



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES

CARRERA DE TELECOMUNICACIONES

“Creación de algoritmos inteligentes basados en la teoría de machine learning tradicional para la clasificación de los eventos sísmicos en el volcán Llaima (multiclase)”

Autores: Cachipueno Yacelga , César Ariel y Tutillo Moyón , Javier Alejandro

Director del Proyecto: Ing. Román Alcides Lara Cueva, PhD.

06 de septiembre de 2023

VERSIÓN: 1.1

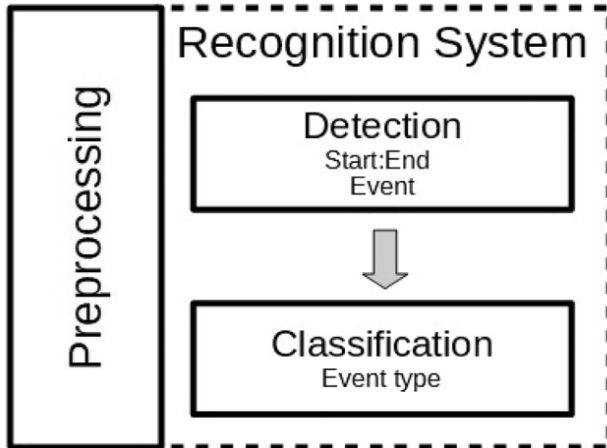


Agenda

- 1.- **Introducción**
- 2.- **Materiales y Métodos**
- 3.- **Pruebas y Resultados**
- 4.- **Conclusiones y Trabajos futuros**

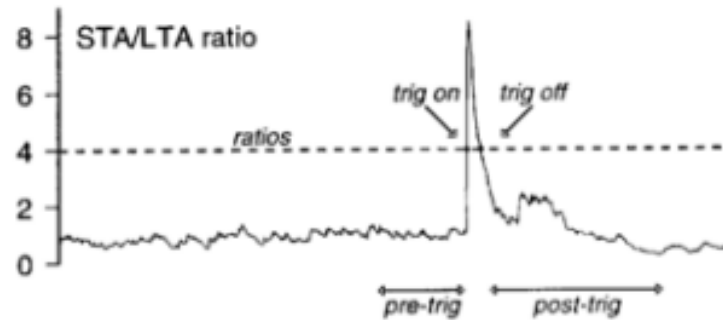
1. Introducción

Antecedentes



STA/LTA

STA: *Short Term Average*
LTA: *Long Term Average*



Bayes

Exactitud :
Clasificador → 84%

Markov

Exactitud :
Clasificador → 90%

Deep Learning

Exactitud :
Clasificador → 99.6%
BER: 0.0027

CNN

Exactitud:
Detector → 99%
Clasificador → 97%

1. Introducción

Motivación e Importancia

Visual



Sistema de Alerta Temprana



Llaima: Estrato volcán activo /Región de La Araucanía (Chile).



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

1. Introducción

Objetivos y Actividades

- Objetivo

Crear algoritmos inteligentes basados en la teoría de *Machine Learning* tradicional para la clasificación de los eventos sísmicos en el volcán Llaima (multiclase).

- Actividades

Actividad 1. - Identificación de las propiedades y variables del problema a tener en cuenta.

Actividad 2. - Reunión de la base de datos representativa, con datos provistos por el Observatorio Volcanológico de los Andes del Sur.

Actividad 3. - Creación de una estructura de datos (temporales, espaciales, otros.) con un soporte común.

Actividad 4. - Pruebas con los algoritmos basados en la teoría de Machine Learning especializados.

Actividad 5. - Pruebas y evaluación del desempeño.



