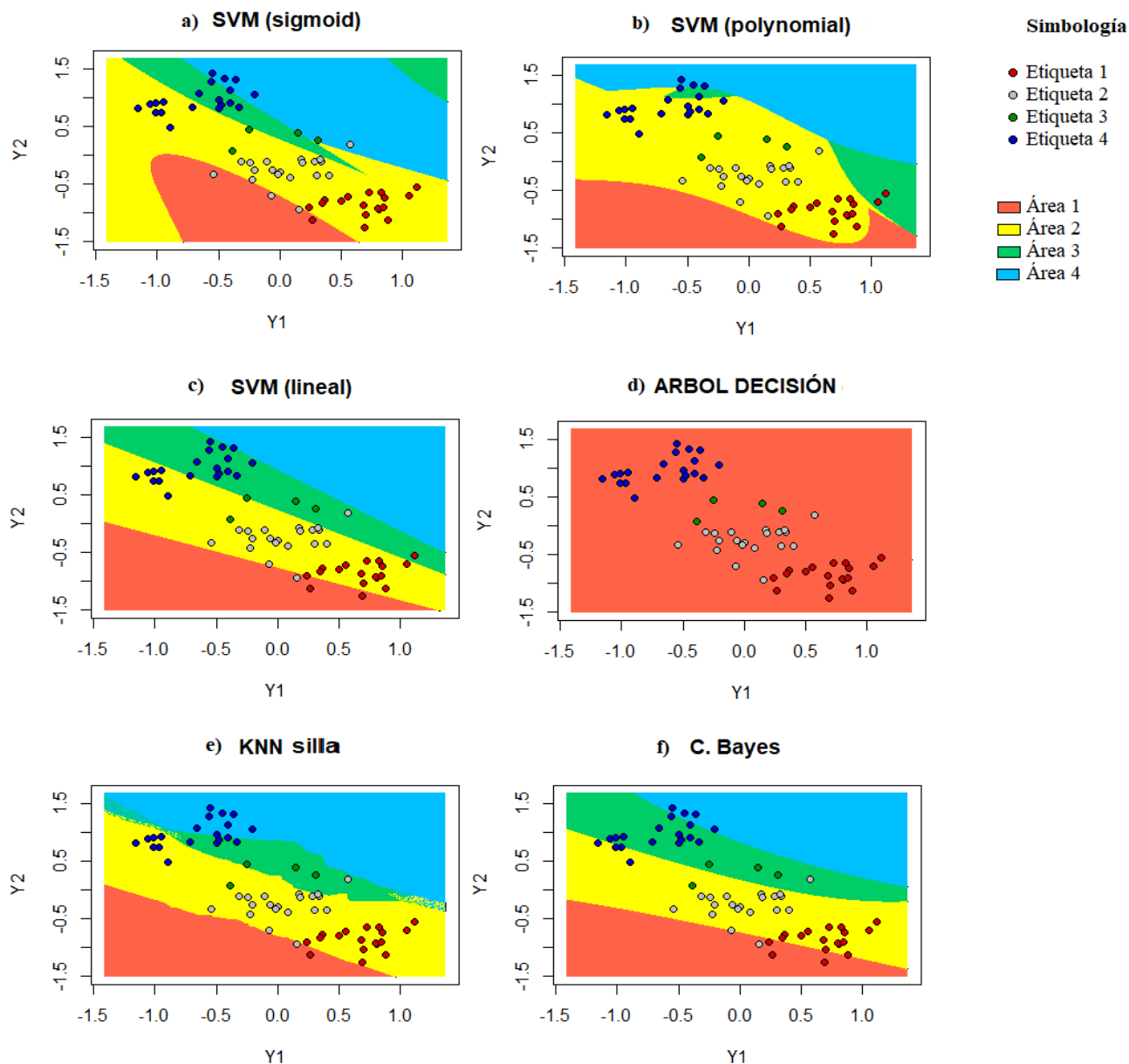


Figura 60 Clasificación con SNE en BDD2  
Fuente: Autoría propia

En la Figura 61 se muestra la clasificación realizada en BDD2 con  $t$ -SNE, donde se observa las mismas características de clasificación obtenidas con LE y LLE, al presentar solo 3 de las 4 áreas de clasificación y manteniendo un área dominante en el literal d).

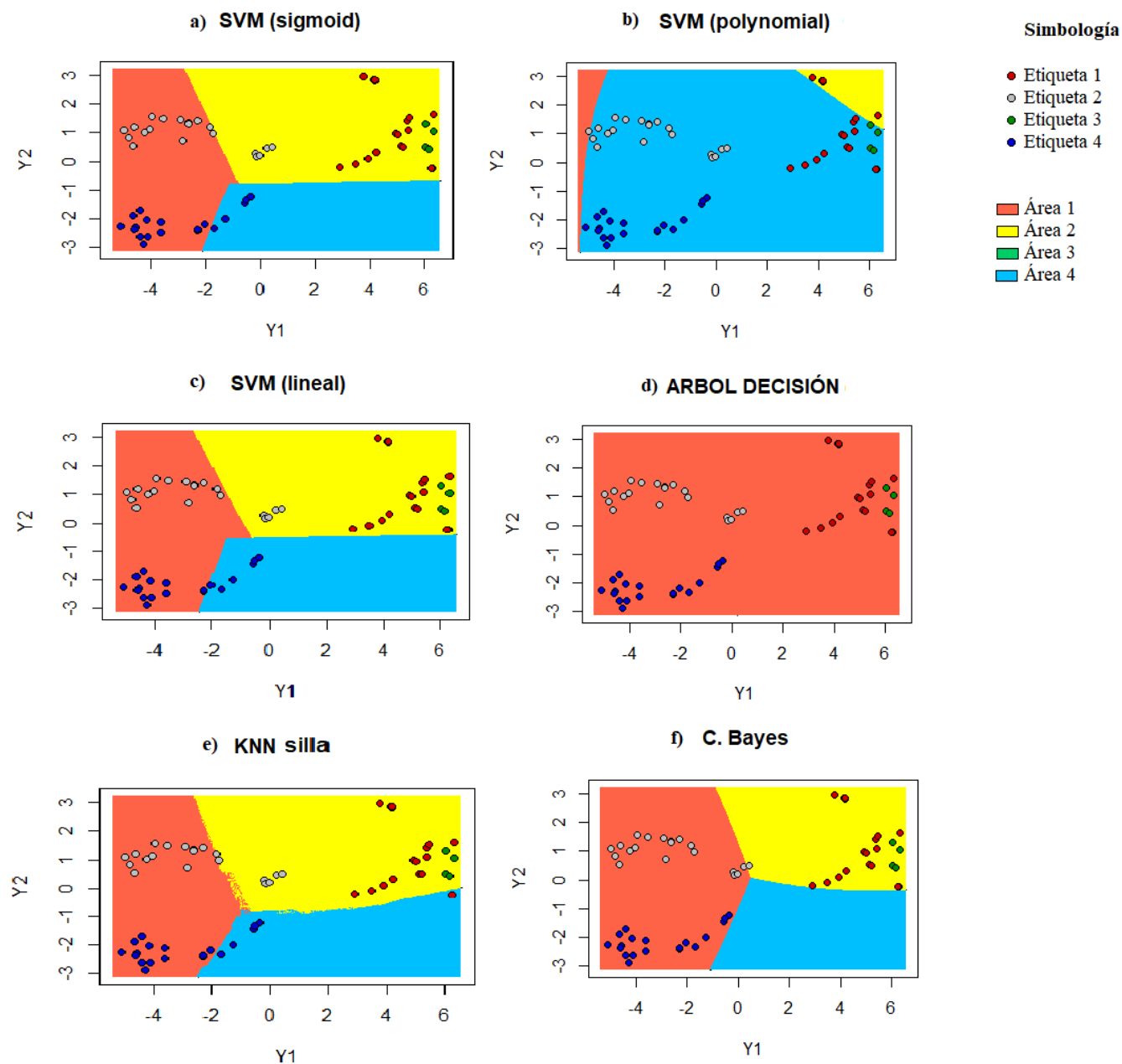


Figura 61 Clasificación con  $t$ -SNE en BDD1  
Fuente: Autoría propia

En la Figura 62 se puede ver la clasificación realizada en BDD3 con PCA, donde se aprecia que existe una buena clasificación de los grupos de etiquetas, y en el que se destaca el literal e) correspondiente al clasificador kNN debido a que tiene la mayor cantidad de puntos de las etiquetas se encuentran dentro de su color de área correspondiente.

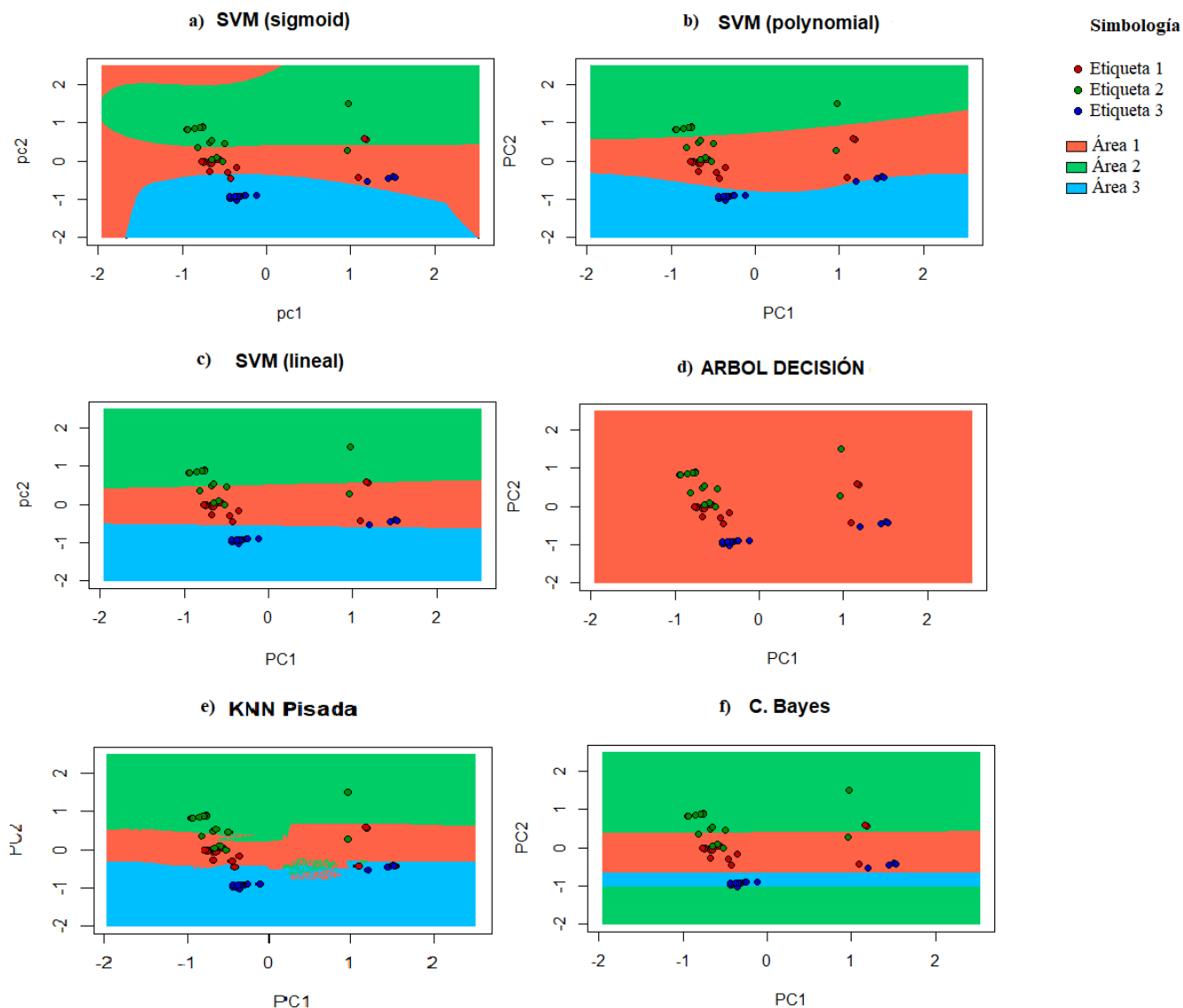


Figura 62 Clasificación con PCA en BDD3  
Fuente: Autoría propia



En la Figura 63 se puede ver la clasificación realizada en BDD3 con LE, donde se comprueba visualmente una mejor clasificación realizada tanto por kNN como por el C. Bayesiano representados en los literales e) y f) respectivamente. Y manteniendo la dominación de una sola área en el literal d) como en las figuras anteriores.

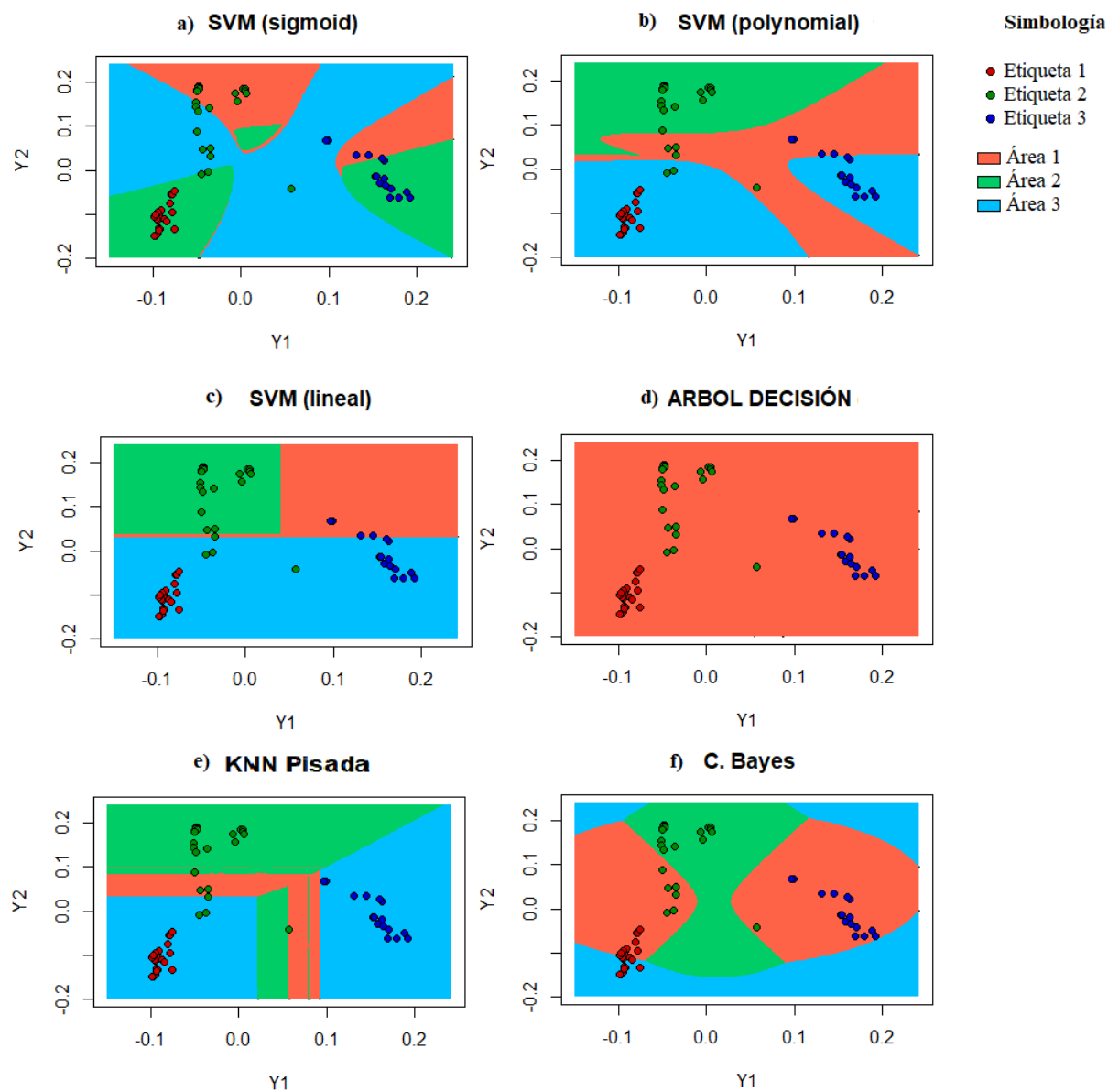


Figura 63 Clasificación con LE en BDD3  
Fuente: Autoría propia

En la Figura 64 se presenta la clasificación realizada en BDD3 con LLE, donde se aprecia que no existe una buena clasificación realizada por los distintos algoritmos presentados. En los literales b) y c) es donde se ve la mejor clasificación realizada gráficamente debido a la mayor correspondencia del grupo de etiquetas 1 y 2 con sus áreas correspondientes.

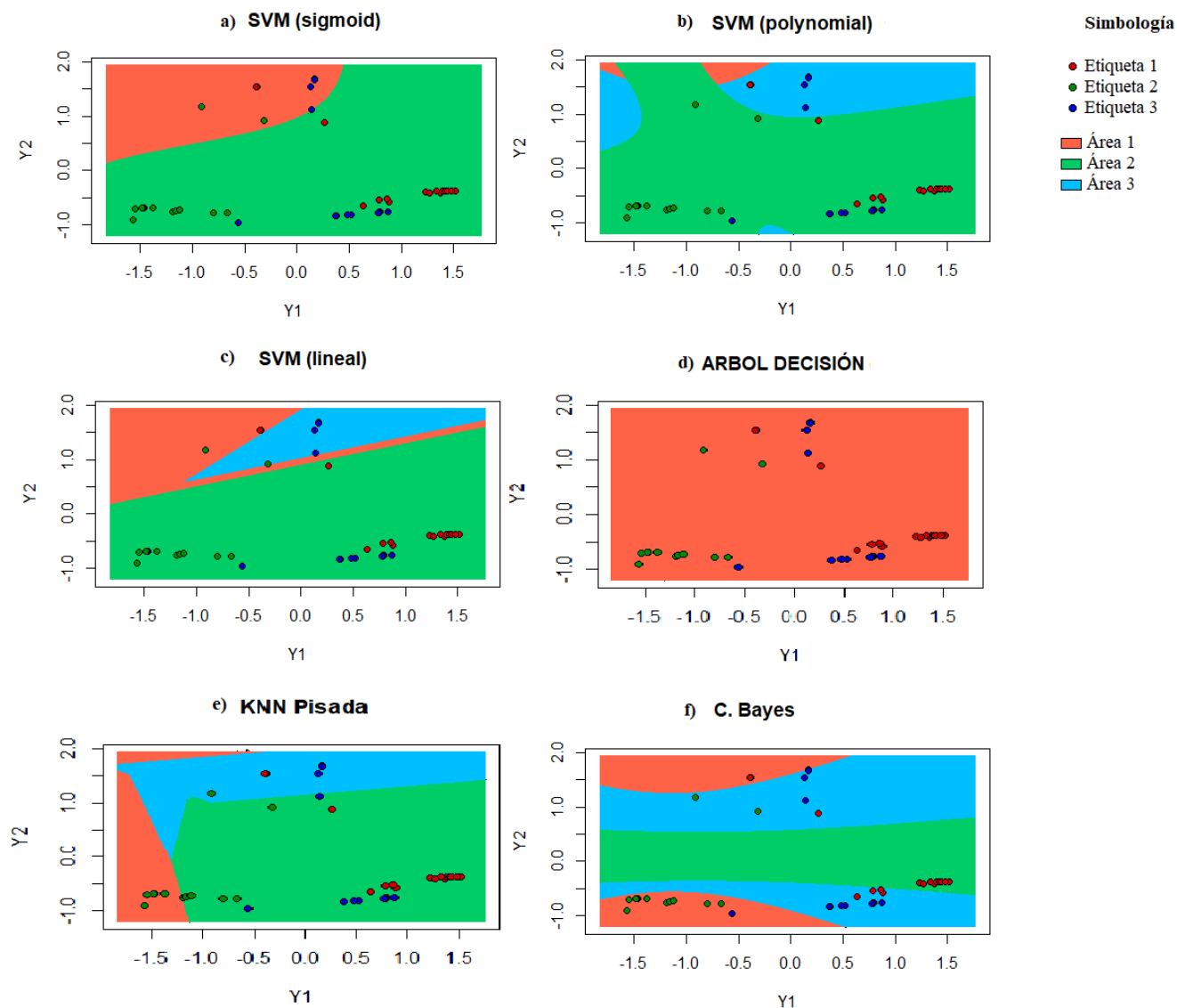


Figura 64 Clasificación con LLE en BDD3

Fuente: Autoría propia

En la Figura 65 se observa la clasificación realizada en BDD3 con SNE, revelando de esta manera un nivel bajo de resultados, ya que en la mayoría de los gráficos correspondientes a cada literal solo coinciden pocos grupos de las etiquetas con sus áreas respectivas.

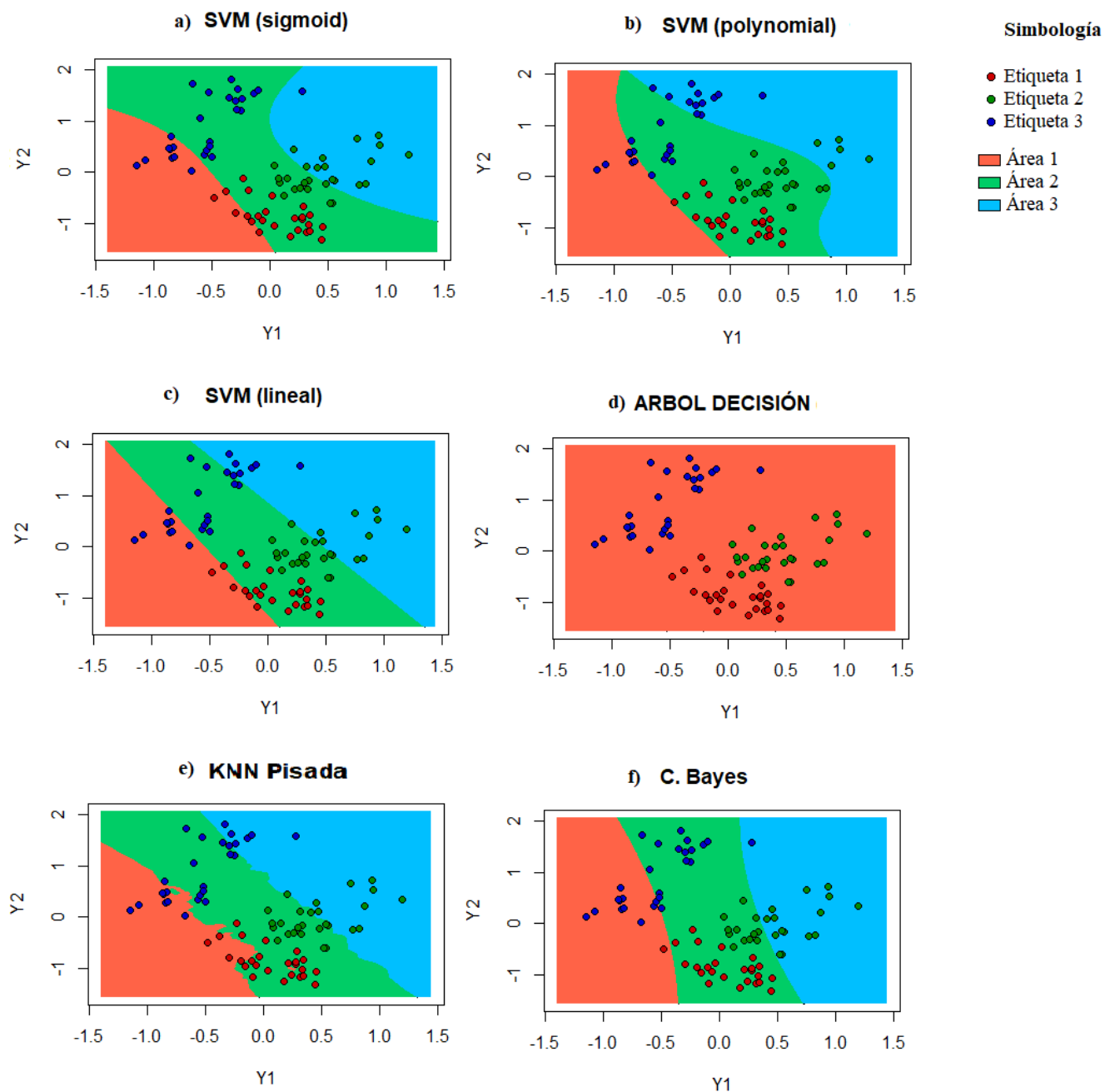


Figura 65 Clasificación con SNE en BDD3  
Fuente: Autoría propia

En la Figura 66 se encuentra la clasificación realizada en BDD3 con  $t$ -SNE, donde visualmente se destaca la clasificación con el algoritmo SVM con kernel sigmoid que se ubica en el literal a), y se muestra que existe la mayor coincidencia del grupo de etiquetas con las áreas de clasificación correspondiente, siendo aun así un nivel bajo de los resultados de clasificación.

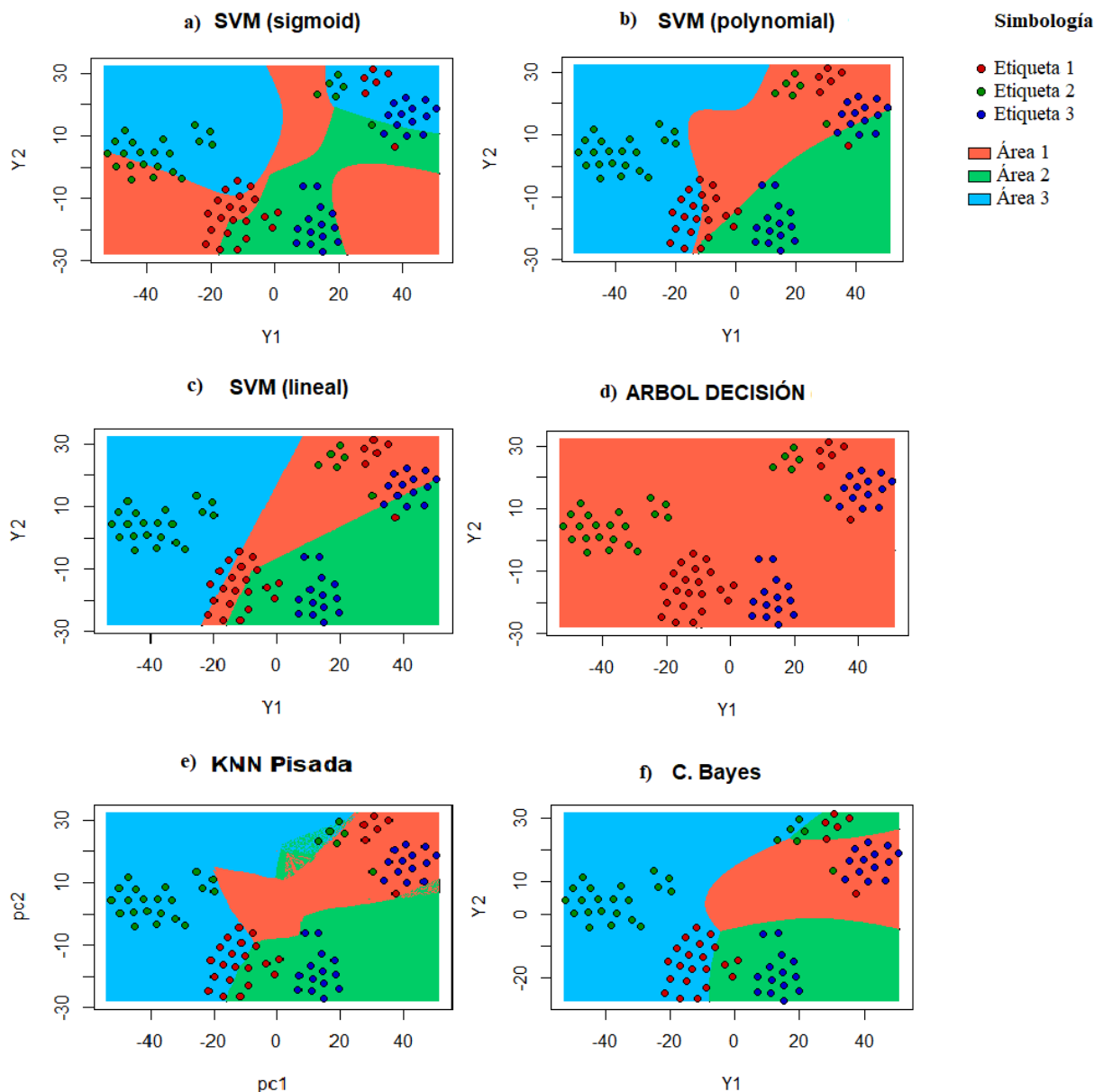


Figura 66 Clasificación con  $t$ -SNE en BDD3  
Fuente: Autoría propia

En la Figura 67 se puede apreciar la clasificación realizada en BDD4 con PCA, donde se muestran buenos resultados por la mayoría de los algoritmos, en especial por el clasificador kNN en el literal e) en el que no existe una división de áreas rígida, por lo que se observa claramente la mejor precisión de clasificación realizada a este set de datos.

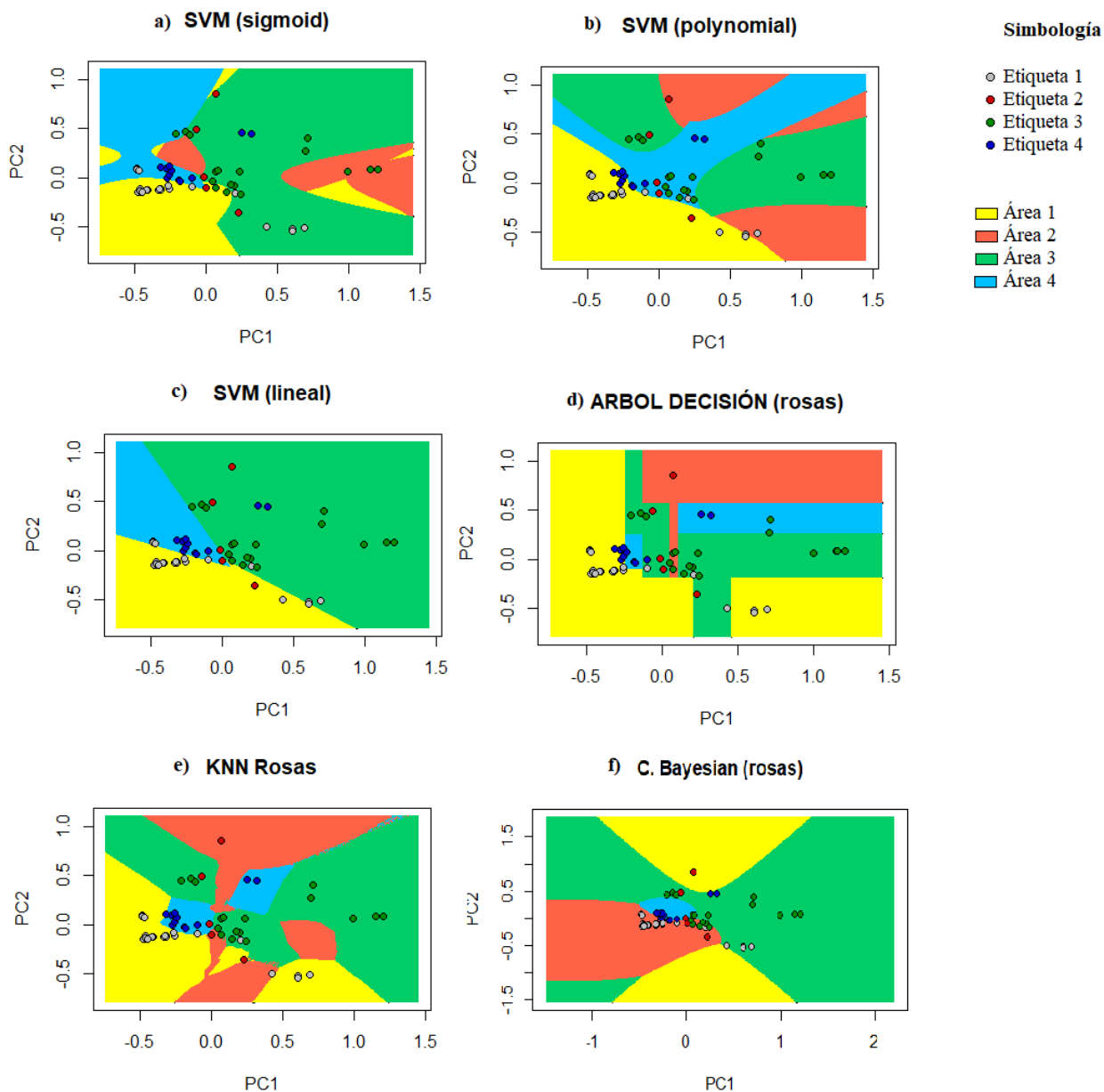


Figura 67 Clasificación con PCA en BDD4  
Fuente: Autoría propia

En la Figura 68 se puede ver la clasificación realizada en BDD4 con LE, en la que claramente destaca que solo en los literales a) y b) correspondientes al clasificador SVM es donde se muestran todas las áreas de clasificación, siendo así el que mejores resultados presenta en esta Figura. También destaca que en el literal d) es el segundo gráfico que no es dominado por una sola área de clasificación.

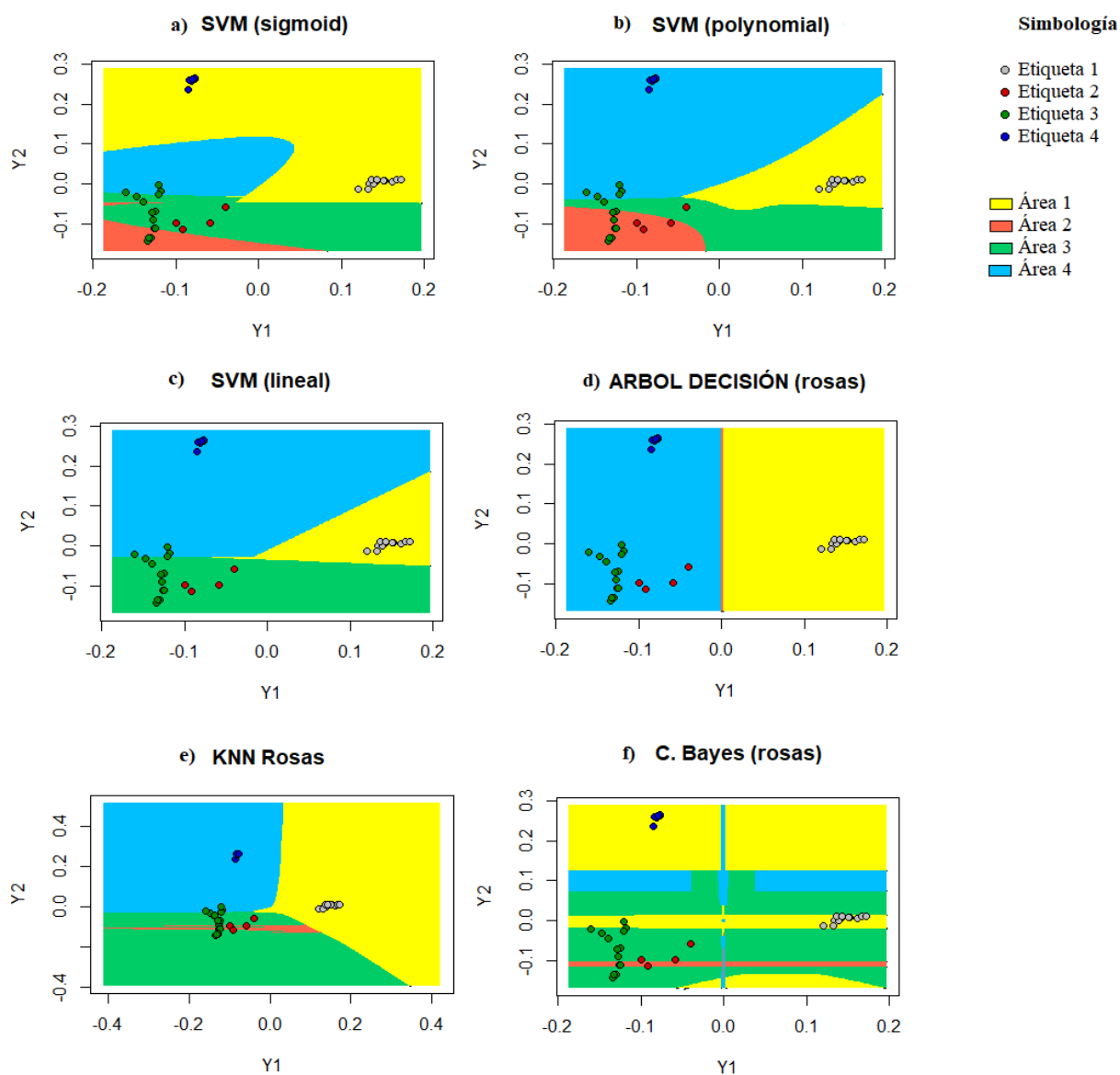


Figura 68 Clasificación con LE en BDD4  
Fuente: Autoría propia

En la Figura 69 se muestra la clasificación realizada en BDD4 con LLE, en la que no existe un buen nivel de clasificación de los grupos de etiquetas con cada área correspondiente. En los literales a) y e) se observa los mejores resultados de clasificación dentro de esta Figura, a pesar de seguir siendo de un bajo porcentaje de precisión tan solo con un análisis visual.

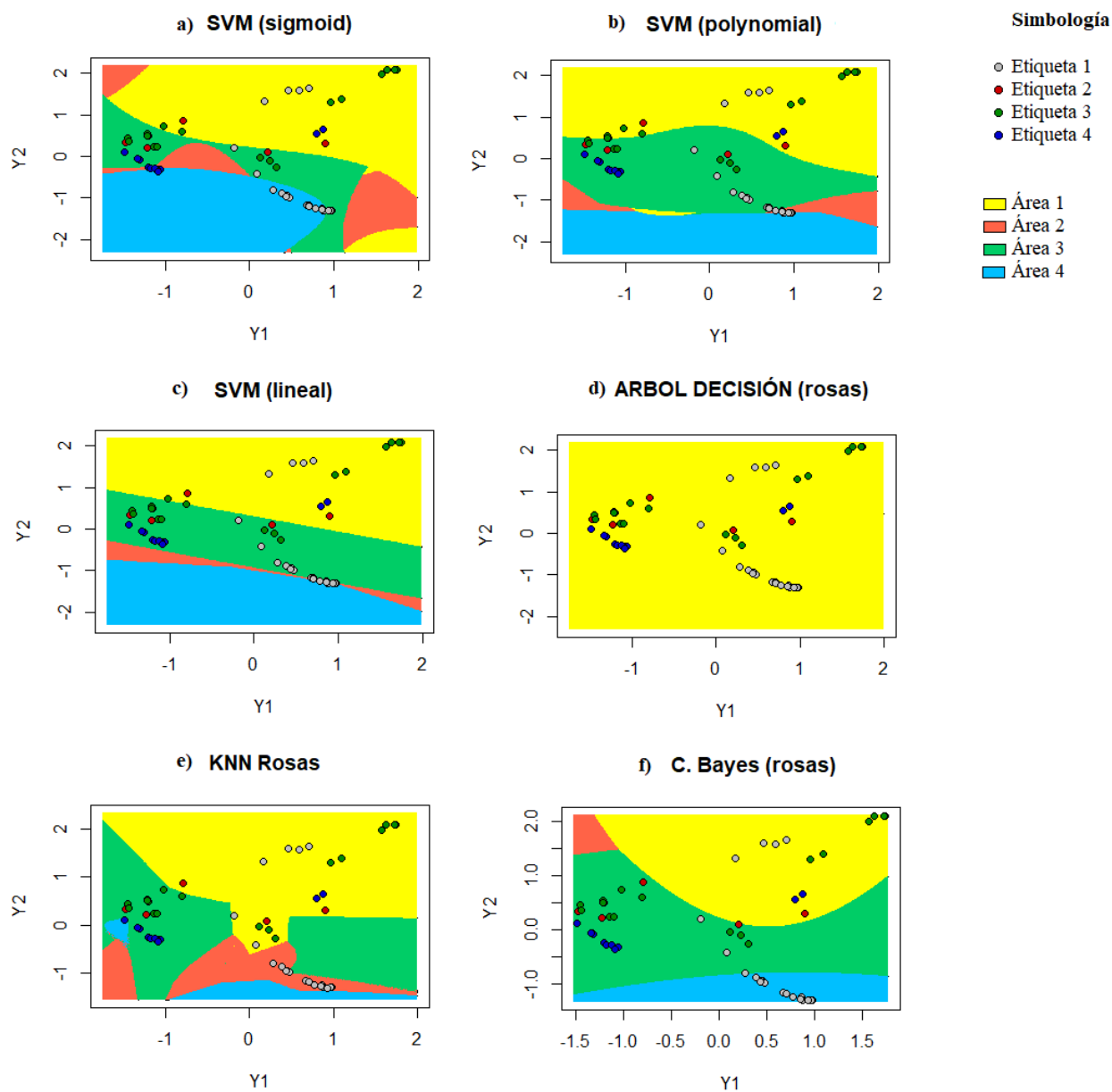


Figura 69 Clasificación con LLE en BDD4  
Fuente: Autoría propia

En la Figura 70 se refleja la clasificación realizada en BDD4 con SNE, en la que los resultados mostrados en los literales b), c) y e) son los que tienen mejor nivel de clasificación pero que quizás no sobrepasen un 50% de precisión. También se visualiza fácilmente que en el literal d) predomina solo un área de clasificación.

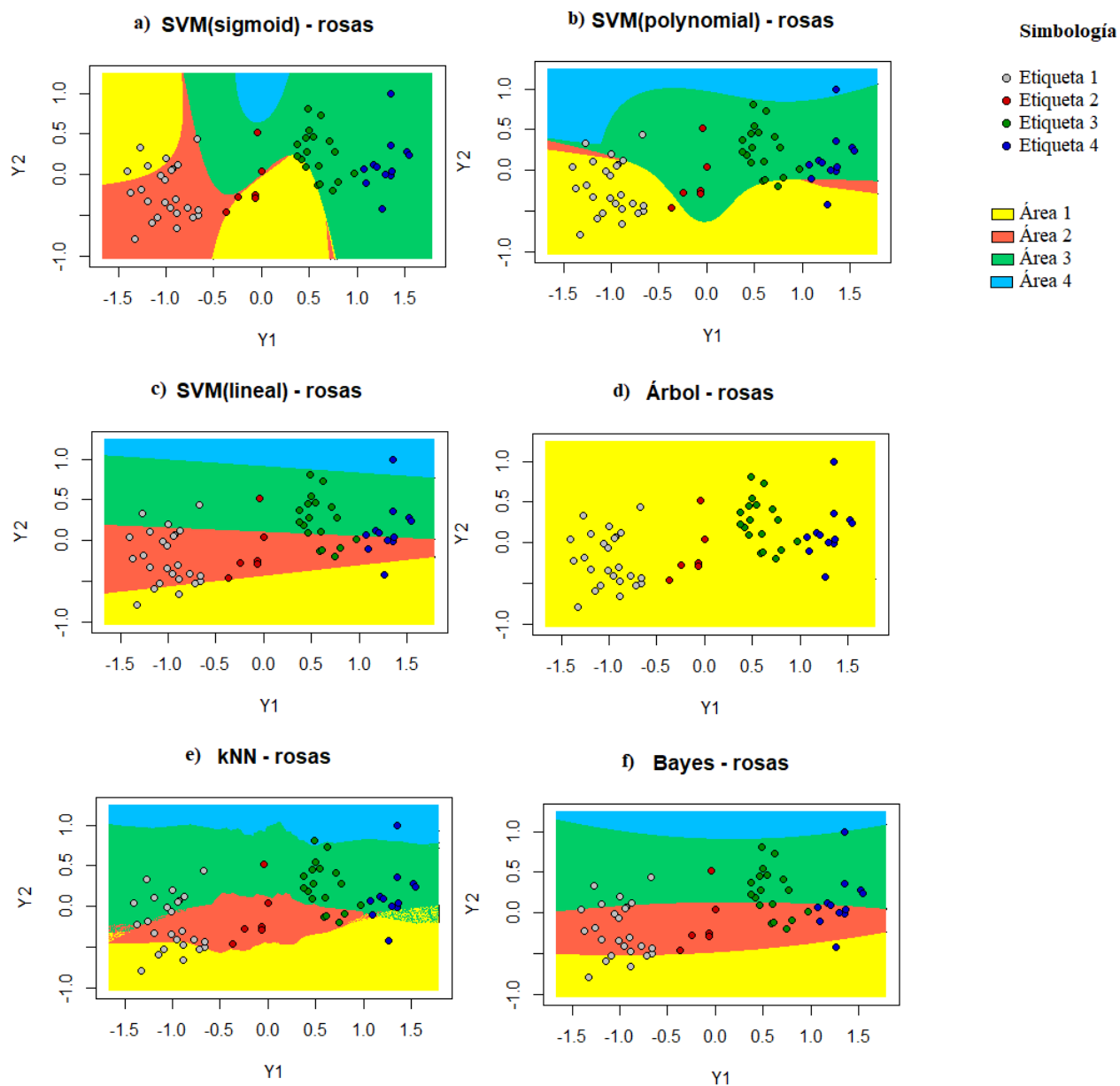


Figura 70 Clasificación con SNE en BDD4

Fuente: Autoría propia



En la Figura 71 se evidencia la clasificación realizada en BDD4 con  $t$ -SNE, donde solo en los literales a) y e) se visualiza las 4 áreas de clasificación, y con un bajo porcentaje de precisión de menos del 50% mediante un análisis visual. Además, se puede observar que en los literales b) y d) existe la dominación de una sola área de clasificación, siendo distintas en cada uno de ellos.