2. Materiales y Métodos

Equipos y Software



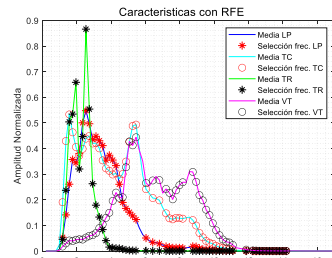
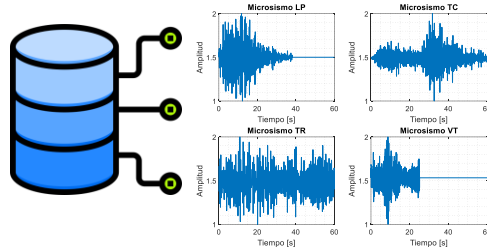
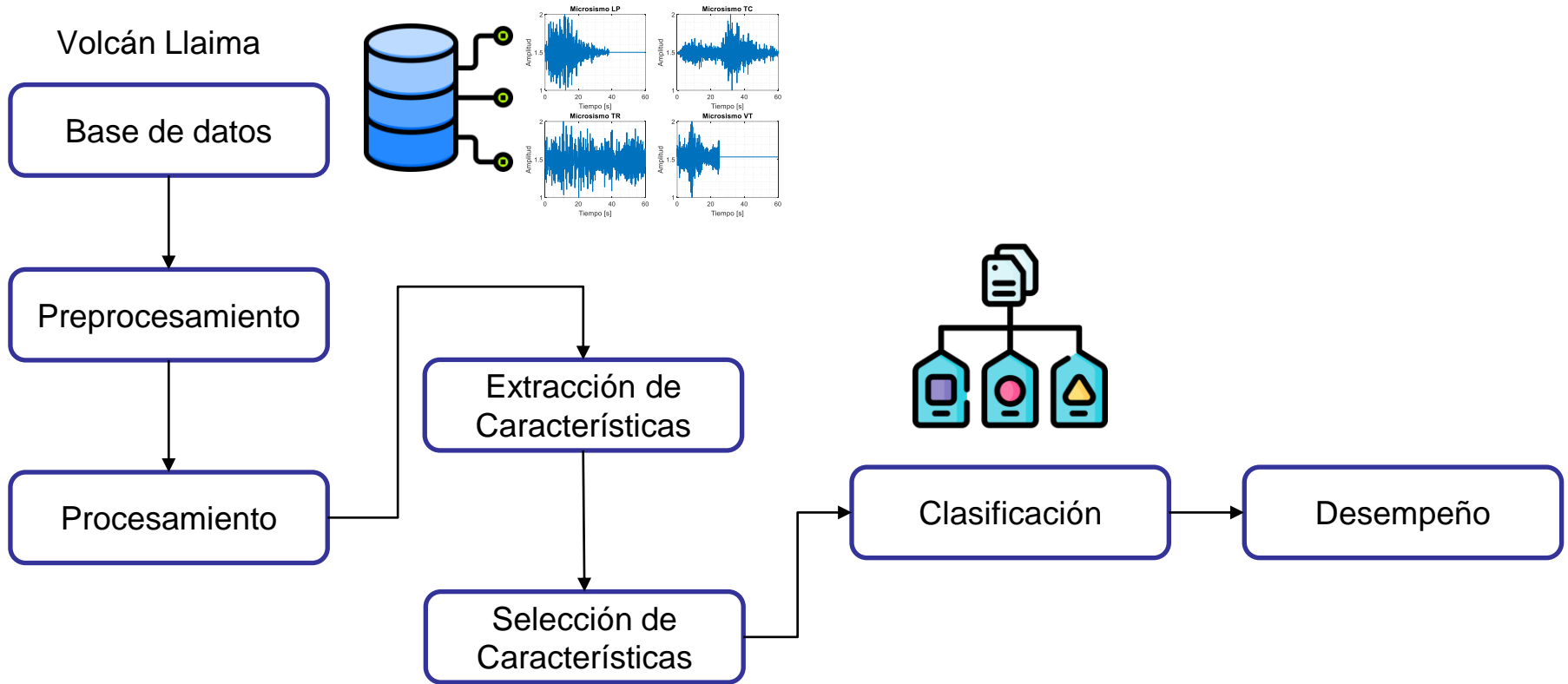
- Procesador:
- Intel® Core i5 @ 2.30 GHz
- Memoria RAM 8 Gb
- Tipo de sistema 64 bits



- Matlab 2022b

2. Materiales y Métodos

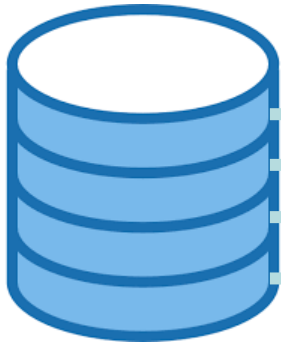
Proceso realizado para el sistema de clasificación



2. Materiales y Métodos

Base de datos (BD)

1	2	3	4	5	6	7	8	9	10	11
Network	Station	SampleRate	Component	Year	Month	Type	Duration	StartPoint	EndPoint	Data
"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	"LP"	38.1100	1	3812	1x6000 double
"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	"LP"	39.6700	1	3968	1x6000 double
"Chile"	"LAV"	100	"SHZ"	"2010-2016"	1	"LP"	47.5300	1	4754	1x6000 double



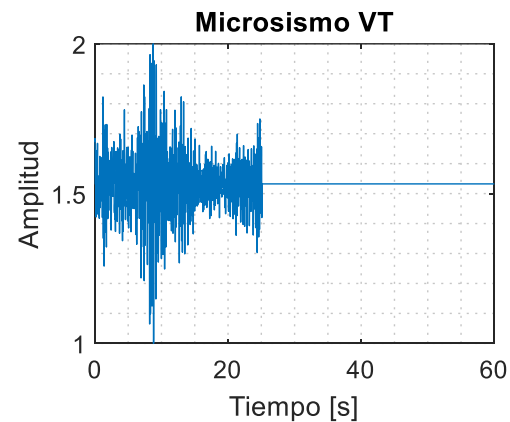
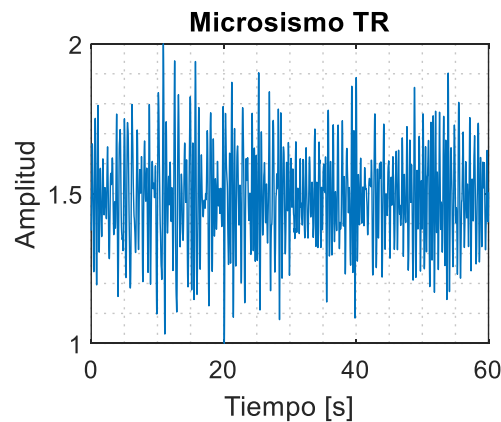
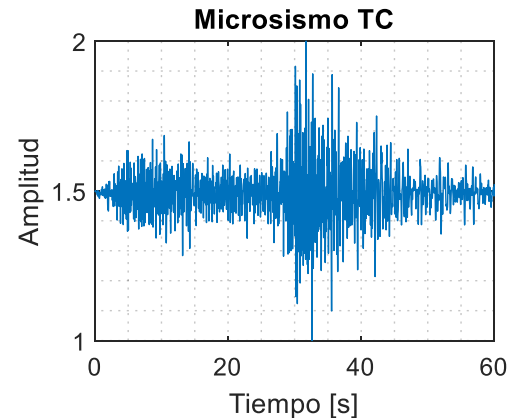
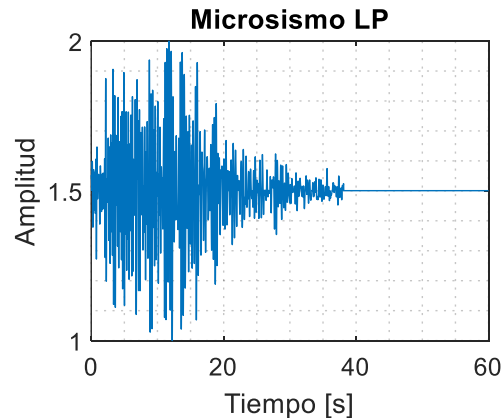
- LP con 1310 señales
- TC con 1488 señales
- TR con 490 señales
- VT con 304 señales

Total: 3592 señales

OVDAS
Estación LAV
2010 - 2016

Características:

- SimpleRate: 100 Hz
- Data: 1x6000 double



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

2. Materiales y Métodos

Preprocesamiento

Filtro

- Mejorar calidad de señales
- Separar componentes
- Eliminar ruido



Filtro digital bidireccional de respuesta al infinito (FIR)

Recorte

- Seleccionar solo señal de interés de las 6 000 muestras.



Puntos de inicio y fin de la BD.

Procesamiento

Extracción características

- Extracción de información significativa y relevante.



- PSD Welch
- 84 Características comprendidas entre tiempo, frecuencia y escala.

2. Materiales y Métodos

Procesamiento

Normalización

- Transformar datos complejos en un conjunto de datos más simples, estables y compactos. Para evitar redundancias y anomalías.