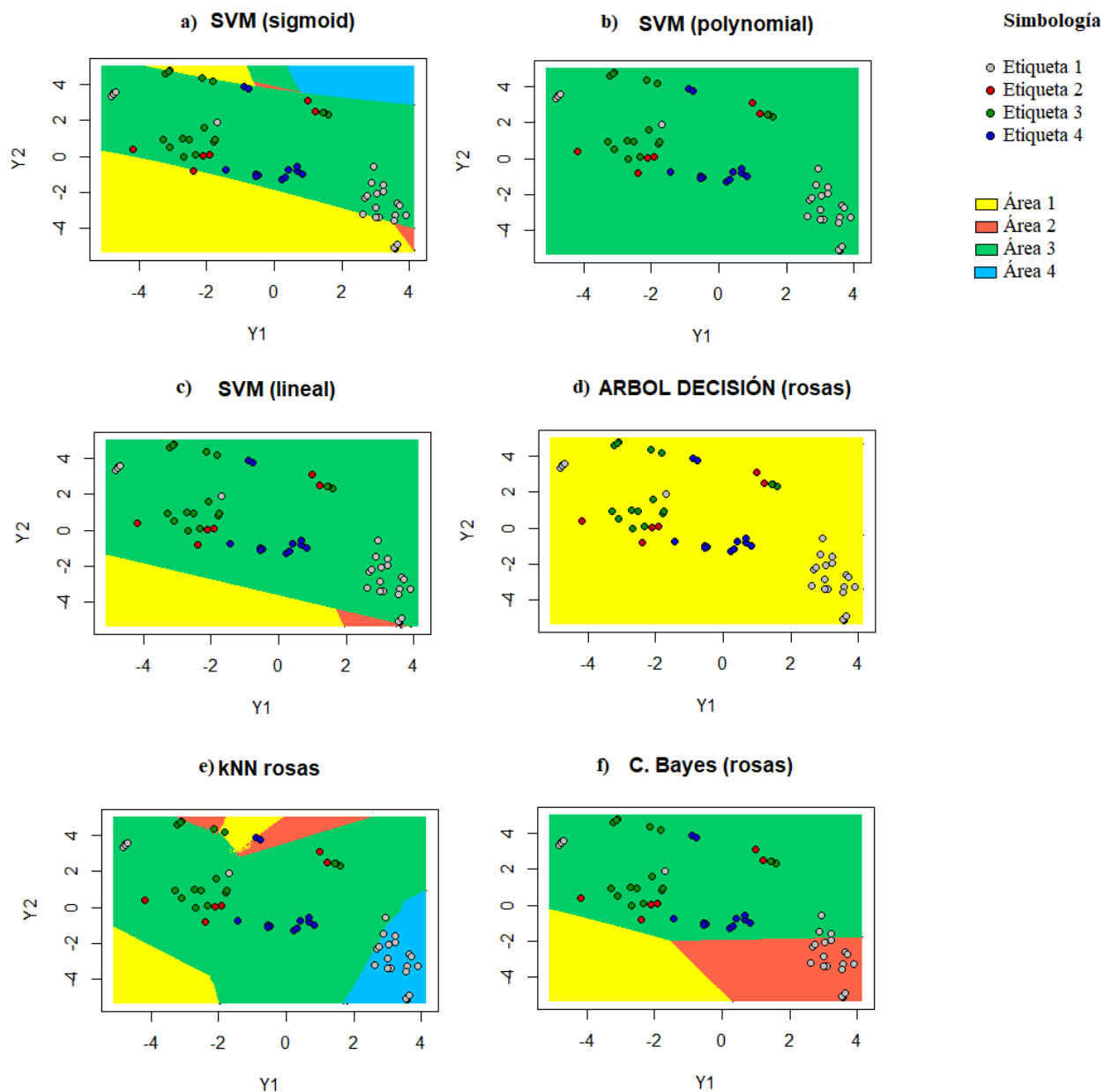


Figura 71 Clasificación con  $t$ -SNE en BDD4  
Fuente: Autoría propia

En la Figura 72 se encuentra la clasificación realizada en BDD5 con PCA, donde se mantiene la dominación del área 1 de clasificación en el literal d). También se puede observar fácilmente que al existir en este set de datos solo 2 tipos de etiquetas los resultados de clasificación son mucho mejores que en los anteriores casos.

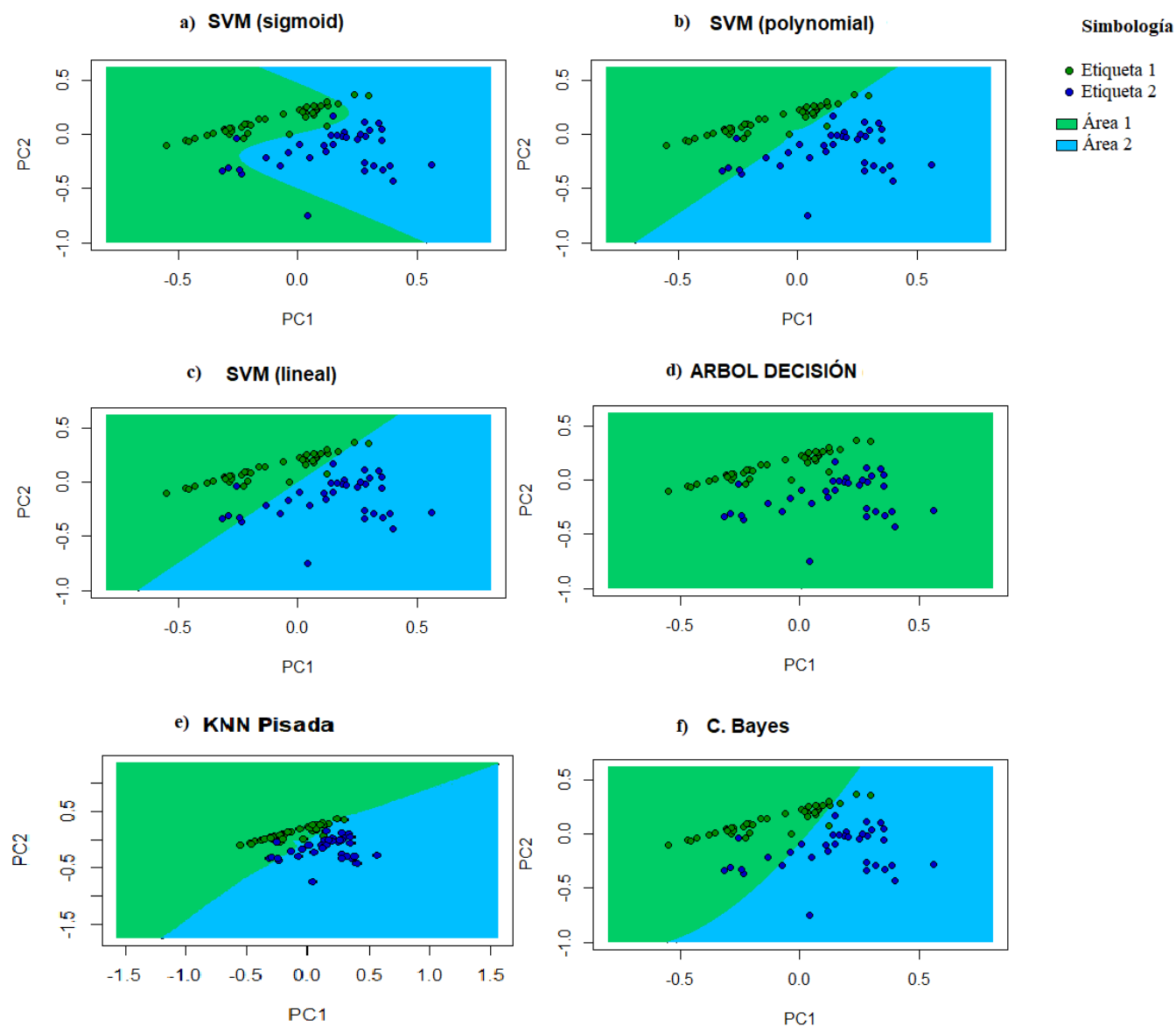


Figura 72 Clasificación con PCA en BDD5  
Fuente: Autoría propia

En la Figura 73 se muestra la clasificación realizada en BDD5 con LE, en los que casi todos obtienen 0% de precisión en su clasificación a excepción de los literales a) donde muy pocos puntos de la etiqueta 1 se encuentran dentro de su área respectiva y d) en la que sigue dominando solo un área de clasificación.

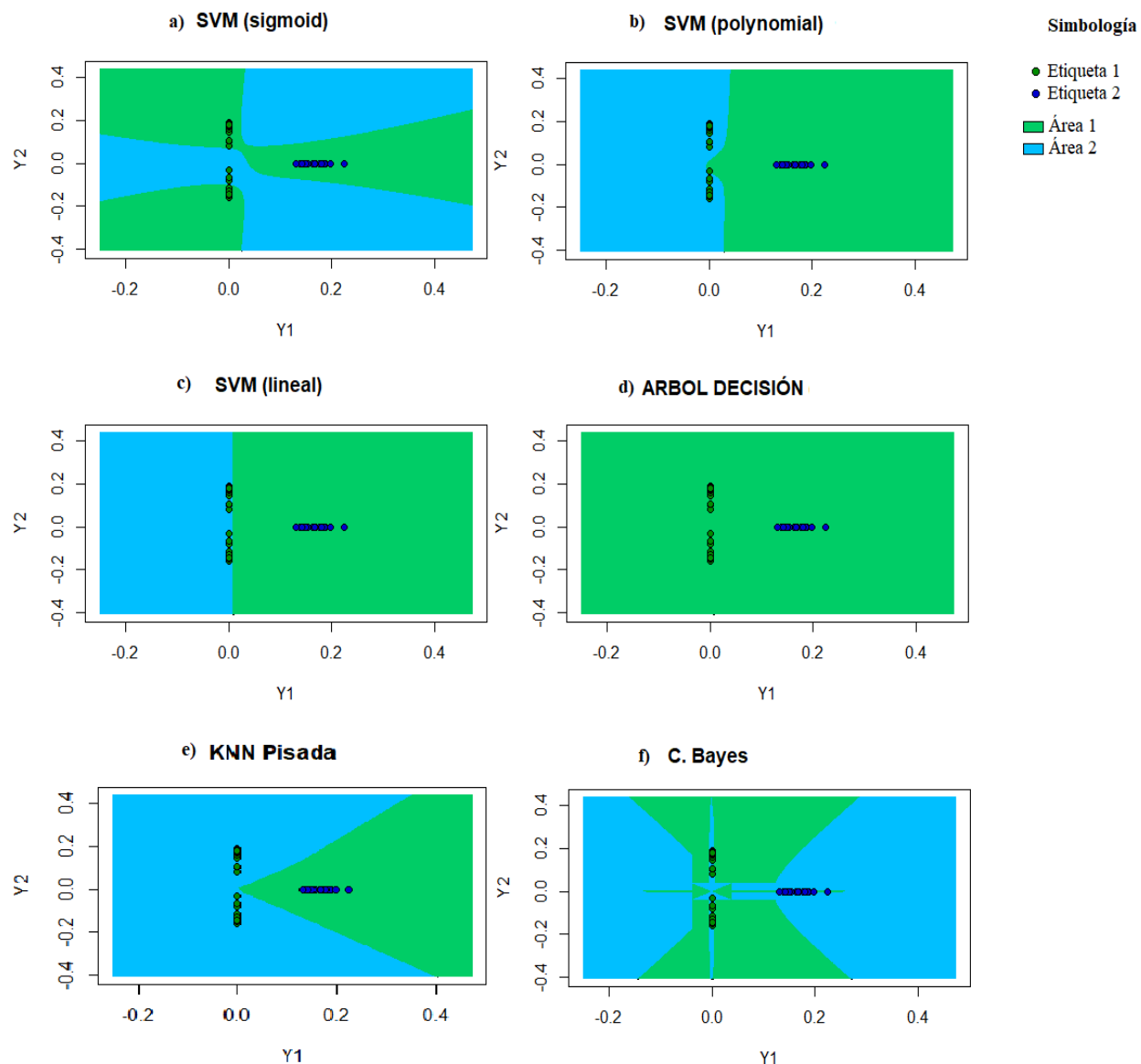


Figura 73 Clasificación con LE en BDD5  
Fuente: Autoría propia

En la Figura 74 se visualiza la clasificación realizada en BDD5 con LLE, en la que las áreas de clasificación solo coinciden en solo un porcentaje del 50% o menos con los puntos de etiquetas respectivas. Y se mantiene la dominación del área 1 en la representación gráfica del clasificador árbol de decisión.

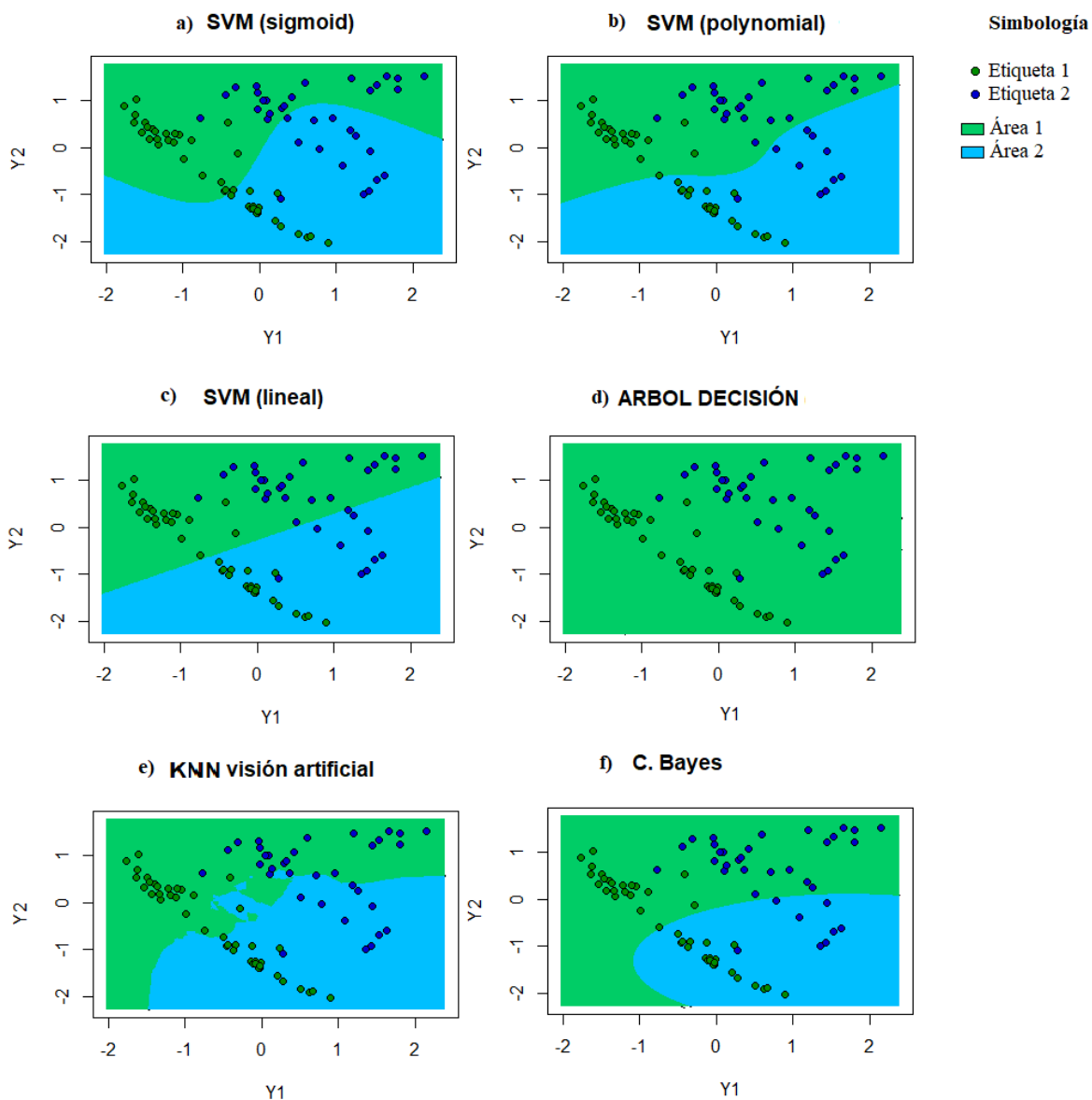


Figura 74 Clasificación con LLE en BDD5  
Fuente: Autoría propia

En la Figura 75 se puede ver la clasificación realizada en BDD5 con SNE, donde se comprueba gráficamente una difícil clasificación de los grupos de etiquetas con sus respectivas áreas ya que se encuentran mezclados entre sí.

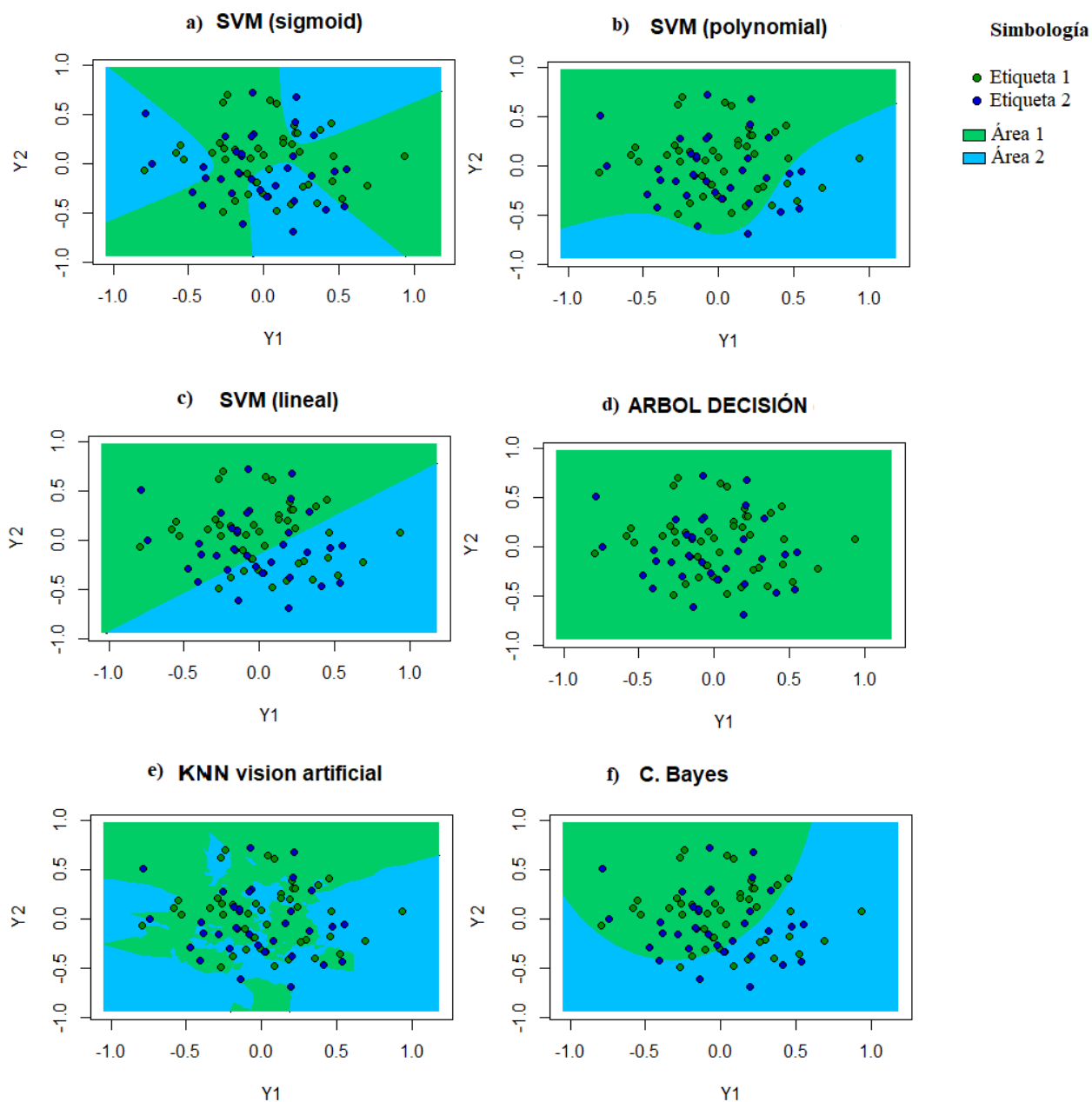


Figura 75 Clasificación con SNE en BDD5  
Fuente: Autoría propia

En la Figura 76 se muestra la clasificación realizada en BDD5 con  $t$ -SNE, y se verifica que no existe una buena clasificación al ser aplicada a este método de RD, los grupos de etiquetas se muestran ubicados en el área opuesta de clasificación.