Rango de 0 a 1

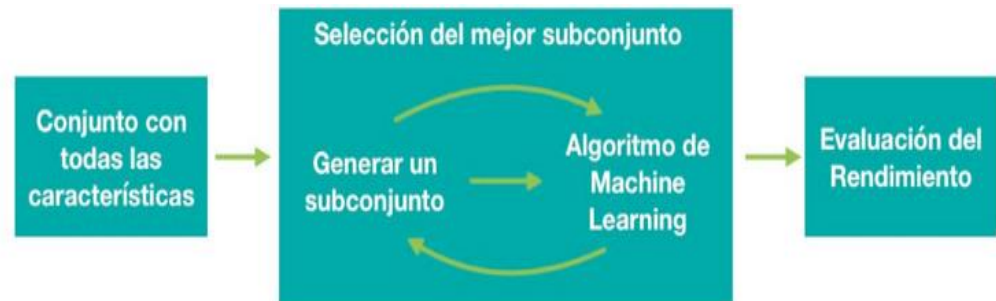
$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Selección de Características (RFE)

- Seleccionar las características más relevantes de un conjunto de datos grande



Método de envoltura hacia adelante.

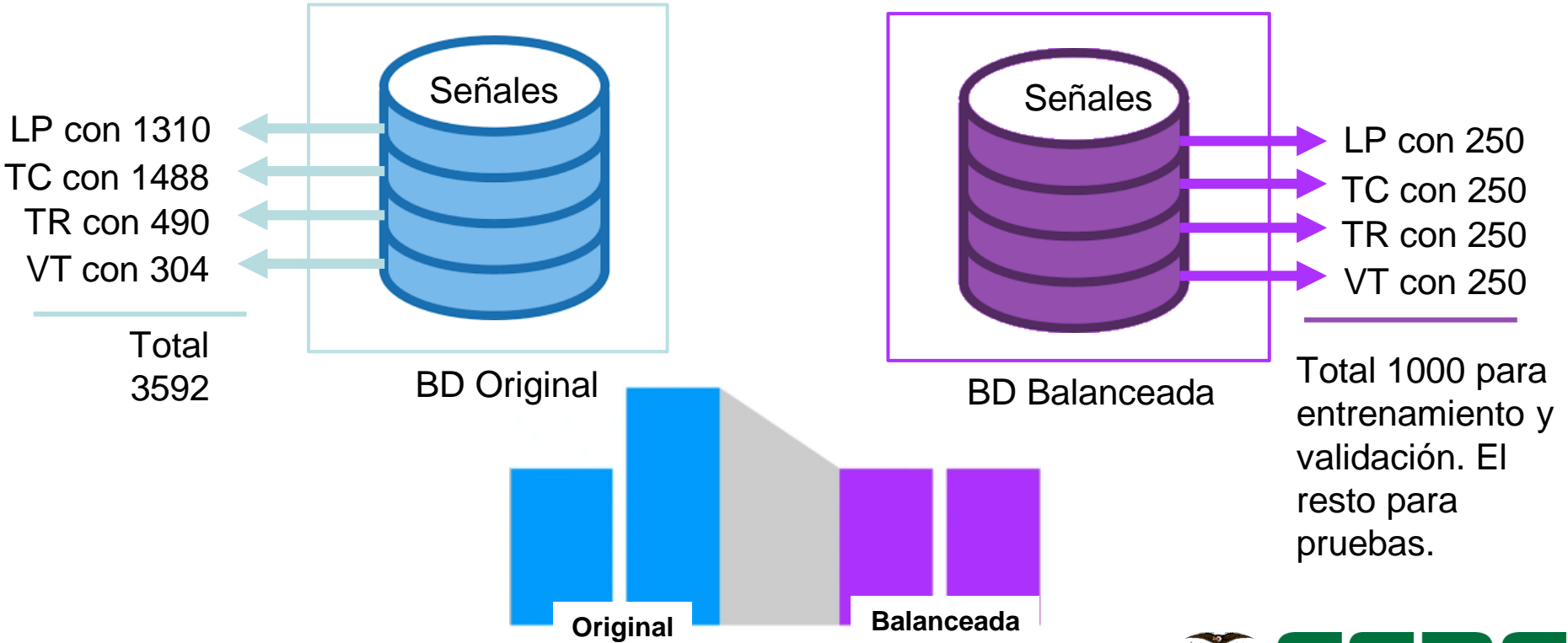


2. Materiales y Métodos

Procesamiento

Balanceo de Datos

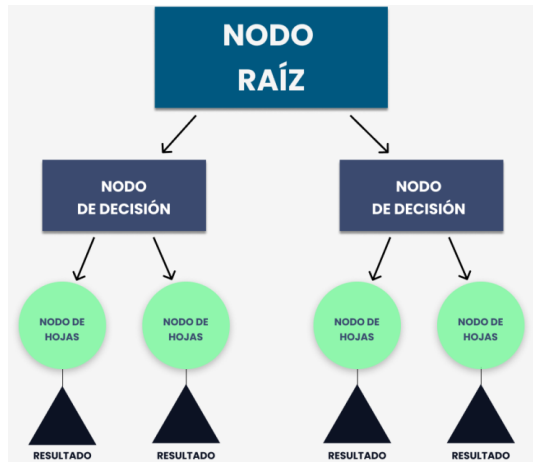
Garantiza que un modelo de ML pueda aprender patrones de todas las clases de una manera homogénea y efectiva, para así replicar predicciones



2. Materiales y Métodos

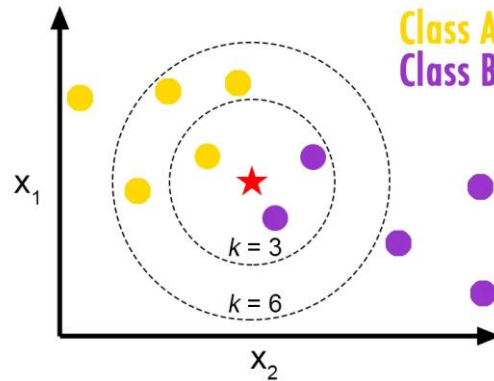
Entrenamiento

ÁRBOL DE DECISIÓN DT



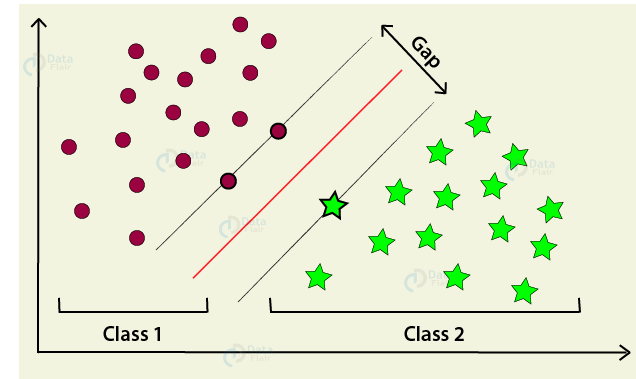
Barrido del número máximo de divisiones del árbol en sus nodos internos (MaxNumSplits).

k - VECINOS MÁS CERCANOS k-NN



Barrido del número de k vecinos más cercanos a considerar para la clasificación

MÁQUINAS DE VECTORES DE SOPORTE SVM



Parámetro C o restricción de caja (BoxConstraints), este controla la penalización de los errores en la clasificación durante el entrenamiento.

“Resto de parámetros por defecto”

2. Materiales y Métodos

Matriz de confusión, modelo 4 clases

Confusion Matrix

	VP			FP	
LP	67 22.3%	3 1.0%	1 0.3%	9 3.0%	83.8% 16.2%
TC	3 1.0%	68 22.7%	2 0.7%	1 0.3%	91.9% 8.1%
TR	2 0.7%	0 0.0%	72 24.0%	0 0.0%	97.3% 2.7%
VT	3 1.0%	4 1.3%	0 0.0%	65 21.7%	90.3% 9.7%
	FN			VN	
	89.3% 10.7%	90.7% 9.3%	96.0% 4.0%	86.7% 13.3%	90.7% 9.3%
	LP	TC	TR	VT	
	Target Class				