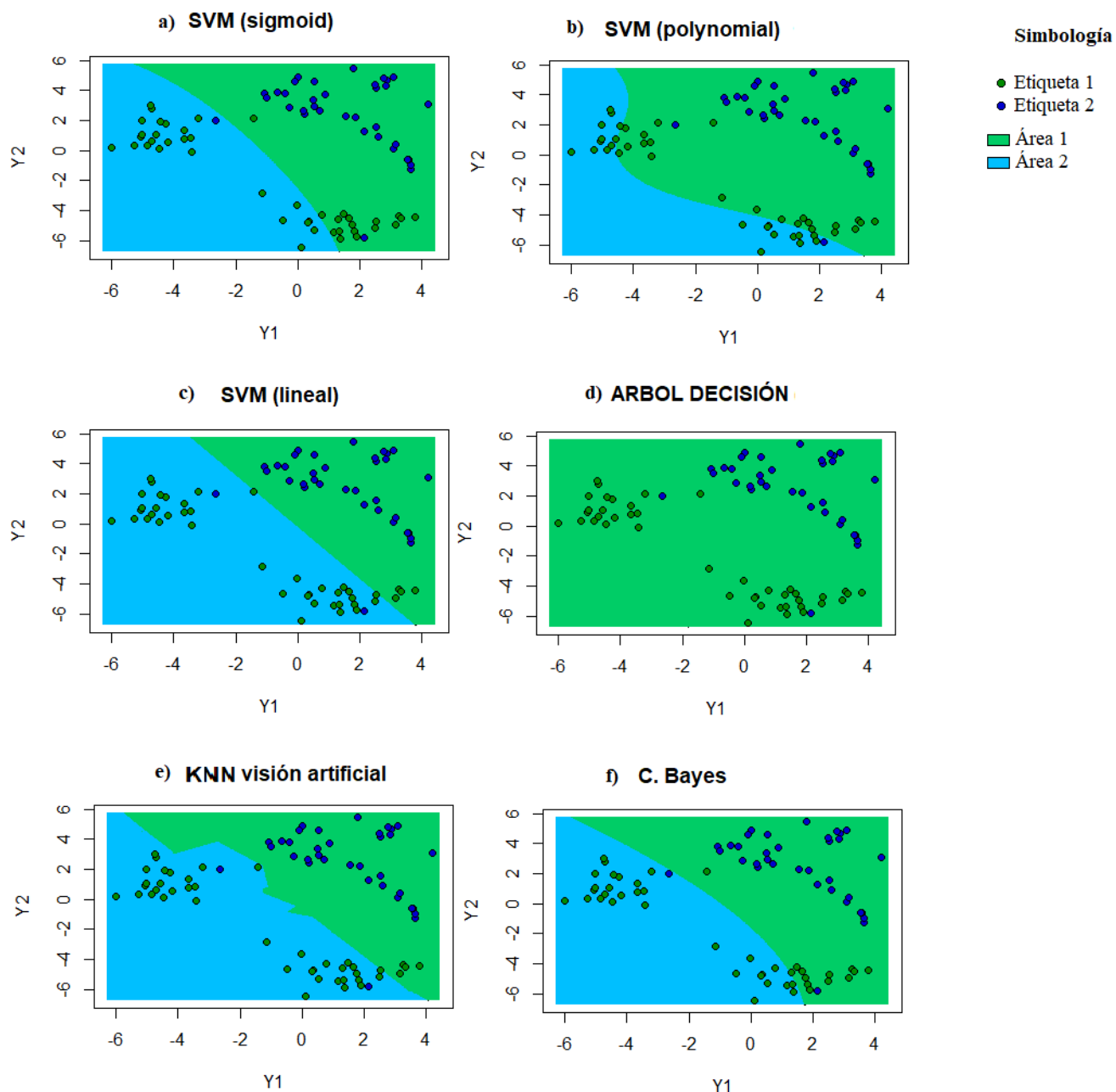


Figura 76 Clasificación con  $t$ -SNE en BDD5  
Fuente: Autoría propia

### 4.3 Resultados

En la Tabla 10 se evidencia la precisión en valor de porcentajes de los resultados de la clasificación realizada en el método de reducción de dimensionalidad PCA. Dando como clasificador con mayor porcentaje de precisión a KNN con un 84,15%.

Tabla 10

Precisión obtenida de clasificadores en PCA.

<b>BDD</b>	<b>PCA</b>			
	<b>SVM</b>	<b>ARBOL DECISIÓN</b>	<b>C. BAYES</b>	<b>KNN</b>
<b>ROSAS</b>	68,85%	70,49%	60,66%	77,05%
<b>POSICION</b>	98,48%	94,70%	99,24%	98,48%
<b>SILLA</b>	69,35%	62,90%	69,35%	67,74%
<b>PISADA</b>	66,67%	67,90%	64,20%	85,19%
<b>VISION</b>	92,31%	84,62%	85,90%	92,31%
	<b>79,13%</b>	<b>76,12%</b>	<b>75,87%</b>	<b>84,15%</b>

*Fuente: Autoría propia*

En la Tabla 11 se evidencia la precisión en valor de porcentajes de los resultados de la clasificación realizada en el método de reducción de dimensionalidad LE. Dando SVM como clasificador con mayor porcentaje de precisión con un 54.15%.

Tabla 11

Precisión obtenida de clasificadores en LE.

<b>LE</b>				
<b>BDD</b>	<b>SVM</b>	<b>ARBOL DECISIÓN</b>	<b>C. BAYES</b>	<b>KNN</b>
<b>ROSAS</b>	81,97%	60,66%	67,21%	83,61%
<b>POSICION</b>	17,42%	17,42%	15,91%	17,42%
<b>SILLA</b>	64,52%	32,26%	32,26%	32,26%
<b>PISADA</b>	55,56%	50,62%	44,44%	60,49%
<b>VISION</b>	51,28%	0,00%	0,00%	0,00%
	<b>54,15%</b>	<b>32,19%</b>	<b>31,96%</b>	<b>38,76%</b>

*Fuente: Autoría propia*

En la Tabla 12 se evidencia la precisión en valor de porcentajes de los resultados de la clasificación realizada en el método de reducción de dimensionalidad LLE. Dando SVM como clasificador con mayor porcentaje de precisión con un 40.07%.

Tabla 12

Precisión obtenida de clasificadores en LLE.

<b>LLE</b>				
<b>BDD</b>	<b>SVM</b>	<b>ARBOL DECISIÓN</b>	<b>C. BAYES</b>	<b>KNN</b>
<b>ROSAS</b>	26,23%	13,11%	26,23%	26,23%
<b>POSICION</b>	34,09%	25,00%	25,00%	25,00%
<b>SILLA</b>	54,84%	22,58%	30,65%	38,71%
<b>PISADA</b>	48,15%	11,11%	37,04%	27,16%
<b>VISION</b>	51,28%	35,90%	38,46%	54,31%
	<b>40,07%</b>	<b>21,54%</b>	<b>31,48%</b>	<b>34,28%</b>

*Fuente: Autoría propia*



En la Tabla 13 se evidencia la precisión en valor de porcentajes de los resultados de la clasificación realizada en el método de reducción de dimensionalidad SNE. Obteniendo SVM como clasificador con mayor porcentaje de precisión con un 64,06%.

Tabla 13

Precisión obtenida de clasificadores en SNE.

BDD	SNE			
	SVM	ARBOL DECISIÓN	C. BAYES	KNN
<b>ROSAS</b>	63,93%	36,07%	34,43%	40,98%
<b>POSICION</b>	82,58%	81,06%	83,33%	64,39%
<b>SILLA</b>	30,65%	61,29%	85,48%	22,58%
<b>PISADA</b>	40,74%	24,69%	24,69%	27,16%
<b>VISION</b>	55,13%	56,41%	53,85%	48,72%
	<b>64,06%</b>	<b>51,90%</b>	<b>56,36%</b>	<b>40,77%</b>

*Fuente: Autoría propia*

En la Tabla 14 se evidencia la precisión en valor de porcentajes de los resultados de la clasificación realizada en el método de reducción de dimensionalidad  $t$ SNE. Dando SVM como clasificador con mayor porcentaje de precisión con un 39.42%.

Tabla 14

Precisión obtenida de clasificadores en  $t$ SNE.

BDD	T-SNE			
	SVM	ARBOL DECISIÓN	C. BAYES	KNN
<b>ROSAS</b>	37,70%	34,43%	29,51%	26,23%
<b>POSICION</b>	22,91%	15,91%	22,51%	16,14%
<b>SILLA</b>	32,26%	0,00%	1,61%	20,97%
<b>PISADA</b>	74,76%	64,08%	9,71%	64,08%
<b>VISION</b>	29,49%	32,05%	17,95%	6,41%
	<b>39,42%</b>	<b>29,29%</b>	<b>16,26%</b>	<b>26,77%</b>

*Fuente: Autoría propia*

Finalmente, después de realizar las pruebas necesarias y al analizar cada una de las tablas de precisión se puede observar que el algoritmo KNN aplicado en el método PCA tiene la mayor precisión que es 84,15%, además, cabe destacar que en este método todos los algoritmos de clasificación superan el 75% de precisión. Mientras que, SNE tiene su mejor porcentaje al trabajar con el algoritmo de clasificación SVM con un valor de 64,06%, y en promedio de este método con todos los clasificadores tiene un 53,27%.

Los métodos de RD como LE, LLE y *t*SNE en promedio al trabajar con los algoritmos clasificadores no superan el 40% de precisión. LE, LLE y *t*SNE se destacan al trabajar con SVM con un porcentaje de 54,15%, 40,07% y 39,42% respectivamente, siendo todos estos valores de precisión muy bajos.

En la representación gráfica de los algoritmos de clasificación se puede comprobar el trabajo realizado por cada uno de ellos aplicados en cada método de RD. Donde es importante mencionar que, por una parte, en la gran mayoría de los gráficos del clasificador Árbol de Decisión solo es posible visualizar un área de clasificación, lo que indica su bajo rendimiento al trabajar con la mayoría de los métodos de RD realizados en este trabajo. Por otra parte, el clasificador SVM tiene la mejor representación gráfica de cada área de clasificación con todas las bases de datos, destacándose al trabajar con PCA. Comprobando así los resultados expuestos en las tablas de precisión.

## CAPITULO V. CONCLUSIONES Y RECOMENDACIONES

### 5.1 CONCLUSIONES

- Mediante el estudio de bases teóricas para la realización de este proyecto se logró analizar la importancia de la reducción de dimensionalidad para obtener conjuntos de datos de menor dimensión con respecto al conjunto de datos original, brindando así una fácil comprensión y representación de estos.
- La limpieza y acondicionamiento de cada set de datos dentro del proceso de RD es primordial, ya que permite un correcto trabajo y mejor rendimiento tanto por parte de los métodos de RD como de los algoritmos de clasificación.
- El algoritmo de clasificación realiza dos fases importantes, en la primera se realiza la clasificación y en la segunda la predicción, las dos hacen uso de dos sets de datos denominados de entrenamiento y de prueba.
- Dividir el set de datos original para obtener un gran set de datos de entrenamiento permite mejorar la precisión de la clasificación y predicción, es decir, el sistema presentará errores en su rendimiento si es entrenada con una pequeña cantidad de datos.
- El clasificador SVM en promedio tiene mejores resultados de precisión en comparación a los demás clasificadores analizados al trabajar con métodos como PCA, LE, LLE, SNE y tSNE.
- PCA al utilizar la varianza para representar la cantidad de información que cada componente representa, es importante normalizar los datos con anticipación para que iguale la participación de cada componente.

- El método PCA es el método con los más altos porcentajes obtenidos al aplicar los algoritmos clasificadores, lo que comprueba que es el mejor método de RD aplicado a los sets de datos trabajados.
- El clasificador kNN tiene la más alta precisión al trabajar con el método PCA, siendo una muestra del más óptimo clasificador aplicado al mejor método de RD.
- El poder visualizar gráficamente los resultados permite tener una comparativa mucho más fácil y comprensible, tanto para un experto como para una persona con conocimientos básicos en el tema.
- La representación gráfica del clasificador Árbol de decisión solo se pudo lograr con buenos resultados al aplicar el método de reducción de dimensionalidad PCA con todas las bases de datos.
- Con el desarrollo de este trabajo de investigación se pudo obtener un amplio conocimiento sobre la importancia de varias técnicas que actualmente se están trabajando para la optimización de procesos en el análisis de datos obtenidos mediante redes de sensores y que a futuro buscan perfeccionarse.

## 5.2 RECOMENDACIONES

- Es importante instalar, investigar y estudiar las herramientas y librerías del Software de análisis y visualización de datos, para que cumplan con su correcto funcionamiento y de esta manera evitar posibles errores durante la ejecución de la programación.
- Se recomienda el análisis e identificación de los tipos de variables de cada BDD original, para poder establecer los parámetros correctos necesarios tanto para el proceso de RD como de clasificación.
- Es necesario usar un gran volumen de datos para el conjunto de entrenamiento, de esta forma el sistema aprende más características del objeto trabajado, mejorando así la precisión de la clasificación.
- Es recomendable el uso de un equipo con al menos 8GB de RAM para el correcto y rápido funcionamiento del Software RStudio ya que realiza varios procesos de cálculo para la representación de gráficos y más aún cuando se trata de escalas muy pequeñas para las áreas de clasificación.

## BIBLIOGRAFÍA

- Aluja, T. (2001). LA MINERÍA DE DATOS, ENTRE LA ESTADÍSTICA Y LA INTELIGENCIA ARTIFICIAL. *Qüestiió: Quaderns d'estadística i Investigació Operativa*, 25, 479–498.
- Anaya, A. (2017). *Metodología de Visualizacion de Datos Utilizando Metodos Espectrales y Basados en Divergencias para la Reducción Interactiva de la Dimensión*. Universidad Tecnológica de Pereira. Retrieved from <http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/9180/T543.5A536.pdf?sequence=1&isAllowed=y>
- Bouza, C., & Santiago, A. (2012). LA MINERÍA DE DATOS: ARBOLES DE DECISIÓN Y SU APLICACIÓN EN ESTUDIOS MÉDICOS. *Modelación Matemática de Fenómenos Del Medio Ambiente y La Salud*, 2.
- Braun, A. C., Weidner, U., & Hinz, S. (2011). SUPPORT VECTOR MACHINES, IMPORT VECTOR MACHINES AND RELEVANCE VECTOR MACHINES FOR HYPERSPECTRAL CLASSIFICATION - A COMPARISON. *2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*.
- Champutiz, W., Rosero, P., Fuentes, E., & Peluffo, D. (2019). Roses greenhouse cultivation database repository (RosesGreenhDB) IEEE DataPort. IEEEDataPort. <https://doi.org/10.21227/89qy-zh32>
- Combarro, E. ., Montañes, E., Díaz, I., Cortina, R., Alonso, P., & Ranilla, J. (2009). Un Framework

para la Gestión Documental en las Administraciones Públicas. Retrieved from [http://www.anteriores.tecnimap.es/recursos/doc/Comunes/1887858652\\_1232009112538.pdf](http://www.anteriores.tecnimap.es/recursos/doc/Comunes/1887858652_1232009112538.pdf)

Dobrin, A., Stamatescu, G., Dragana, C., & Sgarciu, V. (2016). Cloud challenges for networked embedded systems: A review. In *2016 20th International Conference on System Theory, Control and Computing (ICSTCC)* (pp. 866–871). IEEE. <https://doi.org/10.1109/ICSTCC.2016.7790777>

Elmurngi, E., & Gherbi, A. (2017). An Empirical Study on Detecting Fake Reviews Using Machine Learning Techniques. *The Seventh International Conference on Innovative Computing Technolog.*

Fernández, S. de la F. (2011). *Análisis Componentes Principales Santiago de la Fuente Fernández*. Madrid. Retrieved from [http://www.estadistica.net/Master-Econometria/Componentes\\_Principales.pdf](http://www.estadistica.net/Master-Econometria/Componentes_Principales.pdf)

Fuentes, E., Rosero, P., & Peluffo, D. (2019). Footstep analysis using Pressure Sensors | IEEE DataPort. IEEEDataPort. <https://doi.org/10.21227/kqm6-1877>

Galarza Hernandez, J. (2017). Reducción de dimensionalidad en Machine Learning. Valencia.

Gallego, J., & Rengifo, D. (2016). *Comparación de técnicas de reducción de dimensionalidad para la clasificación de actividades físicas humanas utilizando métodos estadísticos*. Universidad Tecnológica de Pereira. Retrieved from <http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/7288/6213678G166.pdf?sequence=1>

- González García, N., & Taborda Londoño, A. (2015). *ANÁLISIS DE COMPONENTES PRINCIPALES SPARSE Formulación, algoritmos e implicaciones en el análisis de datos*. Universidad de Salamanca. Retrieved from [https://gredos.usal.es/jspui/bitstream/10366/126046/4/TFM\\_MAADM\\_González\\_Taborda.pdf](https://gredos.usal.es/jspui/bitstream/10366/126046/4/TFM_MAADM_González_Taborda.pdf)
- Griparis, A., Faur, D., & Datcu, M. (2015). Feature space dimensionality reduction for the optimization of visualization methods. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1120–1123). IEEE. <https://doi.org/10.1109/IGARSS.2015.7325967>
- Hinton, G., & Roweis, S. (2002). *Stochastic Neighbor Embedding*. Toronto. Retrieved from <https://www.cs.toronto.edu/~fritz/absps/sne.pdf>
- Iozu, J. de L. (2017). *Análisis y optimización de algoritmos de clasificación supervisada sobre operaciones impagadas en tarjetas de crédito*. Universidad Internacional de la Rioja. Retrieved from [https://reunir.unir.net/bitstream/handle/123456789/6293/JUAN DE LLANO%2CIOZU DE.pdf?sequence=1&isAllowed=y](https://reunir.unir.net/bitstream/handle/123456789/6293/JUAN_DE_LLANO%2CIOZU_DE.pdf?sequence=1&isAllowed=y)
- Jiang, Q., Lu, Y., & Hong, Z. (2011). A neighborhood parameter optimization method of LLE based on topology preservation. In *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology* (pp. 4231–4234). IEEE. <https://doi.org/10.1109/EMEIT.2011.6023107>
- Kijoeng, N., Hongmo, J., & Seungjin, C. (2004). Fast stochastic neighbor embedding: a trust-



region algorithm. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)* (pp. 123–128). IEEE. <https://doi.org/10.1109/IJCNN.2004.1379883>

Kok-Chin Khor, Choo-Yee Ting, & Somnuk-Phon Amnuaisuk. (2010). Comparing Single and Multiple Bayesian Classifiers Approaches for Network Intrusion Detection. *Second International Conference on Computer Engineering and Applications*.

Kumar, A. (2018). Dimension Reduction - LLE. Retrieved May 27, 2019, from <https://blog.paperspace.com/dimension-reduction-with-lle/>

Lakshen, Guma Abdulkhader; Vraneš, S., & Janev, V. (2016). Big data and quality: A literature review. *Telecommunications Forum (TELFOR)*, 1–4.

Mao, Xue-min; Cai, Chuan-xi; Sun, B. (2012). Comparative research on methods of dimensionality reduction in high-dimension medical data. *Advanced Computational Intelligence (IWACI), 2011 Fourth International Workshop On*.

Niels Aakvaag, J.-E. F. (2006). Redes de sensores inalámbricos. *Revista ABB 2/2006*, 39–42. Retrieved from [http://www.materialelectrico.com.co/pdf/ABB/02-2006/39-42%25202M631\\_SPA72dpi.pdf](http://www.materialelectrico.com.co/pdf/ABB/02-2006/39-42%25202M631_SPA72dpi.pdf)

Ortuño, M., Barandiarán, José María, & Gil, M. (2014). *Texturas*. Trama Editorial. Retrieved from [https://books.google.com.ec/books?id=VT\\_HCQAAQBAJ&pg=PA70&lpg=PA70&dq=La+cantidad+de+información+que+es+almacenada+crece+cuatro+veces+más+rápido&source=bl&ots=iVElvkVTeK&sig=ACfU3U1yphuPpHA5XhWJmdxS4c4w-def8A&hl=es-419&sa=X&ved=2ahUKEwj8hZG](https://books.google.com.ec/books?id=VT_HCQAAQBAJ&pg=PA70&lpg=PA70&dq=La+cantidad+de+información+que+es+almacenada+crece+cuatro+veces+más+rápido&source=bl&ots=iVElvkVTeK&sig=ACfU3U1yphuPpHA5XhWJmdxS4c4w-def8A&hl=es-419&sa=X&ved=2ahUKEwj8hZG)

Parmesh, R. (2014). Scheduling Issues in Embedded Real-Time Systems. IEEE.

Pérez, D. (2009). Sistemas Embebidos y Sistemas Operativos Embebidos. Retrieved from [https://s3.amazonaws.com/academia.edu.documents/39032903/Info\\_Sistemas\\_Embebidos.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1554314933&Signature=w4U%2F2C%2Bslh%2FZ0%2BZ11sy02sK6dkk%3D&response-content-disposition=inline%3B+filename%3DInfo\\_Sistemas\\_Em](https://s3.amazonaws.com/academia.edu.documents/39032903/Info_Sistemas_Embebidos.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1554314933&Signature=w4U%2F2C%2Bslh%2FZ0%2BZ11sy02sK6dkk%3D&response-content-disposition=inline%3B+filename%3DInfo_Sistemas_Em)

Pérez Marquez, M. (2015). BIG DATA - Técnicas, herramientas y aplicaciones - María PÉREZ MARQUÉS - Google Libros, 356. Retrieved from [https://books.google.com.ec/books?hl=es&lr=&id=R6ywDAAAQBAJ&oi=fnd&pg=PT10&dq=aplicaciones+de+big+data&ots=L\\_TODjRLjx&sig=NkLVrfdO3SKWUvIANxX8bNALu\\_A#v=onepage&q&f=false](https://books.google.com.ec/books?hl=es&lr=&id=R6ywDAAAQBAJ&oi=fnd&pg=PT10&dq=aplicaciones+de+big+data&ots=L_TODjRLjx&sig=NkLVrfdO3SKWUvIANxX8bNALu_A#v=onepage&q&f=false)

Popkes, A.-L. (2019). *Kullback-Leibler Divergence*. Retrieved from [http://alpopkes.com/files/kl\\_divergence.pdf](http://alpopkes.com/files/kl_divergence.pdf)

Qing Wu, Zongxian Qi, Zhicang Wang, & Yu zhang. (2018). An improved weighted local linear embedding algorithm. *International Conference on Computational Intelligence and Security (CIS)*.