Métricas de rendimiento

Exactitud: $A\% = \frac{VP + VN}{VP + FP + FN + VN} \cdot 100$

Precisión: $P\% = \frac{VP}{VP + FP} \cdot 100$

Especificidad: $S\% = \frac{VN}{VN + FP} \cdot 100$

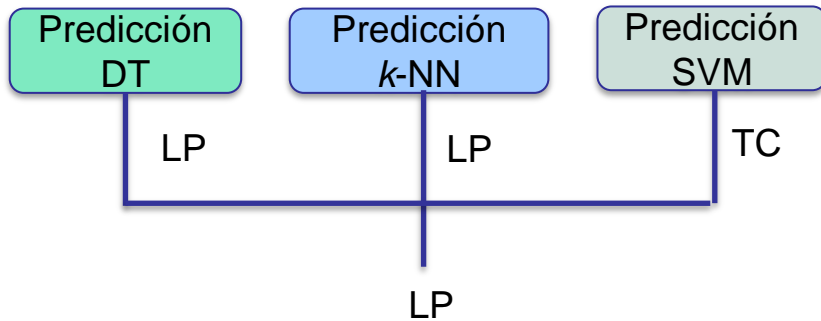
Sensibilidad: $R\% = \frac{VP}{VP + FN} \cdot 100$

Tasa de error balanceado: $BER = 1 - \frac{R + S}{200}$

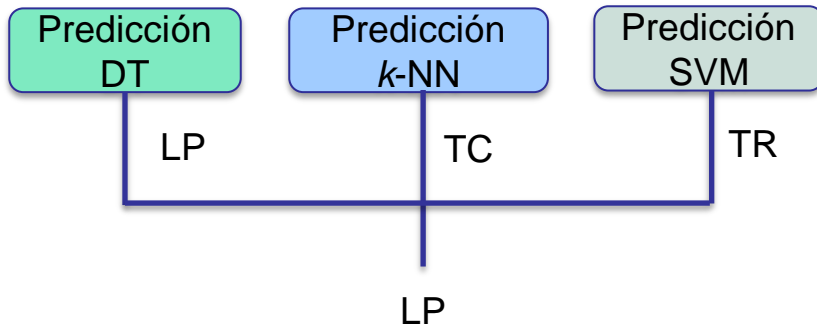
2. Materiales y Métodos

Algoritmo de votación: *Ejemplo*

Votación por mayoría



Votación por prioridad



Confusion Matrix

Output Class	LP	TC	TR	VT	
LP	67 22.3%	3 1.0%	1 0.3%	9 3.0%	83.8% 16.2%
TC	3 1.0%	68 22.7%	2 0.7%	1 0.3%	91.9% 8.1%
TR	2 0.7%	0 0.0%	72 24.0%	0 0.0%	97.3% 2.7%
VT	3 1.0%	4 1.3%	0 0.0%	65 21.7%	90.3% 9.7%
	89.3% 10.7%	90.7% 9.3%	96.0% 4.0%	86.7% 13.3%	90.7% 9.3%
	LP	TC	TR	VT	

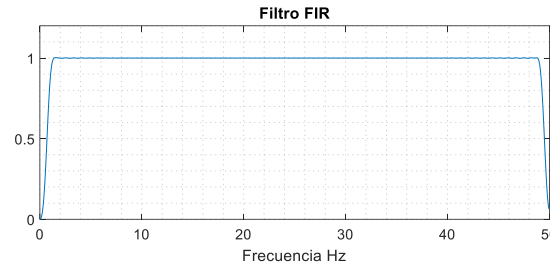
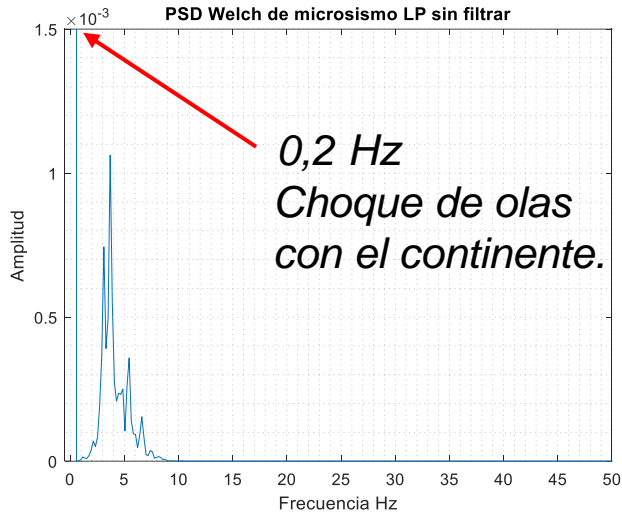
Microsismo	Prioridad de clasificación
LP	DT
TC	DT
TR	k-NN
VT	SVM