Quansheng Jiang, Minping Jia, & Jiayun Lv. (2008). New approach of intelligent fault diagnosis based on LLE algorithm. In *2008 Chinese Control and Decision Conference* (pp. 522–526). IEEE. <https://doi.org/10.1109/CCDC.2008.4597366>

Ramesh, U. B. K., Sentilles, S., & Crnkovic, I. (2012). Energy management in embedded systems:

- Towards a taxonomy. *2012 1st International Workshop on Green and Sustainable Software, GREENS 2012 - Proceedings*, 41–44. <https://doi.org/10.1109/GREENS.2012.6224254>
- Rosero-Montalvo, P. D., López-Batista, V. F., Peluffo-Ordóñez, D. H., Erazo-Chamorro, V. C., & Arciniega-Rocha, R. P. (2019). Multivariate Approach to Alcohol Detection in Drivers by Sensors and Artificial Vision (pp. 234–243). [https://doi.org/10.1007/978-3-030-19651-6\\_23](https://doi.org/10.1007/978-3-030-19651-6_23)
- Rosero-Montalvo, P., Peluffo-Ordóñez, D., López, V., Serrano, J., & Rosero, E. (2019). Intelligent System for Identification of Wheelchair User's Posture Using Machine Learning Techniques. *IEEE Sensors Journal*, *19*(5), 1936–1942. <https://doi.org/10.1109/JSEN.2018.2885323>
- Rosero, P. (2018a). Body position data with accelerometer sensor | IEEE DataPort. Retrieved May 31, 2019, from <https://iee-dataport.org/open-access/body-position-data-accelerometer-sensor>
- Rosero, P. (2018b). Data set Wheelchair sensors | IEEE DataPort. Retrieved May 31, 2019, from <https://iee-dataport.org/open-access/data-set-wheelchair-sensors>
- Rosero, P., & Peluffo, D. (2019). Alcohol Detection in Drivers by Sensors and Computer Vision | IEEE DataPort. Retrieved from <https://iee-dataport.org/open-access/alcohol-detection-drivers-sensors-and-computer-vision>
- Rosero, P., Pijal, J., Vásquez, C., Maya, E., Pupiales, C., Suárez, L., ... Peluffo, D. (2018). Wireless Sensor Networks for Irrigation in Crops Using Multivariate Regression Models. *IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, 1–6. <https://doi.org/10.1109/ETCM.2018.8580322>

- Roweis, S. T., & Saul, L. K. (2000). *Nonlinear Dimensionality Reduction by Locally Linear Embedding*. *New Series* (Vol. 290). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.456.5279&rep=rep1&type=pdf>
- San José Vieco, J., Pastor, J., García, P., & Zangroniz, R. (2011). Implementación del estándar EPCIS en arquitecturas basadas en Servicios Web para aplicaciones de identificación y localización. In *5ª Jornadas Científicas sobre RFID*.
- Sánchez Vítors, R. (2004). Aplicaciones de los sistemas embebidos. *Técnica Industrial Especial Electricidad y Electrónica*. Retrieved from <http://www.tecnicaindustrial.es/tiadmin/numeros/15/07/a07.pdf>
- Taneja, S., Gupta, C., Goyal, K., & Gureja, D. (2014). An Enhanced K-Nearest Neighbor Algorithm Using Information Gain and Clustering. *Fourth International Conference on Advanced Computing & Communication Technologies*.
- Valencia, J., Meza, A., Daza, G., Acosta, C., & Castellanos, G. (2010). *Selección automática de parámetros en LLE Automatic selection of parameters in LLE*. *Rev. Fac. Ing. Univ. Antioquia N.º* (Vol. 56). Diciembre. Retrieved from <http://www.scielo.org.co/pdf/rfiua/n56/n56a17.pdf>
- Van Rijsbergen, C. . (1979). *Information Retrieval*. Retrieved from [www.dcs.gla.ac.uk/Keith/Preface.html](http://www.dcs.gla.ac.uk/Keith/Preface.html)
- Vela, A. (2016). *Study of Precision Agriculture with a Focus on Implementing a Wireless Sensor Network (WSN) for Humidity and Temperature Monitoring in Crops-Case Study Cabalinus Farm Located in Los Rios Province*. *Revista Politécnica-Abril* (Vol. 38). Retrieved from

[http://repositorio.puce.edu.ec/bitstream/handle/22000/11112/Artículo científico Andrés Vela-Revista EPN.PDF?sequence=1&isAllowed=y](http://repositorio.puce.edu.ec/bitstream/handle/22000/11112/Artículo_científico_Andrés_Vela-Revista_EPN.PDF?sequence=1&isAllowed=y)

Verleysen, M., & Lee, J. A. (2013). *Nonlinear Dimensionality Reduction for Visualization*. Retrieved from <http://perso.uclouvain.be/michel.verleysen/>

Wen, Y., Zhou, Z., Wang, X., Zhang, Y., & Wu, R. (2010). AN IMPROVED LOCALLY LINEAR EMBEDDING FOR SPARSE DATA SETS.

Yang, Y., Wan, J., & Yang, Y. (2012). EXPLOITING ROTATION INVARIANCE WITH SVM CLASSIFIER FOR MICROCALCIFICATION DETECTION. *IEEE International Symposium on Biomedical Imaging (ISBI)*.

## GLOSARIO DE TÉRMINOS Y ACRÓNIMOS

**Algoritmo:** Es una secuencia de pasos lógicos que permiten solucionar un problema.

**BDD:** Su traducción al español hace referencia a Base de Datos.

**k-NN:** Algoritmo de aprendizaje automático, conocido como método k vecinos más cercanos.

**LE:** Lapacian Eigenmaps que en español son Mapas propios laplacianos.

**LLE:** Locally Linear Embedding que en español es Incrustación localmente lineal.

**Minería de datos:** Proceso de analizar grandes volúmenes de datos con la ayuda de algoritmos, inteligencia artificial y programas estadísticos.

**PCA:** Su significado en español es Análisis de componentes principales.

**RD:** Reducción de dimensionalidad.

**Set de datos:** Conjunto de datos.

**SNE:** Su significado en español es Incrustación estocástica de vecinos.

**SVM:** Support Vector Machine.

**t-SNE:** Su significado en español es Incrustación estocástica de vecinos distribuidos en T.

## ANEXOS

### ANEXO 1. Bases de datos originales

- Base de datos 1 (Posición del cuerpo).

	A	B	C	D
1	X	Y	Z	HUMAN
2	3188	212	16124	1
3	3188	-376	15460	1
4	1856	-376	15988	1
5	1856	-376	15988	1
6	3048	-376	15988	1
7	3048	-376	15412	1
8	3048	-376	16064	1
9	1800	-376	16064	1
10	1800	-376	14764	1
11	4028	-376	16796	1
12	1808	872	15300	1
13	1808	-188	17004	1
14	1808	-188	15160	1
15	1808	612	15804	1
16	3520	-420	15372	1
17	3520	328	15372	1
18	3520	328	15904	1
19	3520	-176	15904	1
20	3520	-1124	15904	1

Figura 77 Primeros 20 registros originales de la BDD1

Fuente: Autoría propia

	A	B	C	D
639	5436	-2192	-17044	5
640	5436	-2192	-16344	5
641	4296	-1196	-14648	5
642	4296	-1196	-13700	5
643	5636	-1788	-13700	5
644	5636	-1788	-15644	5
645	5636	-1788	-15644	5
646	5636	-1788	-16196	5
647	5636	-1788	-16196	5
648	5636	-1788	-16752	5
649	5636	-1788	-16052	5
650	3620	-1788	-17520	5
651	3620	-1788	-16700	5
652	3620	-1788	-16700	5
653	8404	-1788	-14316	5
654	2460	172	-16464	5
655	2460	172	-16464	5
656	352	-784	-15760	5
657	352	584	-17748	5
658	2352	1196	-17748	5
659	2352	-512	-17748	5
660	-300	1668	-15052	5
661	984	1668	-18204	5
662	984	3076	-15540	5
663	984	1964	-16256	5

Figura 78 Últimos registros originales de la BDD1

Fuente: Autoría propia

- Base de datos 2 (Silla de ruedas).

	A	B	C	D	E
1	s1	s2	s3	s4	label
2	431	243	470	3	1
3	396	247	474	3	1
4	430	237	416	2	1
5	464	255	314	3	1
6	431	221	363	2	1
7	502	261	246	2	1
8	503	259	251	1	1
9	506	261	222	3	1
10	489	258	323	3	1
11	498	246	350	2	1
12	487	266	327	2	1
13	486	264	352	3	1
14	399	277	418	5	1
15	503	265	413	4	1
16	476	251	461	3	1
17	408	203	372	3	1
18	425	208	361	1	1
19	472	270	382	2	1
20	592	204	290	2	1
21	627	252	197	2	1
22	0	278	341	6	1
23	0	278	378	4	1
24	351	470	397	4	1
25	650	345	181	3	1
26	587	332	198	4	1
27	578	340	178	4	1
28	594	357	232	5	1
29	597	368	242	3	1
30	502	262	217	4	1

Figura 79 Primeros 30 registros originales de BDD2

Fuente: Autoría propia

	A	B	C	D	E
281	683	627	0	3	4
282	683	627	0	2	4
283	681	627	0	2	4
284	681	623	0	2	4
285	682	623	0	2	4
286	682	623	0	2	4
287	658	621	0	0	4
288	663	622	0	1	4
289	671	619	0	3	4
290	678	628	0	0	4
291	551	493	9	3	4
292	563	508	10	1	4
293	563	514	11	0	4
294	573	517	8	0	4
295	485	521	11	3	4
296	522	502	13	0	4
297	438	480	7	1	4
298	459	430	7	3	4
299	362	340	9	3	4
300	489	492	11	0	4
301	502	493	2	3	4
302	515	500	10	0	4
303	515	488	12	0	4
304	496	501	5	2	4
305	536	522	9	0	4
306	529	528	7	6	4
307	532	567	9	1	4
308	539	577	9	0	4
309	585	591	11	1	4
310					

Figura 80 Últimos registros originales de BDD2

Fuente: Autoría propia



- Base de datos 3 (Tipo de pisada).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	Clase
2	944	943	891	891	0	936	936	972	969	0	891	184	991	959	1
3	944	942	772	772	0	936	936	971	968	0	772	531	991	959	1
4	944	942	628	628	0	936	936	970	968	0	628	538	991	959	1
5	944	942	489	489	0	936	935	971	968	0	489	531	991	958	1
6	961	957	726	726	0	950	951	986	984	0	726	0	1000	961	1
7	962	959	567	567	0	951	951	568	810	0	567	0	1002	962	1
8	945	943	552	552	0	937	937	971	968	0	552	353	991	958	1
9	948	946	726	726	0	940	940	725	972	0	726	0	992	959	1
10	945	943	892	892	0	937	937	970	967	0	892	295	991	958	1
11	944	943	956	956	0	937	936	970	967	0	956	169	991	958	1
12	944	943	956	956	0	937	936	956	968	0	956	3	991	957	1
13	945	943	835	835	0	937	937	775	968	0	835	103	991	957	1
14	944	943	751	751	0	937	937	970	966	0	751	382	991	957	1
15	940	939	759	759	0	933	932	964	961	0	759	653	988	952	1
16	954	945	488	488	0	936	936	539	970	0	488	0	992	957	1
17	939	936	952	952	0	930	930	963	960	0	952	719	987	950	1
18	945	943	957	957	0	937	937	969	966	0	957	353	993	958	1
19	932	930	946	946	0	924	924	957	954	0	946	609	985	945	1
20	945	943	957	957	0	937	937	968	965	0	957	385	992	957	1
21	945	943	957	957	0	937	937	968	965	0	957	375	992	956	1
22	937	935	874	874	0	928	928	960	957	0	874	592	986	948	1
23	933	931	947	947	0	925	924	957	954	0	947	550	985	945	1
24	941	939	954	954	0	933	933	963	961	0	954	515	988	952	1
25	945	943	957	957	0	937	937	969	966	0	957	197	990	956	1
26	945	943	941	941	0	937	937	967	964	0	941	454	991	955	1

Figura 81 Primeros registros originales de BDD3  
Fuente: Autoría propia

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
377	937	935	874	0	0	928	928	960	957	0	0	592	986	948	3
378	933	931	947	0	0	925	924	957	954	0	0	550	985	945	3
379	941	939	954	0	0	933	933	963	961	0	0	515	988	952	3
380	945	943	957	0	0	937	937	969	966	0	0	197	990	956	3
381	945	943	941	0	0	937	937	967	964	0	0	454	991	955	3
382	946	944	957	0	0	938	938	786	968	0	0	0	992	956	3
383	1023	1023	176	0	0	1023	1023	1023	1023	0	0	175	1023	1023	3
384	1023	1023	109	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
385	1023	1023	435	0	0	959	1023	1023	1023	0	0	1023	1023	1023	3
386	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	457	3
387	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
388	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
389	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
390	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
391	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
392	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
393	873	1023	698	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
394	628	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
395	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	474	3
396	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	603	3
397	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
398	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	1023	3
399	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	474	3
400	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	603	3
401	1023	1023	1023	0	0	1023	1023	1023	1023	0	0	1023	1023	603	3
402															

Figura 82 Últimos registros originales de BDD3  
Fuente: Autoría propia

- Base de datos 4 (Patrones en cultivos de rosas).

	A	B	C	D	E	F
1	HS (Analog)	L (Lux)	T (°)	CO2 (Analog)	HR (%)	clase
2	713	0.83	16.2	76	96.8	1
3	717	2.5	16.2	75	96.5	1
4	717	561.67	16.1	78	96.7	1
5	716	640	15.9	79	96.5	1
6	718	568.33	16.1	79	96.6	1
7	712	935.83	16.4	80	96.1	1
8	717	971.67	16.2	80	96.5	1
9	716	675	16.2	81	96.9	1
10	712	835	16.2	84	96.9	1
11	703	835	16.8	85	94.7	1
12	713	335.83	16.3	90	95.3	1
13	717	420	16.3	91	95.9	1
14	716	380	16.4	90	95.2	1
15	787	2357.5	22.7	157	73	1
16	788	2334.17	23.1	158	72	1
17	788	2420	23.4	154	69.3	1
18	788	2578.33	23.8	156	69	1
19	787	2683.33	24.4	156	71.3	1
20	785	2875	25.2	159	76	1

Figura 83 Primeros 20 registros originales de BDD4

Fuente: Autoría propia

	A	B	C	D	E	F
278	569	4181.67	21.3	158	89	4
279	562	5141.67	21.1	163	89.5	4
280	567	5918.33	21.2	163	89.3	4
281	560	4555	21.3	157	89.8	4
282	562	6480.83	21.4	162	89.8	4
283	555	6013.33	21.4	159	87.2	4
284	555	2853.33	21.3	158	85.3	4
285	549	2811.67	21.3	156	85.4	4
286	554	2160.83	21.1	154	84.9	4
287	554	1880	20.9	154	85.9	4
288	547	1760	20.9	154	87.3	4
289	553	2052.5	20.7	156	86.8	4
290	553	2418.33	20.8	154	87.7	4
291	552	3710.83	20.8	152	87.1	4
292	552	2827.5	20.7	154	87.1	4
293	550	3030	20.8	154	85.8	4
294	549	3824.17	20.9	153	86.1	4
295	552	2950	21	154	85.9	4
296	550	1140	21	154	84.8	4
297	550	280.83	21	153	86.7	4
298	549	260.83	20.8	151	84.3	4
299	550	215	20.6	148	86.3	4
300	549	215.83	20.4	151	87.8	4
301	544	171.67	20.5	147	90.2	4
302	140	0	22	36	68.9	4

Figura 84 Últimos registros originales de BDD4

Fuente: Autoría propia

- Base de datos 5 (Detección de alcohol mediante visión artificial).

	A	B	C	D	E	F
1	ALCOHOL	TEMP_AMB	TEMP_FAC_MAX	TEMP_FAC_MIN	EYES	LABEL
2	0.2	27.94	27.23	11.75	353.553	1
3	0.2	27.83	27.26	12.62	353.553	1
4	0.2	27.94	27.49	12.09	353.553	1
5	0.2	28.17	26.7	13.01	353.553	1
6	0.2	28.11	27.15	13.26	353.553	1
7	0.03	27.42	27.34	11.54	430.116	1
8	0	28.16	26.81	14.56	430.116	1
9	0.42	27.68	28.1	11.49	4.301.196	1
10	0.88	27.69	27.34	12.19	4.301.116	1
11	0	28.42	27.1	16.27	380.789	1
12	1.1	27.89	27.26	16.2	380.789	1
13	0.04	27.99	27.74	16.1	380.789	1
14	0.06	28.27	27.1	16.25	353.553	1
15	0	28.74	27.8	13.26	291.548	1
16	0	28.3	27.17	12.66	380.789	1
17	0.19	28.69	27.43	15.69	291.548	1
18	0	28.27	27.8	13.48	430.116	1
19	0.12	28.21	27.28	12.99	291.548	1
20	0	28.42	27.32	16.24	380.789	1

Figura 85 Primeros 20 registros originales de BDD5  
Fuente: Autoría propia

	A	B	C	D	E	F
367	1.99	26.34	28.47	11.11	853.553	2
368	1.84	26.45	28.94	13.59	865.083	2
369	1.76	26.32	28.88	10.14	863.134	2
370	2.36	26.45	28.81	10.79	874.643	2
371	1.98	26.34	29.42	12.82	919.239	2
372	2.24	26.32	28.32	10.89	804.152	2
373	2.16	26.45	29.57	12.77	982.344	2
374	2.24	27.12	28.82	12.58	961.769	2
375	2.33	27.43	27	12.42	804.152	2
376	2.26	25.89	27.86	12.74	863.134	2
377	2.31	25.89	28.17	11.97	804.152	2
378	2.23	26.45	27.69	17.47	804.152	2
379	1.75	27.32	27.76	17.2	804.152	2
380	1.84	25.78	27.33	13.8	853.553	2
381	0.55	25.78	27.84	8.11	804.152	2
382	1.89	26.75	28.24	8.57	804.152	2
383	2.57	26.54	28.64	10.13	121.037	2
384	2.34	25.89	28.9	11.38	804.152	2
385	3.11	25.89	29.36	12.56	804.152	2
386	2.36	26.55	28.43	10.08	863.134	2
387	2.32	26.66	28.17	11.5	604.152	2
388	1.98	26.67	28.8	11.52	504.152	2
389	2.63	27.23	28.08	12.26	853.553	2
390	1.98	28.01	28.64	11.75	982.344	2
391	0.58	26.56	28.7	12.2	961.769	2
392						

Figura 86 Últimos registros originales de BDD5  
Fuente: Autoría propia

## ANEXO 2. Bases de datos con RD.

- Base de datos 1 con PCA.

	A	B	C	D
1	,"pc1","pc2","HUMAN_POSITION"			
2	1,-0.396395471806844,-0.12768253136641,1			
3	2,-0.380738657266163,-0.136504208117535,1			
4	3,-0.440051512980107,-0.140672895840633,1			
5	6,-0.385202848097547,-0.135893957126913,1			
6	7,-0.396352804888559,-0.142498666618587,1			
7	9,-0.421233710874299,-0.128224264776334,1			
8	10,-0.371875528714946,-0.150781872876107,1			
9	12,-0.460613663740121,-0.145951266038793,1			
10	13,-0.429079123367994,-0.127271688887494,1			
11	14,-0.445944847500438,-0.112641691906115,1			
12	15,-0.366378753027438,-0.137070318648977,1			
13	17,-0.380948729678203,-0.122680758704905,1			
14	18,-0.37726160640306,-0.136007568973361,1			
15	19,-0.370326303099815,-0.161074664478314,1			
16	22,-0.492240275807654,-0.165501521646162,1			
17	23,-0.466549817533742,-0.196200857705489,1			
18	25,-0.47592365755191,-0.184765363241837,1			
19	26,-0.428358117314738,-0.172325962283249,1			
20	27,-0.410552617201491,-0.182583109231656,1			

Figura 87 Primeros 20 registros de BDD1 aplicados PCA

Fuente: Autoría propia

	A	B	C	D
509	635,0.282076371943738,0.124566022611681,5			
510	636,0.237934953001387,0.140418326423298,5			
511	638,0.273265567195595,0.142748647703118,5			
512	639,0.261294754689907,0.135657701929848,5			
513	640,0.181969349185905,0.145823502903207,5			
514	641,0.165757448821059,0.136220336341693,5			
515	642,0.220673801037481,0.119379624763775,5			
516	644,0.253918457481849,0.139072194168398,5			
517	645,0.26335829820062,0.144663911406748,5			
518	646,0.26335829820062,0.144663911406748,5			
519	647,0.27286654356228,0.150296148335231,5			
520	648,0.260895731056592,0.143205202561961,5			
521	649,0.209895564733832,0.159861740650745,5			
522	650,0.195872612941455,0.151555204173486,5			
523	651,0.195872612941455,0.151555204173486,5			
524	653,0.133707540606947,0.202018577279191,5			
525	654,0.133707540606947,0.202018577279191,5			
526	655,0.0490844516772159,0.17147578683011,5			
527	658,0.156592348083828,0.197034680937362,5			
528	659,-0.00557453424830207,0.229717345532292,5			
529	660,0.0967997512426007,0.260509469311427,5			
530	661,0.0409417242142057,0.270753809766174,5			
531	662,0.0613212368032753,0.248603237041002,5			

Figura 88 Últimos registros de BDD1 aplicados PCA

Fuente: Autoría propia

- Base de datos 1 con LE.

	A	B
1	,"Y1","Y2","HUMAN_POSITIO	
2	4,0,-0.111439981506691,1	
3	5,0,-0.111439981506691,1	
4	8,0,-0.111439981506691,1	
5	11,0,-0.111439981506691,1	
6	16,0,-0.111439981506691,1	
7	20,0,-0.111439981506691,1	
8	21,0,-0.111439981506691,1	
9	24,0,-0.111439981506691,1	
10	31,0,-0.111439981506691,1	
11	32,0,-0.111439981506691,1	
12	50,0,-0.111439981506691,1	
13	53,0,-0.111439981506691,1	
14	59,0,-0.111439981506691,1	
15	65,0,-0.111439981506691,1	
16	67,0,-0.111439981506691,1	
17	68,0,-0.111439981506691,1	
18	87,0,-0.111439981506691,1	
19	88,0,-0.111439981506691,1	
20	89,0,-0.111439981506691,1	

Figura 89 Primeros registros de BDD1 aplicados LE  
Fuente: Autoría propia

	A	B	C
108	538,-0.0967505640439538,0,4		
109	543,-0.0957230453055206,0,4		
110	546,-0.12097007259836,0,4		
111	548,-0.124151898165174,0,4		
112	554,-0.0921688557225778,0,4		
113	562,-0.0777600460055489,0,5		
114	570,-0.108521636242242,0,5		
115	572,-0.0809598616953328,0,5		
116	575,-0.0747189571674903,0,5		
117	581,-0.0747189571674903,0,5		
118	583,-0.0747189571674903,0,5		
119	585,-0.0747189571674903,0,5		
120	589,-0.0747189571674903,0,5		
121	593,-0.0747189571674903,0,5		
122	596,-0.0747189571674903,0,5		
123	599,-0.0747189571674903,0,5		
124	614,-0.0747189571674903,0,5		
125	617,-0.0747189571674903,0,5		
126	619,-0.0747189571674903,0,5		
127	621,-0.075308627077664,0,5		
128	622,-0.0747189571674904,0,5		
129	637,-0.0747189571674906,0,5		
130	643,-0.0747189571674904,0,5		
131	652,-0.0984466631780384,0,5		
132	656,-0.0754040232457878,0,5		
133	657,-0.0754040232457878,0,5		

Figura 90 Últimos registros de BDD1 aplicados LE  
Fuente: Autoría propia

- Base de datos 1 con LLE.

	A	B	C	D
1	,"Y1","Y2","HUMAN_POSITION"			
2	4,-1.78896624887627,-0.504384722619402,1			
3	5,-1.66533038793605,-0.528460967664462,1			
4	8,-1.79729858894591,-0.503115700983976,1			
5	11,-1.67280215722905,-0.551797806543952,1			
6	16,-1.5535782515454,-0.564058683339502,1			
7	20,-1.79465938171506,-0.485296881693487,1			
8	21,-1.968746642538,-0.453832535466464,1			
9	24,-1.8907405703237,-0.18647809405957,1			
10	31,-1.83298671633745,-0.331157317241974,1			
11	32,-2.13427976531204,-0.400074269281531,1			
12	50,-1.8085722020152,-0.50387037634261,1			
13	53,-1.90645628675506,-0.482251074442901,1			
14	59,-1.85342625352335,-0.484645026593896,1			
15	65,-2.04528628743145,-0.431794047107537,1			
16	67,-2.00134278343115,-0.346706767118537,1			
17	68,-1.99899712384536,-0.426791133132236,1			
18	87,-1.54939731553672,-0.630133681167389,1			
19	88,-1.4175958199049,-0.65218287279975,1			
20	89,-1.22536489650352,-0.70049047794257,1			

Figura 91 Primeros registros de BDD1 aplicados LLE

Fuente: Autoría propia

	A	B	C	D
109	543,0.896830801705585,0.236928456570433,4			
110	546,0.759922682481908,0.443077242432982,4			
111	548,0.736694407482267,0.47933826565382,4			
112	554,0.855825409106991,0.320051765703652,4			
113	562,0.522158372292969,0.523256117765588,5			
114	570,0.767209809875091,0.227856510707036,5			
115	572,0.796329146888799,0.159571560392705,5			
116	575,0.773572321574864,0.0413736083600021,5			
117	581,0.831980416711312,0.090322342377889,5			
118	583,0.743431915949851,0.165969957361693,5			
119	585,0.741469265780941,0.21493647852619,5			
120	589,0.719904961971816,0.203374405779825,5			
121	593,0.784287893765574,0.156219429058251,5			
122	596,0.784220824679922,0.0904416225892055,5			
123	599,0.747401894321152,0.140088936487377,5			
124	614,0.766247107836928,0.145595364395609,5			
125	617,0.784739340819911,0.088038477049039,5			
126	619,0.783161127614519,0.0523769709669922,5			
127	621,0.783857105946178,-0.0130463450355933,5			
128	622,0.782427849760857,0.0203300903350994,5			
129	637,0.796595233318941,0.130518078175077,5			
130	643,0.760203596213879,0.121821628657132,5			
131	652,0.790969699911402,0.169778705373379,5			
132	656,0.795274721819188,-0.0758297029341515,5			
133	657,0.778937747413656,-0.0435906855115885,5			

Figura 92 Últimos registros de BDD1 aplicados LLE

Fuente: Autoría propia

- Base de datos 1 con SNE.

	A	B	C	D
1	,"Y1","Y2","HUMAN_POSITION"			
2	4,-0.0180104312702032,-1.74041821276551,1			
3	5,0.378691292886357,-1.548931478755,1			
4	8,0.0957731823089065,-1.40151962349664,1			
5	11,-0.139609740339703,-1.536942980233,1			
6	16,0.0403867407250834,-1.25171287168501,1			
7	20,0.819558435748452,-0.826808259463447,1			
8	21,0.893757938293757,-1.40901349833039,1			
9	24,0.627238944542496,-1.33889051388938,1			
10	31,0.405124129181652,-1.30467102003675,1			
11	32,0.508311494089312,-1.59650129461243,1			
12	50,0.438136061511784,-1.55531598007558,1			
13	53,0.510447337896746,-1.47437027083384,1			
14	59,0.775644282002129,-0.99828420693401,1			
15	65,0.274341444357256,-1.0924448579169,1			
16	67,0.0907042219653046,-1.2796777260828,1			
17	68,0.525339918572942,-1.14953898244479,1			
18	87,0.121382282772371,-1.20877393068059,1			
19	88,0.151080420454868,-1.46269131052209,1			
20	89,-0.288955718771363,-1.22960138799536,1			

Figura 93 Primeros registros de BDD1 aplicados SNE

Fuente: Autoría propia

	A	B	C	D
109	543,-0.0133167975492524,0.798347236458137,4			
110	546,0.228255872043992,1.06628625419938,4			
111	548,0.132376724607706,0.622401784739143,4			
112	554,0.0701948304912901,0.780539810589789,4			
113	562,-0.22404272416271,1.37227440357335,5			
114	570,-0.345863597299,1.40234027077454,5			
115	572,-0.451652663029426,1.89257545362709,5			
116	575,-0.0086628434968736,1.79392606439514,5			
117	581,-0.277714261591274,1.48491517860065,5			
118	583,-0.0175197606395611,1.66097850800433,5			
119	585,-0.425775224597743,1.75455513465162,5			
120	589,-0.381916808471699,1.59279967795416,5			
121	593,-0.240560916527538,1.5934725354493,5			
122	596,-0.0260532046659561,1.82414630737543,5			
123	599,0.199772993045854,1.67926578907775,5			
124	614,-0.459502159721575,1.57618852919322,5			
125	617,-0.532677839283671,1.86514093849668,5			
126	619,0.323150265423939,1.54959263317565,5			
127	621,-0.184777958520963,1.88460369089378,5			
128	622,-0.260783344589214,1.71468159199491,5			
129	637,-0.21616157907903,1.86429539840963,5			
130	643,0.0631010257848388,1.79002866531778,5			
131	652,-0.170919739460583,1.18658440361363,5			
132	656,-0.655877890113507,1.68715767968108,5			
133	657,0.244013708411268,1.59544886332286,5			

Figura 94 Últimos registros de BDD1 aplicados SNE

Fuente: Autoría propia

- Base de datos 1 con *t*SNE.

	A	B	C	D
1	,"V1","V2","cl_kmeans"			
2	1,-17.7676882157704,0.114566978572528,"3"			
3	2,-18.5090115640295,0.233433947735053,"3"			
4	3,-17.6970295575112,0.111839127175196,"3"			
5	4,-17.822121783179,-0.349022824226284,"3"			
6	5,-18.7185381326826,-0.0256486863745917,"3"			
7	6,-18.0225989057782,0.455742184110743,"3"			
8	7,-16.8889000953406,0.163892272175989,"3"			
9	8,-16.574808243093,0.901959356968658,"3"			
10	9,-17.6319681192683,0.772692587887006,"3"			
11	10,-16.1904156049462,0.443184383210039,"3"			
12	11,-16.7795102752505,-0.459860539220048,"3"			
13	12,-16.7047668259467,-0.206723971703038,"3"			
14	13,-17.9480256944378,0.531197495699487,"3"			
15	14,-16.414404583064,0.208745791156779,"3"			
16	15,-17.1401364285022,0.910184795156107,"3"			
17	16,-17.1673803991778,0.662808244144203,"3"			
18	17,-17.6675148858151,-1.07863959262993,"3"			
19	18,-18.5450340623124,-0.810296173169954,"3"			
20	19,-18.7585391677425,-0.726348729150937,"3"			

Figura 95 Primeros registros de BDD1 aplicados *t*SNE

Fuente: Autoría propia

	A	B	C	D
105	104,-3.73596665975511,-17.8013594842545,"2"			
106	105,-4.32293792305892,-18.4838441357533,"4"			
107	106,-3.42212760081044,-18.1886483735535,"2"			
108	107,-2.87338079949671,-18.0272320649061,"2"			
109	108,-3.82815451969184,-18.5693496233844,"2"			
110	109,-4.95903478606827,-18.0377534146363,"4"			
111	110,-5.1239108706939,-18.0334275301273,"4"			
112	111,-4.87121748734848,-18.6117559614419,"4"			
113	112,18.4611352707033,8.69797398329828,"5"			
114	113,17.9601633219823,9.55491026315625,"5"			
115	114,17.7259654305628,9.47766358027235,"5"			
116	115,17.5513940421412,8.07514012172533,"5"			
117	116,17.4505899959136,9.35953651195058,"5"			
118	117,18.0618898406967,8.89858888189323,"5"			
119	118,18.1300017371517,8.4375270991507,"5"			
120	119,18.1057489773473,8.92670684553863,"5"			
121	120,17.9933776558986,9.20569486292895,"5"			
122	121,17.2575818537751,8.99706944454022,"5"			
123	122,17.9968711978992,8.52109522749372,"5"			
124	123,17.9504093544451,8.5390895622821,"5"			
125	124,17.1716776292106,8.63165877337214,"5"			
126	125,17.173805230658,8.59730661606919,"5"			
127	126,17.380435509761,8.30895363995222,"5"			
128	127,17.4988587230405,8.22687782356729,"5"			
129	128,17.2530290416704,8.85516084853287,"5"			
130	129,17.334004087546,9.00688471079036,"5"			

Figura 96 Últimos registros de BDD1 aplicados *t*SNE

Fuente: Autoría propia



- Base de datos 2 con PCA.

	A	B	C	D
1	"pc1","pc2","label"			
2	4,-0.466133414854839,0.0641289823411076,1			
3	5,-0.585773287440931,0.0691878278863993,1			
4	8,-0.301360975489921,0.0321881007203597,1			
5	11,-0.468919781776637,0.0962641222663195,1			
6	16,-0.622974262638039,0.0558692223339183,1			
7	20,-0.221979546412964,0.139362331756642,1			
8	21,-0.67485897678686,-0.405663758439015,1			
9	24,-0.10757754761661,0.146979198226992,1			
10	31,-0.44924364604725,0.134942881882408,1			
11	32,-0.461824675746334,0.142122134207011,1			
12	50,-0.700371889436953,0.309667509446233,1			
13	53,-0.568047784554716,0.0603467035769455,1			
14	59,-0.091841548847868,-0.0332668651205991,1			
15	65,-0.448954578165478,-0.572942314548289,1			
16	67,-0.694745335889905,-0.426335563849944,1			
17	68,-0.700091214109806,-0.423071400826546,1			
18	87,-0.56772941958663,0.289662745793671,1			
19	88,-0.595305347618546,0.291270109156743,1			
20	89,0.218155988587554,-0.00196475478525734,2			

Figura 97 Primeros registros de BDD2 aplicados PCA

Fuente: Autoría propia

	A	B	C	D
38	179,0.333069509742578,0.047380386874631,2			
39	181,0.28896417932891,0.0320142322888688,2			
40	189,-0.595668225237062,0.220367580856923,3			
41	193,-0.637211326123539,0.114406931745939,3			
42	195,-0.635937999640672,0.267797267380705,3			
43	202,-0.589439743708352,0.252750704216782,3			
44	219,0.378867428974385,-0.0023928677968987,4			
45	220,0.379549891729122,-0.0966551730861907,4			
46	222,0.40976122555837,0.000622928249093484,4			
47	230,0.341020623873184,-0.0558308333414041,4			
48	238,0.356482311345706,0.0155008431359173,4			
49	240,0.367296699984712,0.0151701980094563,4			
50	248,0.385515368262552,0.0125754692112811,4			
51	249,0.385926320945344,0.0136501955799377,4			
52	256,0.146779212069909,-0.0865293177445551,4			
53	260,0.15011662470659,-0.0823685943089758,4			
54	261,0.100952027283479,-0.149285624375863,4			
55	262,0.122670105009075,-0.124459465801434,4			
56	264,0.34851770550105,0.0300014629799958,4			
57	271,0.408379344975884,0.0860714083970725,4			
58	277,0.411020969591942,-0.0027997827819577,4			
59	294,0.236841092245657,-0.184173968861645,4			
60	296,0.186146392673689,-0.240499912388209,4			
61	297,0.155290114964922,-0.209162825712993,4			
62	300,0.234050389314434,-0.171468893644395,4			
63	301,0.228654377129115,-0.159284502838555,4			

Figura 98 Últimos registros de BDD2 aplicados PCA

Fuente: Autoría propia

- Base de datos 2 con LE.

	A	B	
1	,"Y1", "Y2", "label"		
2	4,0.238211791239862,0,1		
3	5,0.238211791239862,0,1		
4	8,0.238211791239862,0,1		
5	11,0.238211791239862,0,1		
6	16,0.238211791239862,0,1		
7	20,0.238211791239862,0,1		
8	21,0.229338152149385,0,1		
9	24,0.239929556962008,0,1		
10	31,0.239929556962008,0,1		
11	32,0.239929556962008,0,1		
12	50,0.23251049043223,0,1		
13	53,0.240745897674594,0,1		
14	59,0.240745897674594,0,1		
15	65,0.220156888131632,0,1		
16	67,0.242394894809971,0,1		
17	68,0.242394894809971,0,1		
18	87,0.22854727263725,0,1		
19	88,0.214579836437623,0,1		
20	89,0,0.167269733718999,2		

Figura 99 Primeros registros de BDD2 aplicados LE  
Fuente: Autoría propia

	A	B	C
38	179,0,0.19946951065611,2		
39	181,0,0.19946951065611,2		
40	189,0,0.031980470564605,3		
41	193,0,0.031980470564605,3		
42	195,0,0.0319804705646049,3		
43	202,0,0.0319804705646049,3		
44	219,0,-0.127769989890768,4		
45	220,0,-0.133263737413918,4		
46	222,0,-0.133263737413918,4		
47	230,0,-0.13148571568191,4		
48	238,0,-0.13148571568191,4		
49	240,0,-0.13148571568191,4		
50	248,0,-0.13148571568191,4		
51	249,0,-0.13148571568191,4		
52	256,0,-0.0958960345821674,4		
53	260,0,-0.0958960345821674,4		
54	261,0,-0.0924063143179899,4		
55	262,0,-0.09240631431799,4		
56	264,0,-0.130984374259423,4		
57	271,0,-0.115857620952179,4		
58	277,0,-0.12242340420445,4		
59	294,0,-0.134306309273013,4		
60	296,0,-0.127750331797878,4		
61	297,0,-0.116625657229962,4		
62	300,0,-0.13384592089922,4		
63	301,0,-0.13384592089922,4		

Figura 100 Últimos registros de BDD2 aplicados LE  
Fuente: Autoría propia

- Base de datos 2 con LLE.

	A	B	C	D
1	,"Y1","Y2","label"			
2	4,0.143511194669342,-1.36228965959093,1			
3	5,0.140092208705791,-1.65230476457666,1			
4	8,0.145811549506015,-0.924482986180512,1			
5	11,0.143221010417629,-1.37406501720088,1			
6	16,0.139662874219328,-1.72702817828449,1			
7	20,0.153104348289979,-0.668360359907952,1			
8	21,-0.41395252679008,-1.10676445648202,1			
9	24,0.141435282250725,-0.448133588927619,1			
10	31,0.14661621015787,-1.32970228481632,1			
11	32,0.146268877233074,-1.37756880395766,1			
12	50,0.143309693507852,-1.86703162553894,1			
13	53,0.142433467520498,-1.60244983696298,1			
14	59,0.158714229616732,-0.18291907145127,1			
15	65,-0.880486773680573,-0.36154846326874,1			
16	67,-0.421342541115619,-1.08711353117298,1			
17	68,-0.41879193681554,-1.09574389820699,1			
18	87,0.160693777311957,-1.52090864969369,1			
19	88,0.159884207814342,-1.53935874866239,1			
20	89,-1.4787581264325,0.661230647385164,2			

Figura 101 Primeros registros de BDD2 aplicados LLE  
Fuente: Autoría propia

	A	B	C	D
38	179,-1.29405123148281,0.772824077005725,2			
39	181,-1.03316370759109,0.728103456758369,2			
40	189,0.14561784364223,-1.76372270992552,3			
41	193,0.151852641498657,-1.72882064291303,3			
42	195,0.149981396016512,-1.83214308862031,3			
43	202,0.155443984515622,-1.66460655064269,3			
44	219,1.5300336208408,0.824220621883594,4			
45	220,1.62684180907753,0.805316342242677,4			
46	222,1.74716845005885,0.853659127733385,4			
47	230,1.52659867664286,0.785247784662503,4			
48	238,1.5031518751813,0.814194440653025,4			
49	240,1.64953673871713,0.825328971305978,4			
50	248,1.7111495713957,0.838887663668971,4			
51	249,1.71181630354669,0.839493733668152,4			
52	256,0.187257036982764,0.523281698412444,4			
53	260,0.206092684516902,0.526203885034054,4			
54	261,0.0447106713809043,0.483522218488905,4			
55	262,0.0968133143217087,0.502335888044382,4			
56	264,1.46534972190388,0.793041728636066,4			
57	271,1.67808391410924,0.860132689118472,4			
58	277,1.55274784917386,0.849647108897914,4			
59	294,1.05472958604863,0.671716886433279,4			
60	296,0.783314893341127,0.609297573758966,4			
61	297,0.457648518961446,0.561047994425336,4			
62	300,0.921065225376711,0.648255503911931,4			
63	301,0.992694491448797,0.661033374927974,4			

Figura 102 Últimos registros de BDD2 aplicados LLE  
Fuente: Autoría propia

- Base de datos 2 con SNE.

	A	B	C	D
1	"Y1", "Y2", "label"			
2	4,1.19118304054958e+87,2.44840788414653e+87,1			
3	5,1.20895847861013e+87,2.48494426986584e+87,1			
4	8,1.16778931636456e+87,2.40032343836531e+87,1			
5	11,1.19282603059944e+87,2.45178495584514e+87,1			
6	16,1.21377514493464e+87,2.4948446494484e+87,1			
7	20,1.15900215700111e+87,2.38226193875605e+87,1			
8	21,1.24256321019439e+87,2.55401685351502e+87,1			
9	24,1.14704746368399e+87,2.35768975765749e+87,1			
10	31,1.18829163718063e+87,2.44246477227127e+87,1			
11	32,1.18988855114107e+87,2.44574713660971e+87,1			
12	50,1.23545798721829e+87,2.53941247843474e+87,1			
13	53,1.20481275360454e+87,2.47642297211126e+87,1			
14	59,1.14382784774161e+87,2.3510720232004e+87,1			
15	65,1.21292996795857e+87,2.49310743692664e+87,1			
16	67,1.24608193307647e+87,2.56124938499955e+87,1			
17	68,1.24697767154821e+87,2.56309052365251e+87,1			
18	87,1.20680586984774e+87,2.48051970726864e+87,1			
19	88,1.21303686522207e+87,2.49332715809177e+87,1			
20	89,6.23530246771561e+86,1.28163037970736e+87,2			

Figura 103 Primeros registros de BDD2 aplicados SNE

Fuente: Autoría propia

	A	B	C	D
37	174,6.15579376668102e+86,1.26528782580165e+87,2			
38	179,6.12148737185209e+86,1.25823634464788e+87,2			
39	181,6.14220706926673e+86,1.26249515868499e+87,2			
40	189,-4.27359111655736e+86,-8.78411950957612e+86			
41	193,-4.23097732732249e+86,-8.6965293291055e+86,3			
42	195,-4.14522297935432e+86,-8.5202662238236e+86,3			
43	202,-4.27531825831766e+86,-8.78766954965883e+86			
44	219,-1.62210207520637e+87,-3.33413705167058e+87			
45	220,-1.62990169209954e+87,-3.35016871334548e+87			
46	222,-1.62957355721232e+87,-3.34949425105175e+87			
47	230,-1.62087839069753e+87,-3.33162183889733e+87			
48	238,-1.61894162050378e+87,-3.32764091971681e+87			
49	240,-1.62181535217602e+87,-3.33354770905579e+87			
50	248,-1.6248692297106e+87,-3.33982477780143e+87,4			
51	249,-1.62489272915768e+87,-3.33987307955652e+87			
52	256,-1.59838323603829e+87,-3.28538434880884e+87			
53	260,-1.59854764393719e+87,-3.28572227974153e+87			
54	261,-1.59680404394896e+87,-3.28213841075409e+87			
55	262,-1.59761389806965e+87,-3.28380301908572e+87			
56	264,-1.61842512754522e+87,-3.32657929829607e+87			
57	271,-1.62614957673332e+87,-3.34245645709699e+87			
58	277,-1.62764931423563e+87,-3.34553908084194e+87			
59	294,-1.61158648116742e+87,-3.31252285596912e+87			
60	296,-1.61050516688978e+87,-3.31030027697696e+87			
61	297,-1.60237180121186e+87,-3.29358261396747e+87			
62	300,-1.60920271934134e+87,-3.30762317132678e+87			

Figura 104 Últimos registros de BDD2 aplicados SNE

Fuente: Autoría propia

- Base de datos 2 con tSNE.

	A	B	C	D
1	,"V1","V2","cl_kmeans"			
2	1,-3.96430178835739,-5.88153574759806,"2"			
3	2,-4.01593461573114,-6.09871744022412,"2"			
4	3,-3.90852561992835,-5.58604884026446,"2"			
5	4,-4.01635517910299,-5.88510778103313,"2"			
6	5,-4.0082510624358,-6.13731008249994,"2"			
7	6,-4.07302580995835,-5.43045501529663,"2"			
8	7,-3.34069215164172,-6.40628351417933,"2"			
9	8,-4.00192333714757,-5.3383310822999,"2"			
10	9,-4.07538014333229,-5.78825673447271,"2"			
11	10,-4.11761576313209,-5.82665812050084,"2"			
12	11,-4.35548192034756,-6.26565360755169,"3"			
13	12,-4.02730662145014,-6.04056531119259,"2"			
14	13,-3.8363550548904,-5.34158555532615,"2"			
15	14,-3.20008231763429,-6.30299095432104,"2"			
16	15,-3.34158504741971,-6.40370562637997,"2"			
17	16,-3.34167631866912,-6.40547967455204,"2"			
18	17,-4.57878452101778,-5.78429174641664,"3"			
19	18,-4.57831630553123,-5.81560634897277,"3"			
20	19,-0.0184063720851601,-0.794388719186495,"4"			

Figura 105 Primeros registros de BDD2 aplicados tSNE  
Fuente: Autoría propia

	A	B	C	D
30	29,0.236605671904181,-1.17647504408468,"4"			
31	30,0.223312461418551,-1.19254323362663,"4"			
32	31,0.246231504133611,-1.19683538632089,"4"			
33	32,-0.877611282825851,-0.495121351740835,"4"			
34	33,-0.625333086364618,-0.894866385914028,"4"			
35	34,-0.731557609602449,-0.509148863179711,"4"			
36	35,-0.382511706477476,-0.759413658238657,"4"			
37	36,-0.445085420772494,-0.809974936114116,"4"			
38	37,-0.289657023120781,-0.315140965650131,"4"			
39	38,-0.135104770592351,-0.454236332609698,"4"			
40	39,1.308882021813,1.94018887728305,"4"			
41	40,1.31832793492499,1.95481870899986,"4"			
42	41,1.30863686615837,1.95970722945175,"4"			
43	42,1.28574170752729,1.90450840518892,"4"			
44	43,3.91216547668758,5.73631397172297,"1"			
45	44,4.01588407644213,5.5878744321465,"1"			
46	45,3.91911737114438,5.9290597367974,"1"			
47	46,3.88102250003968,5.60780786185664,"1"			
48	47,3.35840041870705,5.99859011214085,"1"			
49	48,3.36809632027385,6.02602704150301,"1"			
50	49,3.64621639761545,6.02501355037532,"1"			
51	50,3.6124680544249,6.06617493139349,"1"			
52	51,3.04305909285862,5.39465711716797,"1"			
53	52,3.04326357982423,5.48444356858996,"1"			
54	53,3.08265608824022,5.30329324872419,"1"			
55	54,3.07558480176812,5.37737923975919,"1"			

Figura 106 Últimos registros de BDD2 aplicados tSNE  
Fuente: Autoría propia

- Base de datos 3 con PCA.