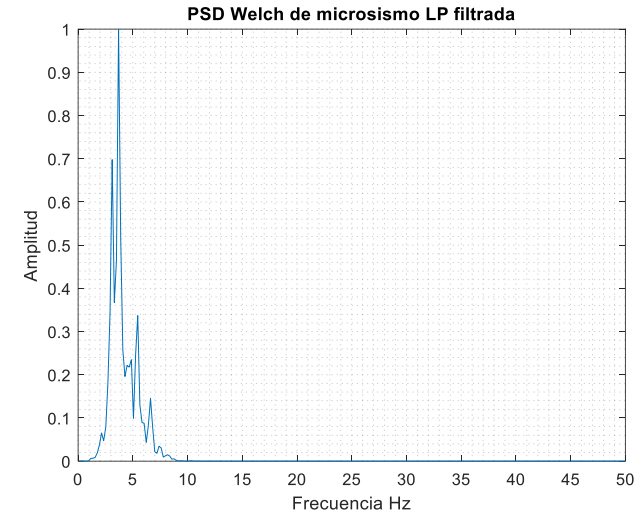
3. Pruebas y Resultados

Preprocesamiento

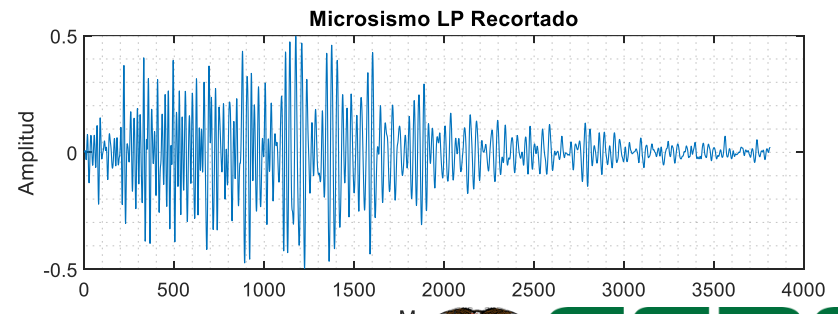
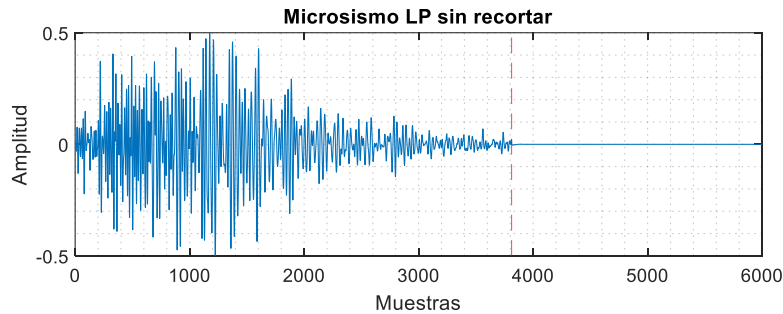
Filtro



- Filtro
FIR
- $f_{min} = 0,7 \text{ Hz}$
 - $f_{max} = 49,5 \text{ Hz}$
 - $Orden = 256 \text{ coef.}$