	A	B	C	D
1	,"pc1","pc2","Clase"			
2	4,-0.428534498168269,-0.443778848600821,1			
3	5,-0.35930305048571,-0.180297763250495,1			
4	8,-0.678735613406581,-0.280280539548114,1			
5	11,-0.72912099687748,-0.0148437924151342,1			
6	16,-0.612078002232064,0.0434414512451186,1			
7	20,-0.580809538735048,0.0413282390492013,1			
8	21,-0.673254230776027,-0.0671705744744522,1			
9	24,-0.664638787588432,0.0124629587283483,1			
10	31,1.16573117966374,0.597570411401208,1			
11	32,1.16573117966374,0.597570411401208,1			
12	50,-0.678735613406581,-0.280280539548114,1			
13	53,-0.72912099687748,-0.0148437924151342,1			
14	59,-0.576651313862678,0.0414822125852137,1			
15	65,-0.632853853530231,0.030982252534331,1			
16	67,-0.560617386921031,0.0318704385358616,1			
17	68,-0.751474819698231,-0.0323545458314539,1			
18	87,-0.458376241032579,-0.293066284052588,1			
19	88,-0.428534498168269,-0.443778848600821,1			
20	89,-0.35930305048571,-0.180297763250495,1			

Figura 107 Primeros registros de BDD3 aplicados PCA

Fuente: Autoría propia

	A	B	C	D
35	179,-0.656539286518724,0.0416087424373487,2			
36	181,-0.817717091740322,0.365390499531445,2			
37	189,-0.794323687319835,0.89010650524476,2			
38	190,-0.759854171666246,0.89259403682645,2			
39	193,-0.764012396538615,0.892440063290437,2			
40	195,-0.952640695153694,0.830728148695237,2			
41	202,0.966930695503225,0.282805128612595,2			
42	206,0.969893642011653,1.50737960283149,2			
43	219,-0.498284528819451,0.465373275829063,2			
44	220,-0.656539286518724,0.0416087424373487,2			
45	222,-0.817717091740322,0.365390499531445,2			
46	230,-0.794323687319835,0.89010650524476,2			
47	238,-0.847841645391999,0.863574782969584,2			
48	240,-0.934677677501799,0.818757278409782,2			
49	248,0.969893642011653,1.50737960283149,2			
50	249,0.969893642011653,1.50737960283149,2			
51	256,0.969893642011653,1.50737960283149,2			
52	260,-0.498284528819451,0.465373275829063,2			
53	261,-0.656539286518724,0.0416087424373487,2			
54	262,-0.6460856757758,0.546821131450262,2			
55	264,-0.522145989773227,-0.0088847189728538,2			
56	271,-0.358886451000178,-1.01874121180861,3			
57	277,-0.432265843159949,-0.975369589906992,3			
58	294,1.19528869100687,-0.53069086803113,3			
59	296,1.50692865212712,-0.405738568585816,3			
60	297,1.50692865212712,-0.405738568585816,3			

Figura 108 Últimos registros de BDD3 aplicados PCA

Fuente: Autoría propia

- Base de datos 3 con LE.

	A	B	C	D
1	,"Y1","Y2","Clase"			
2	4,-0.0933057286448643,-0.131932704035797,1			
3	5,-0.0924340143204487,-0.133496153641444,1			
4	8,-0.0792693402652812,-0.0552396536582229,1			
5	11,-0.0963215598798822,-0.146518829833367,1			
6	16,-0.0949014043789862,-0.113584323224562,1			
7	20,-0.0927005844870014,-0.0945841444948969,1			
8	21,-0.0906141816464509,-0.0908643189189487,1			
9	24,-0.0936849181312376,-0.103502071593252,1			
10	31,-0.0867225238876284,-0.110646600056855,1			
11	32,-0.0867225238876284,-0.110646600056855,1			
12	50,-0.0787616894519697,-0.0544964995912084,1			
13	53,-0.0966591486554247,-0.147649118155516,1			
14	59,-0.0961011635727144,-0.101264463706199,1			
15	65,-0.0951655842146666,-0.0948656120663269,1			
16	67,-0.0959578368948547,-0.111156021523574,1			
17	68,-0.0760319539898342,-0.0466447329289063,1			
18	87,-0.094998962666992,-0.143844429116795,1			
19	88,-0.0842590996087057,-0.115966911202998,1			
20	89,-0.0934422572550335,-0.136070471550796,1			

Figura 109 Primeros registros de BDD3 aplicados LE

Fuente: Autoría propia

	A	B	C	D
34	173,-0.0517266816766629,0.155813770436993,2			
35	179,0.00437169094101049,0.184102574048445,2			
36	181,-0.00430831179366953,0.157380560512114,2			
37	189,-0.0491827971835147,0.190571773743315,2			
38	190,-0.0480787734117838,0.191088936059899,2			
39	193,-0.0480787734117838,0.191088936059899,2			
40	195,-0.0491827971835147,0.190571773743315,2			
41	202,0.0572096299269959,-0.0413193591137547,2			
42	206,-0.0507772460169767,0.14505403000674,2			
43	219,-0.0436522566522442,0.0484917609033979,2			
44	220,0.00523839458655657,0.180093295368937,2			
45	222,-0.00672777803458994,0.176106799314636,2			
46	230,-0.0485906365513637,0.187611981546036,2			
47	238,-0.0478706645276654,0.185532003544796,2			
48	240,-0.0504385061104897,0.181379760313949,2			
49	248,-0.04867238972357,0.133220318641765,2			
50	249,-0.0486723897235701,0.133220318641765,2			
51	256,-0.0486723897235701,0.133220318641765,2			
52	260,-0.0500488065476916,0.0893096887161729,2			
53	261,0.00631285186413762,0.174206307680392,2			
54	262,-0.035539526360526,0.141233630892528,2			
55	264,-0.0353068410755518,0.0488859946034484,2			
56	271,0.0989518103204006,0.067825173575133,3			
57	277,0.131687730807663,0.0344167002593527,3			
58	294,0.188200221816743,-0.0490092269747277,3			
59	296,0.192164452256028,-0.0624212291179418,3			

Figura 110 Últimos registros de BDD3 aplicados LE

Fuente: Autoría propia

- Base de datos 3 con LLE.

	A	B	C	D
1	,"Y1", "Y2", "Clase"			
2	4,0.63077982806162,-0.65547663018545,1			
3	5,0.855770322135292,-0.52059231375502,1			
4	8,0.77897559234545,-0.543573993920937,1			
5	11,1.33555044312324,-0.372820215948015,1			
6	16,1.51249032651026,-0.378871960768927,1			
7	20,1.43526092332621,-0.37417608851829,1			
8	21,1.37448257003884,-0.420290880709949,1			
9	24,1.39184910613228,-0.373056566769002,1			
10	31,-0.395749224964017,1.54330852443693,1			
11	32,-0.395749224964019,1.54330852443693,1			
12	50,0.778975592345453,-0.543573993920939,1			
13	53,1.33555044312324,-0.372820215948015,1			
14	59,1.42200809124432,-0.375486558581588,1			
15	65,1.48075700105323,-0.374088812752439,1			
16	67,1.42133809790472,-0.383897601770774,1			
17	68,1.22495437533303,-0.398715655625937,1			
18	87,0.880003710411558,-0.571789517264883,1			
19	88,0.630779828061623,-0.655476630185453,1			
20	89,0.855770322135293,-0.520592313755018,1			

Figura 111 Primeros registros de BDD3 aplicados LLE  
Fuente: Autoría propia

	A	B	C	D
35	179,-1.57404948798132,-0.898617514173614,2			
36	181,-1.19186565659427,-0.754378781073817,2			
37	189,-1.37907941698006,-0.680647314117173,2			
38	190,-1.47793915307316,-0.691934488818958,2			
39	193,-1.47194186699035,-0.691493672650781,2			
40	195,-1.37631370664993,-0.684032380543647,2			
41	202,-0.315539194883516,0.914267327605599,2			
42	206,-0.916259630926908,1.17623845777103,2			
43	219,-1.16556454892001,-0.74115329544665,2			
44	220,-1.57404948798132,-0.898617514173613,2			
45	222,-1.19186565659427,-0.754378781073816,2			
46	230,-1.37907941698006,-0.680647314117173,2			
47	238,-1.48812503508626,-0.691499636732038,2			
48	240,-1.55366965725209,-0.710607352209001,2			
49	248,-0.916259630926906,1.17623845777103,2			
50	249,-0.916259630926905,1.17623845777103,2			
51	256,-0.916259630926906,1.17623845777103,2			
52	260,-1.16556454892001,-0.741153295446649,2			
53	261,-1.57404948798132,-0.898617514173614,2			
54	262,-1.12119133245022,-0.719938999580617,2			
55	264,-0.670593178476852,-0.786445264186178,2			
56	271,-0.562691571405941,-0.959660282554495,3			
57	277,0.519358108952164,-0.808063818719116,3			
58	294,0.135526293798353,1.12444900856803,3			
59	296,0.167994271631199,1.69062008635911,3			
60	297,0.167994271631199,1.69062008635911,3			

Figura 112 Últimos registros de BDD3 aplicados LLE  
Fuente: Autoría propia



- Base de datos 3 con SNE.

	A	B	C	D
1	,"Y1", "Y2", "Clase"			
2	4,0.197246123443165,-0.912039099903641,1			
3	5,0.0919800789497499,-0.769161258522565,1			
4	8,0.422741548825138,-0.940636609830749,1			
5	11,0.617102028582591,-1.10915146566393,1			
6	16,0.554894177381147,-0.954921728147436,1			
7	20,0.780743530920805,-0.641217372182782,1			
8	21,0.807919395843,-0.969923384305894,1			
9	24,0.631230416054395,-0.710097767017099,1			
10	31,0.368659307952656,-0.524925770764839,1			
11	32,-0.0459421486798968,-0.503125526219087,1			
12	50,0.619179400453328,-1.00931858075218,1			
13	53,0.454780898026061,-0.935331878880979,1			
14	59,0.378557015114816,-0.623664441846845,1			
15	65,0.857314226345086,-0.846986465717444,1			
16	67,0.847986319695208,-1.13905011480065,1			
17	68,0.701006505356361,-0.435153743506857,1			
18	87,0.26377429398627,-0.732213003687025,1			
19	88,0.718892334260103,-0.843747255245465,1			
20	89,0.281183115626414,-0.856602328922042,1			

Figura 113 Primeros registros de BDD3 con SNE

Fuente: Autoría propia

	A	B	C	D
57	277,-0.667257787743273,-0.0188593374757882,3			
58	294,-1.01879167067827,0.811742898259396,3			
59	296,-0.950889410410816,1.30822644613592,3			
60	297,-0.237345858194703,1.45492434987539,3			
61	300,-1.07035061882392,1.54760841334992,3			
62	301,-0.666461407194012,1.00090359087049,3			
63	316,-1.18265748885509,-0.171014793638949,3			
64	317,-1.10873436880332,0.364782376703888,3			
65	320,-0.915947252974206,0.178850981281287,3			
66	321,-0.971793362231425,-0.0154007849013295,3			
67	327,-0.686215958915376,0.00797508469216998,3			
68	330,-0.729663154494231,0.0986559042232271,3			
69	334,-0.69108818215916,0.274636529576042,3			
70	340,-0.810292389092084,1.65150431434865,3			
71	347,-0.730221739279017,1.15418054959591,3			
72	352,-0.646363756449584,1.3652677936329,3			
73	356,-0.710937991772959,1.07325010503535,3			
74	360,-1.26445286962989,-0.159198425768943,3			
75	363,-0.769573844508431,-0.30971317742367,3			
76	373,-1.00115783273142,0.104927427347628,3			
77	376,-0.982028265983153,-0.0488226603292438,3			
78	380,-0.647935483979159,0.188906257502272,3			
79	386,-0.694289078395756,1.21809630952658,3			
80	391,-0.83794727926015,1.23391109568507,3			
81	394,-0.727903347070657,1.4122773938861,3			
82	400,-0.552288174229704,1.40917503872985,3			

Figura 114 Últimos registros de BDD3 con SNE

Fuente: Autoría propia

- Base de datos 3 con *t*SNE.

	A	B	C	D
1	,"V1","V2","cl_kmeans"			
2	1,-3.96430178835739,-5.88153574759806,"1"			
3	2,-4.01593461573114,-6.09871744022412,"1"			
4	3,-3.90852561992835,-5.58604884026446,"1"			
5	4,-4.01635517910299,-5.88510778103313,"1"			
6	5,-4.0082510624358,-6.13731008249994,"1"			
7	6,-4.07302580995835,-5.43045501529663,"1"			
8	7,-3.34069215164172,-6.40628351417933,"1"			
9	8,-4.00192333714757,-5.3383310822999,"1"			
10	9,-4.07538014333229,-5.78825673447271,"1"			
11	10,-4.11761576313209,-5.82665812050084,"1"			
12	11,-4.35548192034756,-6.26565360755169,"1"			
13	12,-4.02730662145014,-6.04056531119259,"1"			
14	13,-3.8363550548904,-5.3415855532615,"1"			
15	14,-3.20008231763429,-6.30299095432104,"1"			
16	15,-3.34158504741971,-6.40370562637997,"1"			
17	16,-3.34167631866912,-6.40547967455204,"1"			
18	17,-4.57878452101778,-5.78429174641664,"1"			
19	18,-4.57831630553123,-5.81560634897277,"1"			
20	19,-0.0184063720851601,-0.794388719186495,"2"			

Figura 115 Primeros registros de BDD3 con *t*SNE

Fuente: Autoría propia

	A	B	C	D
38	37,-0.289657023120781,-0.315140965650131,"2"			
39	38,-0.135104770592351,-0.454236332609698,"2"			
40	39,1.308882021813,1.94018887728305,"2"			
41	40,1.31832793492499,1.95481870899986,"2"			
42	41,1.30863686615837,1.95970722945175,"2"			
43	42,1.28574170752729,1.90450840518892,"2"			
44	43,3.91216547668758,5.73631397172297,"3"			
45	44,4.01588407644213,5.5878744321465,"3"			
46	45,3.91911737114438,5.9290597367974,"3"			
47	46,3.88102250003968,5.60780786185664,"3"			
48	47,3.35840041870705,5.99859011214085,"3"			
49	48,3.36809632027385,6.02602704150301,"3"			
50	49,3.64621639761545,6.02501355037532,"3"			
51	50,3.6124680544249,6.06617493139349,"3"			
52	51,3.04305909285862,5.39465711716797,"3"			
53	52,3.04326357982423,5.48444356858996,"3"			
54	53,3.08265608824022,5.30329324872419,"3"			
55	54,3.07558480176812,5.37737923975919,"3"			
56	55,3.31620456957425,5.98943728754282,"3"			
57	56,3.74857174246984,6.14319361898455,"3"			
58	57,3.93875414492504,5.87717601265426,"3"			
59	58,3.7744294652109,5.27673147220109,"3"			
60	59,3.57470223100191,5.09564582619497,"3"			
61	60,3.38604846736501,5.13246664354784,"3"			
62	61,3.63443309555905,5.29950683954349,"3"			
63	62,3.64782728912525,5.30347528477966,"3"			

Figura 116 Últimos registros de BDD3 con *t*SNE

Fuente: Autoría propia

- Base de datos 4 con PCA.

	A	B	C	D
1	,"pc1", "pc2", "clase"			
2	1,-1.86599932110646,1.20327887594498,1			
3	2,-1.85735084934593,1.20649706688832,1			
4	3,-1.86953475878533,1.1502615397013,1			
5	4,-1.88372424817102,1.14635053508278,1			
6	5,-1.87050815485068,1.13017931626514,1			
7	6,-1.79782099363141,1.11169501744643,1			
8	7,-1.84410233653207,1.10369032955344,1			
9	8,-1.87118029326846,1.09593720914678,1			
10	9,-1.86968267879585,1.0566040809648,1			
11	10,-1.70125970197665,1.04251923480995,1			
12	11,-1.8326114356901,0.960454054650731,1			
13	12,-1.86421485524653,0.927430596981468,1			
14	13,-1.82091408419216,0.944056158690429,1			
15	14,-0.504912292222443,-0.756864879316489,1			
16	15,-0.425125460777544,-0.796181418811076,1			
17	16,-0.266448875761474,-0.743333043708886,1			
18	17,-0.210627534510243,-0.800518975186853,1			
19	18,-0.228743564121529,-0.834213184262461,1			
20	19,-0.325718971831558,-0.930147434883209,1			

Figura 117 Primeros registros de BDD4 aplicados PCA  
Fuente: Autoría propia

	A	B	C	D
96	95,-1.71761410280878,-0.37923844816697,1			
97	96,-1.72459634841132,-0.375616805451179,1			
98	97,-1.71522317728882,-0.401474315455402,1			
99	98,-1.73813622257218,-0.393590880374336,1			
100	99,-1.76767904110874,-0.353701475490747,1			
101	100,-1.79486282672046,-0.355538706935437,1			
102	101,-1.79909112516597,-0.355530994916178,1			
103	102,-1.83937097695535,-0.387543792925911,1			
104	103,-1.82183717488094,-0.395876542201433,1			
105	104,-1.81446002015845,-0.379044664362246,1			
106	105,-1.81348397445021,-0.41198007280928,1			
107	106,-1.83000484372289,-0.396046564270241,1			
108	107,-1.85292391686336,-0.407670499299764,1			
109	108,-1.83748444479279,-0.390600980396302,1			
110	109,-1.81059982109423,-0.378816114801538,1			
111	110,-1.8105972313714,-0.356606069631795,1			
112	111,-1.78573110675469,-0.191224037490387,1			
113	112,-1.80366867154874,-0.32333179407937,1			
114	113,-1.80155886831507,-0.389303837014556,1			
115	114,-1.77919342614468,-0.361170087728411,1			
116	115,-1.77467424461652,-0.377518971419792,1			
117	116,-1.78160280443918,-0.410793246972217,1			
118	117,-1.77480540092834,-0.452662428406522,1			
119	118,-1.74055224008723,-0.465763100252902,1			
120	119,-1.73204391756216,-0.504177902071608,1			
121	120,-1.74064095451622,-0.531178178694617,1			

Figura 118 Últimos registros de BDD4 aplicados PCA  
Fuente: Autoría propia

- Base de datos 4 con LE.

	A	B	C	D
1	,"Y1", "Y2", "clase"			
2	4,0.142017049795992,0.00973749734804664,1			
3	5,0.142017049795992,0.00973749734804664,1			
4	8,0.142017049795992,0.00973749734804664,1			
5	11,0.133848938004828,0.00173327465460191,1			
6	16,0.119414659009271,-0.0123754978286075,1			
7	20,0.138509226869005,0.000212745581712641,1			
8	21,0.138509226869005,0.000212745581712641,1			
9	24,0.138509226869005,0.000212745581712643,1			
10	31,0.138509226869005,0.000212745581712643,1			
11	32,0.152333083700677,0.0098656618903357,1			
12	50,0.131787983927496,-0.0116227479791456,1			
13	53,0.161623242075474,0.00751652393622681,1			
14	59,0.166127800670668,0.0111838950166219,1			
15	65,0.166127800670668,0.0111838950166219,1			
16	67,0.166127800670668,0.0111838950166219,1			
17	68,0.150300425417332,0.00961408029865888,1			
18	87,0.166976323993449,0.0112658634454861,1			
19	88,0.166976323993449,0.0112658634454861,1			
20	89,0.166976323993449,0.0112658634454861,1			

Figura 119 Primeros registros de BDD4 aplicados LE  
Fuente: Autoría propia

	A	B	C	D
35	174,-0.120529490701353,-0.00135642269240255,3			
36	179,-0.12757266991868,-0.0893684997255418,3			
37	181,-0.125288628800011,-0.0686822401056001,3			
38	189,-0.126583903364698,-0.110501427928973,3			
39	190,-0.128826331954375,-0.0700059230025249,3			
40	193,-0.125086186105974,-0.109284337195926,3			
41	195,-0.125086186105974,-0.109284337195926,3			
42	202,-0.139881196857336,-0.0440752073615231,3			
43	206,-0.147455818461849,-0.0305740658619207,3			
44	219,-0.139068287357522,-0.0438200858344619,3			
45	220,-0.129845206464515,-0.13431953897742,3			
46	222,-0.134515175092186,-0.141980049039167,3			
47	223,-0.129570608546284,-0.133883076966335,3			
48	230,-0.132205822167902,-0.134514174593189,3			
49	238,-0.160664876664068,-0.0212278465351377,3			
50	240,-0.160664876664068,-0.0212278465351377,3			
51	248,-0.0851810915856167,0.235538228820611,4			
52	249,-0.080453309918967,0.256758093994286,4			
53	260,-0.0777401331396696,0.260873024481829,4			
54	261,-0.0777401331396696,0.260873024481829,4			
55	262,-0.0777401331396697,0.260873024481829,4			
56	264,-0.0777401331396699,0.260873024481829,4			
57	271,-0.0777401331396693,0.260873024481829,4			
58	277,-0.0777401331396696,0.260873024481828,4			
59	294,-0.0777278799222694,0.26372446092258,4			
60	296,-0.0836495067405617,0.259692196365788,4			

Figura 120 Últimos registros de BDD4 aplicados LE  
Fuente: Autoría propia

- Base de datos 4 con LLE.

	A	B	C	D
1	"X1","X2","clase"			
2	4,-0.401182538076522,-1.69052501674644,1			
3	5,-0.398001612188596,-1.68626384376654,1			
4	8,-0.397462136327896,-1.68915927402597,1			
5	11,-0.362938504716933,-1.63949220129868,1			
6	16,0.294077676682511,-0.408295453250137,1			
7	20,-0.564994517951425,1.22349617243082,1			
8	21,-1.04913585710955,1.26991237061262,1			
9	24,-0.94277235304631,1.28852900749643,1			
10	31,-1.18124578098076,1.25989497100447,1			
11	32,0.575822385758317,0.299432516215241,1			
12	50,0.139651132587585,-0.842626153558017,1			
13	53,0.0909407369470083,-0.931169129855803,1			
14	59,0.0481082085877821,-1.03101641610924,1			
15	65,0.0580165868464911,-1.0355139530414,1			
16	67,0.0332011592190137,-1.07983760145248,1			
17	68,0.0527482663675657,-1.04570485751311,1			
18	87,-0.0655975128557251,-1.28871315629792,1			
19	88,-0.0791383861472677,-1.31425207603544,1			
20	89,-0.0872465700156114,-1.32617138628385,1			

Figura 121 Primeros registros de BDD4 aplicados LLE

Fuente: Autoría propia

	A	B	C	D
28	132,1.11293287279448,0.795954786806002,2			
29	137,-1.14655061143225,-0.245274840070264,2			
30	139,1.11293287279447,0.795954786806001,2			
31	145,-0.420832658551054,-0.0576838698311408,2			
32	151,1.17183272987082,0.842825534232438,2			
33	167,-0.446019142394239,-0.517586891893785,3			
34	173,-0.271640244398486,-0.115285761953482,3			
35	174,-0.223102098220031,-0.0682976262959934,3			
36	179,-2.41384979744638,1.27935003757937,3			
37	181,-1.43526591313646,0.71730843326947,3			
38	189,-2.35409243684576,1.1903002470592,3			
39	190,-1.58318803435453,0.797115521868339,3			
40	193,-2.56168763597209,1.22082922784595,3			
41	195,-2.5343391849412,1.23206972777663,3			
42	202,1.16577697256173,0.951408236048398,3			
43	206,1.12584709990285,0.881255774899949,3			
44	219,0.992664619245912,0.899517018450804,3			
45	220,0.928454388020063,0.954726312380134,3			
46	222,1.03977903637283,0.911028639616893,3			
47	223,1.05453455004662,0.901462279039814,3			
48	230,0.802249328577993,0.863671194683992,3			
49	238,0.990750698828955,0.741495251873642,3			
50	240,0.966529843767656,0.718627705246444,3			
51	248,-1.13498077201171,0.142325248329747,4			
52	249,-1.02636938852808,0.10234751629799,4			
53	260,1.04300089327503,0.472042346322388,4			

Figura 122 Últimos registros de BDD4 aplicados LLE

Fuente: Autoría propia

- Base de datos 4 con SNE.

	A	B	C	D
1	,"Y1", "Y2", "clase"			
2	4,-0.424318027892999,0.76806874689951,1			
3	5,-0.212154745692543,1.38770038095256,1			
4	8,0.132346742360504,0.891106968530192,1			
5	11,0.322999934913355,1.0840844802747,1			
6	16,0.0239874307653528,0.622227797955705,1			
7	20,-0.583938261840879,0.553711126037382,1			
8	21,0.452831377079874,0.85533023713178,1			
9	24,0.188659859478371,0.556975316922042,1			
10	31,0.304409322699955,1.14344354635149,1			
11	32,0.353715668518703,0.796852386678515,1			
12	50,-0.34038197901311,0.854555795033687,1			
13	53,-0.145796480548913,0.96597195321945,1			
14	59,-0.0959704922203335,0.642074505171286,1			
15	65,-0.1634222294594,1.10303287573858,1			
16	67,0.0844752402591433,1.10639387363881,1			
17	68,-0.0390874505444137,0.83476901421208,1			
18	87,-0.431141439715813,0.498610005221014,1			
19	88,-0.438499099827182,1.24497595932356,1			
20	89,-0.0843733813902643,1.18590692809317,1			

Figura 123 Primeros registros de BDD4 con SNE

Fuente: Autoría propia

	A	B	C	D
37	181,0.162393084881656,-0.237888880554564,3			
38	189,0.388958385433379,-0.673101819055234,3			
39	190,-0.279589803170806,-0.559065792817911,3			
40	193,0.306196416403831,-0.609326321254001,3			
41	195,0.504456358899298,-0.311785576480937,3			
42	202,0.231897831679196,-0.7367432251087,3			
43	206,0.146688454192127,-0.208344389910444,3			
44	219,-0.12589227026946,-0.504145581199899,3			
45	220,0.012338445011054,-0.474760281178581,3			
46	222,0.689799774019084,-0.466560125001075,3			
47	223,0.00477544505501348,-0.44799660933124,3			
48	230,0.391017543548286,-0.43864882337786,3			
49	238,-0.119363587635232,-0.435291751555907,3			
50	240,0.222670223441684,-0.551197120340706,3			
51	248,0.544137756840041,-0.68168799421797,4			
52	249,0.04750243457085,-1.06493217980045,4			
53	260,-0.255720512031058,-1.21895285263488,4			
54	261,-0.138385849215604,-1.33093508331669,4			
55	262,-0.077641217043576,-1.16971964970152,4			
56	264,-0.0636211803780302,-1.16081323658165,4			
57	271,0.311925694439491,-1.07984405034999,4			
58	277,0.0258100249090404,-1.37599086160491,4			
59	294,0.312337981237059,-1.16184597375528,4			
60	296,-0.166538643278496,-1.48982674815468,4			
61	297,0.170701174133359,-1.16449872832751,4			
62	300,0.130603154489522,-1.35436998087052,4			

Figura 124 Últimos registros de BDD4 con SNE

Fuente: Autoría propia

- Base de datos 4 con *t*SNE.

	A	B	C	D
1	"V1", "V2", "cl_kmeans"			
2	1,-20.1490061467382,-4.9949005624967,"3"			
3	2,-20.1597346180711,-5.49639439990928,"3"			
4	3,-20.145871377373,-5.48713423134368,"3"			
5	4,-20.035586201628,-4.94840034967084,"3"			
6	5,-18.0023489951407,-5.71597408499884,"3"			
7	6,-16.5249231681694,-5.26420423947774,"3"			
8	7,-16.5028652062354,-5.25950579826625,"3"			
9	8,-16.4993954022645,-5.25820922825775,"3"			
10	9,-16.5012729524406,-5.26168889280186,"3"			
11	10,-17.5115925765934,-5.72405074844634,"3"			
12	11,-18.3531901468991,-4.9790286597715,"3"			
13	12,-18.4526229659365,-5.39028327667091,"3"			
14	13,-18.6487095329444,-5.02927328746656,"3"			
15	14,-18.5843712060484,-5.16131169369919,"3"			
16	15,-18.7524940184107,-5.83398405644191,"3"			
17	16,-18.6706078690462,-5.90296820016821,"3"			
18	17,-19.0424701675408,-4.88911261273166,"3"			
19	18,-19.1549342106608,-5.90489458520498,"3"			
20	19,-19.079638468796,-4.84254623403541,"3"			

Figura 125 Primeros registros de BDD4 aplicados *t*SNE  
Fuente: Autoría propia

	A	B	C	D
37	36,12.3105021640197,5.30611528950773,"1"			
38	37,11.8579317514358,5.80766737024581,"1"			
39	38,12.1514800950371,5.33348562597652,"1"			
40	39,11.8629301530458,5.8772902212899,"1"			
41	40,11.8608886877493,5.87656952686874,"1"			
42	41,12.4230105538706,3.82243022967235,"4"			
43	42,12.7772881007669,3.88330479223358,"4"			
44	43,12.3232363748432,3.91972508714492,"4"			
45	44,12.2644627914051,4.13313204841261,"4"			
46	45,12.4911606627618,4.1662795019493,"4"			
47	46,12.6151270125835,4.16897246172124,"4"			
48	47,12.7866185271211,4.39294152293171,"4"			
49	48,12.8232227423184,3.89581923407425,"4"			
50	49,12.8049258915349,3.87615569171063,"4"			
51	50,15.1439856720155,2.40850781925433,"2"			
52	51,15.160764295876,2.44246344700953,"2"			
53	52,15.7071888943884,3.69680357549505,"4"			
54	53,15.420461189143,3.71191977645621,"4"			
55	54,15.5140293058085,3.73521769491084,"4"			
56	55,15.8392915032109,3.4130777654101,"4"			
57	56,15.2656862801327,3.53232421882037,"4"			
58	57,15.2091309266019,3.39239410478582,"4"			
59	58,15.2354058487443,3.22113154943315,"4"			
60	59,15.7290971053903,3.13636260604924,"4"			
61	60,15.5419848673401,3.08050366916688,"4"			
62	61,15.7809368485354,3.23663530190165,"4"			

Figura 126 Últimos registros de BDD4 aplicados *t*SNE  
Fuente: Autoría propia

- Base de datos 5 con PCA

	A	B	C	D
1	,"pc1","pc2","LABEL"			
2	2,0.0321390036500863,0.200992449705993,1			
3	4,0.0817080468910824,0.22676127195285,1			
4	5,0.0755402080006815,0.226021269892985,1			
5	8,0.034786807031037,0.162941044171288,1			
6	11,0.124088930056506,0.0711554167488367,1			
7	16,0.168256866536146,0.284877909742188,1			
8	20,0.125335792232671,0.266203203618629,1			
9	21,0.0769826226050362,0.254774154698237,1			
10	24,-0.0598502572018926,0.184889587463833,1			
11	31,0.00701738690472128,0.225248578482005,1			
12	32,0.0285697976076445,0.233362532326026,1			
13	34,0.0653366398901979,0.174717073187736,1			
14	50,0.0651429010309009,0.2594646532887,1			
15	53,0.0198446255114329,0.205358504835116,1			
16	59,0.237645974908769,0.367501722310119,1			
17	65,0.122715470730287,0.295442967025638,1			
18	67,0.0630060013239741,0.199993204116272,1			
19	68,0.0957608727241396,0.260050326949003,1			
20	69,0.297610943343187,0.358361391023606,1			

Figura 127 Primeros registros de BDD5 aplicados PCA

Fuente: Autoría propia

	A	B	C	D
47	222,0.301806544816032,0.0343168123539091,2			
48	230,0.00647883961993869,-0.0932382050346813,2			
49	238,0.351585756604927,0.0514738101385479,2			
50	240,0.250506423289084,-0.0474668426737171,2			
51	248,0.194956584050196,0.0218593256765375,2			
52	249,0.147129996670235,0.164617442418021,2			
53	260,0.140430996389654,-0.00836750248313262,2			
54	261,0.188240210159887,-0.0217195604992071,2			
55	262,0.20398069174977,-0.0296420586368257,2			
56	264,0.16671128965505,-0.00491542200309554,2			
57	271,0.109230571907161,-0.104743396568131,2			
58	277,0.146431876346281,-0.0897007134927208,2			
59	294,0.28272795922928,-0.0141251324426619,2			
60	296,0.265008603521459,-0.000620263711246936,2			
61	297,0.353372590941445,-0.057129941713342,2			
62	300,0.279118411511946,-0.265284626958572,2			
63	316,0.0419412771390526,-0.748901492167294,2			
64	317,0.56014574105288,-0.279625308863081,2			
65	320,0.401390206831696,-0.427803476868719,2			
66	321,0.357391620312417,-0.323723745850656,2			
67	327,0.385789677288931,-0.286756692414508,2			
68	330,0.317748255546333,-0.289257580152198,2			
69	334,0.279097097404548,-0.339058690670636,2			
70	340,-0.315354850293606,-0.332542895061625,2			
71	347,-0.292991005621938,-0.312080014145142,2			
72	352,-0.0399288434625826,-0.164524268908183,2			

Figura 128 Últimos registros de BDD5 aplicados PCA

Fuente: Autoría propia



- Base de datos 5 con LE.

	A	B
1	"Y1", "Y2", "LABEL"	
2	2,0,-0.12346895286559,1	
3	4,0,-0.155053849512944,1	
4	5,0,-0.155053849512944,1	
5	8,0,-0.114424009033346,1	
6	11,0,-0.144462739584746,1	
7	16,0,-0.157576031236802,1	
8	20,0,-0.157576031236802,1	
9	21,0,-0.143448312958884,1	
10	24,0,-0.0793544204831027,1	
11	31,0,-0.12775810516919,1	
12	32,0,-0.12775810516919,1	
13	34,0,-0.136856764065325,1	
14	50,0,-0.15129861960411,1	
15	53,0,-0.130895903792685,1	
16	59,0,-0.151294320716454,1	
17	65,0,-0.151294320716454,1	
18	67,0,-0.145971940238913,1	
19	68,0,-0.144070604942597,1	
20	69,0,-0.150544459631582,1	

Figura 129 Primeros registros de BDD5 aplicados LE

Fuente: Autoría propia

	A	B
52	249,0.168184702874085,0,2	
53	260,0.185268625795367,0,2	
54	261,0.184289770402881,0,2	
55	262,0.185439578212265,0,2	
56	264,0.185439578212265,0,2	
57	271,0.183946051224085,0,2	
58	277,0.187171258927432,0,2	
59	294,0.16329832339946,0,2	
60	296,0.176034969627485,0,2	
61	297,0.166732476674592,0,2	
62	300,0.185259772004398,0,2	
63	316,0.150511269381271,0,2	
64	317,0.150790239750958,0,2	
65	320,0.147629196591815,0,2	
66	321,0.152342315647872,0,2	
67	327,0.153794600625838,0,2	
68	330,0.179308731178477,0,2	
69	334,0.166918236649816,0,2	
70	340,0.140002160088807,0,2	
71	347,0.139069393295603,0,2	
72	352,0.149027688657505,0,2	
73	356,0.139237070853406,0,2	
74	360,0.138474987696614,0,2	
75	363,0.138474987696614,0,2	
76	373,0.223311719163853,0,2	
77	376,0.141408845855392,0,2	

Figura 130 Últimos registros de BDD5 aplicados LE

Fuente: Autoría propia

- Base de datos 5 con LLE.

	A	B	C	D
1	"Y1", "Y2", "LABEL"			
2	2,-1.19455295440499,0.160396476665739,1			
3	4,-1.39739075181847,0.385659516649545,1			
4	5,-1.3547528854334,0.355327423462603,1			
5	8,-0.88261959271834,0.160304159402922,1			
6	11,-0.408024642233072,0.533242704712842,1			
7	16,-1.62659023986472,0.7008039140766,1			
8	20,-1.49139569901676,0.532666853676604,1			
9	21,-1.04826523050177,0.279520700552896,1			
10	24,-0.991505551097191,-0.241474333803987,1			
11	31,-1.32131214556291,0.0657929203948789,1			
12	32,-1.43667938296311,0.185244471325301,1			
13	34,-1.10781682023451,0.306165459214972,1			
14	50,-1.53432872797828,0.33049580753057,1			
15	53,-1.12022311758152,0.0978023541563792,1			
16	59,-1.76539064496633,0.880803234317387,1			
17	65,-1.64082414361372,0.530774657363461,1			
18	67,-1.22516241166898,0.301973344257868,1			
19	68,-1.46386282573865,0.450068421034519,1			
20	69,-1.60564810280983,1.0317134667484,1			

Figura 131 Primeros registros de BDD5 aplicados LLE  
Fuente: Autoría propia

	A	B	C	D
53	260,0.106368779234462,0.614141053022513,2			
54	261,0.283007162838289,0.840640500953342,2			
55	262,0.316079482521295,0.889847624766474,2			
56	264,0.127049994689296,0.728763669210087,2			
57	271,0.710435275697186,0.580956615507725,2			
58	277,0.364514503631027,0.623782904423955,2			
59	294,0.0889044428415178,0.995406908978653,2			
60	296,0.0413537689069344,1.01140691577891,2			
61	297,0.587144632270551,1.38813585035116,2			
62	300,1.44476093320383,1.21910850685868,2			
63	316,1.24639048782083,0.255717239455937,2			
64	317,1.19655842296304,1.47123345915588,2			
65	320,2.13556190004995,1.52869885541014,2			
66	321,1.79892010135245,1.46497786831243,2			
67	327,1.65586603140751,1.52406602105071,2			
68	330,1.53059426136472,1.32994586610953,2			
69	334,1.80292152831562,1.22728620208105,2			
70	340,1.35882689915504,-0.986443002707026,2			
71	347,1.42348105537985,-0.929610401605858,2			
72	352,0.774770965463288,-0.0373874368784496,2			
73	356,1.62581139107748,-0.608218466972877,2			
74	360,1.53018140837258,-0.692921517252413,2			
75	363,1.07937188070539,-0.392055824267083,2			
76	373,0.94600904509449,0.616636769142358,2			
77	376,1.43280597125814,-0.0762215119574117,2			
78	380,0.281554586079995,-1.09719367886838,2			

Figura 132 Últimos registros de BDD5 aplicados LLE  
Fuente: Autoría propia

- Base de datos 5 con SNE.

	A	B	C	D
1	,"Y1","Y2","LABEL"			
2	2,-0.122094959091504,0.0154035975958175,1			
3	4,-0.226188461670893,0.101619622735875,1			
4	5,0.789532629510611,0.310632539231882,1			
5	8,0.0220661015546138,-0.317319358587583,1			
6	11,-0.569608787606801,0.225324041649305,1			
7	16,0.127197877526283,-0.124565598690639,1			
8	20,-0.374681213515455,-0.278457363845158,1			
9	21,0.0245591733029974,-0.398067314700869,1			
10	24,-0.375053589012855,-0.115830657494421,1			
11	31,-0.326247737133112,-0.157646639211521,1			
12	32,-0.333407785337195,0.117039490966373,1			
13	34,0.303121588049607,-0.210205679840299,1			
14	50,-0.123657956570219,-0.335738454724753,1			
15	53,-0.489935423828531,0.288841882535739,1			
16	59,-0.254502003593134,0.158805450164503,1			
17	65,-0.133060026366703,-0.152514843437939,1			
18	67,-0.515047109357393,-0.435580909167307,1			
19	68,0.260493320680584,-0.329541457009808,1			
20	69,0.650057817251963,0.267808358754057,1			

Figura 133 Primeros registros de BDD5 aplicados SNE  
Fuente: Autoría propia

	A	B	C	D
33	139,0.217126214042023,0.0330567834544804,1			
34	145,0.127932326149478,-0.0433800201573922,1			
35	151,-0.0910702320124612,-0.352269387093528,1			
36	173,-0.0193573684922986,0.0541938936405218,1			
37	179,-0.584164561928107,0.289704456552748,1			
38	181,-0.05515350631809,-0.022155498013351,1			
39	189,0.018408169744176,0.379485188204443,1			
40	190,-0.176241919824468,0.615612335984919,1			
41	193,-0.126468360634076,-0.528550895037206,1			
42	195,-0.167141933717122,-0.15456526561268,1			
43	202,0.260578835089435,0.220020731262631,1			
44	206,0.0345429015156257,-0.294876593504549,1			
45	219,-0.0687890936369427,0.0763131157719156,2			
46	220,-0.144068539412073,-0.0623787343696681,2			
47	222,0.251953048639801,0.470643834990382,2			
48	230,0.070858286025196,0.094528952040034,2			
49	238,0.137097621906407,0.611165717436051,2			
50	240,0.392194881549387,0.142279002166541,2			
51	248,-0.504937005698478,0.151743665487426,2			
52	249,0.0537004510584093,0.393204738949293,2			
53	260,0.263246839819785,0.0305398701574037,2			
54	261,0.336480766844435,0.289255462455605,2			
55	262,0.184775362890709,0.359869360115194,2			
56	264,-0.336363558915314,-0.114930341638781,2			
57	271,0.161821565538861,0.272263920441203,2			
58	277,0.357374718981291,-0.429325484471746,2			

Figura 134 Últimos registros de BDD5 aplicados SNE  
Fuente: Autoría propia

- Base de datos 5 con tSNE.

	A	B	C	D
1	,"V1","V2","cl_kmeans"			
2	1,14.9574359932923,5.4804685391117,"1"			
3	2,14.2163176512019,5.7615793243162,"1"			
4	3,14.4552418411984,5.43302994127509,"1"			
5	4,15.2435381119455,5.13409831589279,"1"			
6	5,15.250521531484,6.18974812472013,"1"			
7	6,13.7028176869899,5.23942856130076,"1"			
8	7,14.1639682306997,4.86267026725571,"1"			
9	8,14.7314866388606,3.96983015638061,"1"			
10	9,15.8380852104443,5.35910517672069,"1"			
11	10,14.9577224500227,5.60425902853059,"1"			
12	11,14.3945227611785,6.07976475168498,"1"			
13	12,14.6988615381327,6.05404400386012,"1"			
14	13,14.2014810073243,5.5740686921343,"1"			
15	14,15.1267529493397,5.0383272147578,"1"			
16	15,13.4286045109722,4.57702177635848,"1"			
17	16,14.0514320464064,4.8462175008006,"1"			
18	17,14.617982514683,5.77795173503916,"1"			
19	18,13.7058864035909,6.50038742975016,"1"			
20	19,13.2441782861977,4.65675047669174,"1"			

Figura 135 Primeros registros de BDD5 aplicados tSNE  
Fuente: Autoría propia

	A	B	C	D
52	51,-22.5845829138506,-6.29297514238046,"2"			
53	52,-21.6655937920097,-6.74007818595017,"2"			
54	53,-21.76200691294,-6.76661377203938,"2"			
55	54,-21.8219598322518,-7.01140941142655,"2"			
56	55,-21.8172906650088,-6.78288235195228,"2"			
57	56,-20.9823817031447,-6.77744840354628,"2"			
58	57,-21.4141411767207,-7.68111099993023,"2"			
59	58,-22.6637320587214,-8.06924297901142,"2"			
60	59,-22.5661906269289,-7.77004649826316,"2"			
61	60,-22.1959189585923,-7.87594759423006,"2"			
62	61,-20.9736009288217,-8.71576594754187,"2"			
63	62,-19.256108103235,-8.7363637548637,"2"			
64	63,-20.0648651549925,-9.37667278645336,"2"			
65	64,-21.1841144661645,-9.54289402939444,"2"			
66	65,-21.1687461936707,-9.24056033001385,"2"			
67	66,-21.2851736296119,-9.22900856424273,"2"			
68	67,-20.9171363961055,-8.90926665231902,"2"			
69	68,-20.927601801215,-8.99135080250752,"2"			
70	69,-18.7961513824454,-6.67173910344862,"2"			
71	70,-18.8401636517416,-6.49617839430818,"2"			
72	71,-20.0253460178506,-6.57738203214286,"2"			
73	72,-18.8159088988569,-6.57801671928608,"2"			
74	73,-18.8394522894561,-6.47084522040487,"2"			
75	74,-19.4595113274035,-6.44095140776408,"2"			
76	75,-20.8084327191011,-7.02947644415276,"2"			
77	76,-19.4911455559397,-6.71898999756147,"2"			

Figura 136 Últimos registros de BDD5 aplicados tSNE  
Fuente: Autoría propia