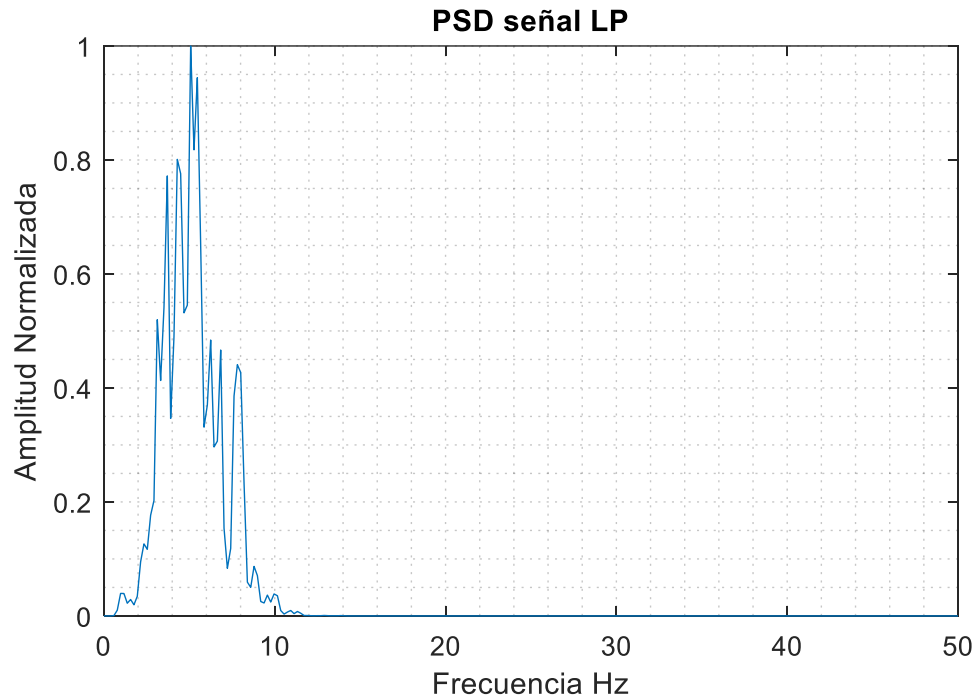
Recorte



3. Pruebas y Resultados

Procesamiento

Extracción de características PSD Welch y normalización



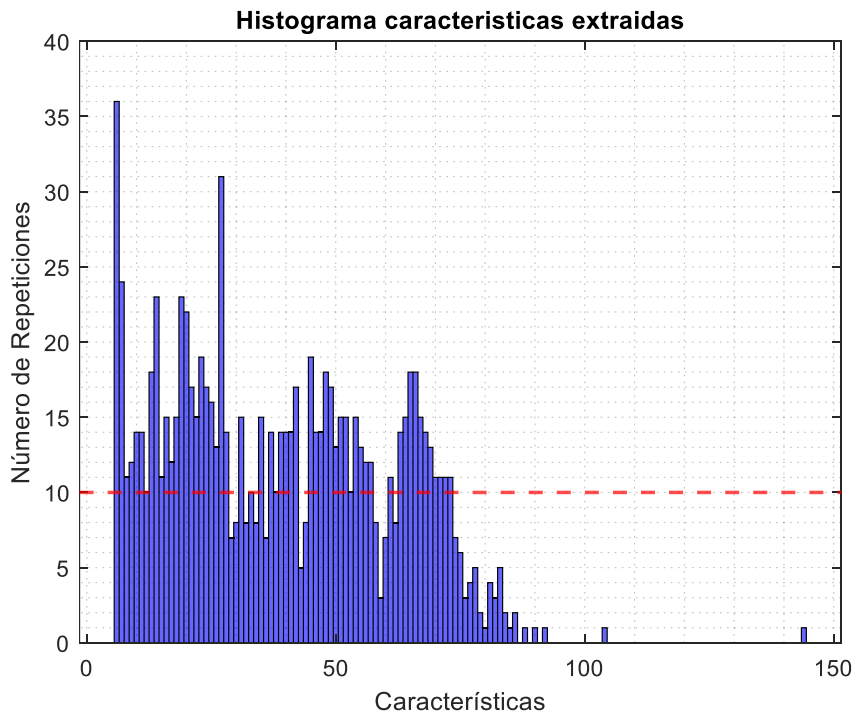
- Densidad espectral de potencia (PSD) método Welch:
 - $f_s = 100$ Hz
 - Resolución = 512
 - Ventana = Hamming
 - Solapamiento = 50%
 - Rango de frecuencias = $[0, f_s/2]$

3. Pruebas y Resultados

Procesamiento

RFE: Wrapper hacia adelante para PSD Welch

- Conjunto de datos: PSD Welch
- Algoritmo ML: SVM (multiclase)
- Métricas de evaluación: A%, P%, S%, R%, y BER.



Nro. Caract.	Frecuencia Hz
6 al 28	1,16 al 5,44
31	6,03
33	6,42
35	6,80
37 al 42	7,19 al 8,17
45 al 57	8,75 al 11,08
61	11,86
63 al 73	12,25 al 14,20

Total = 57 características seleccionadas



3. Pruebas y Resultados

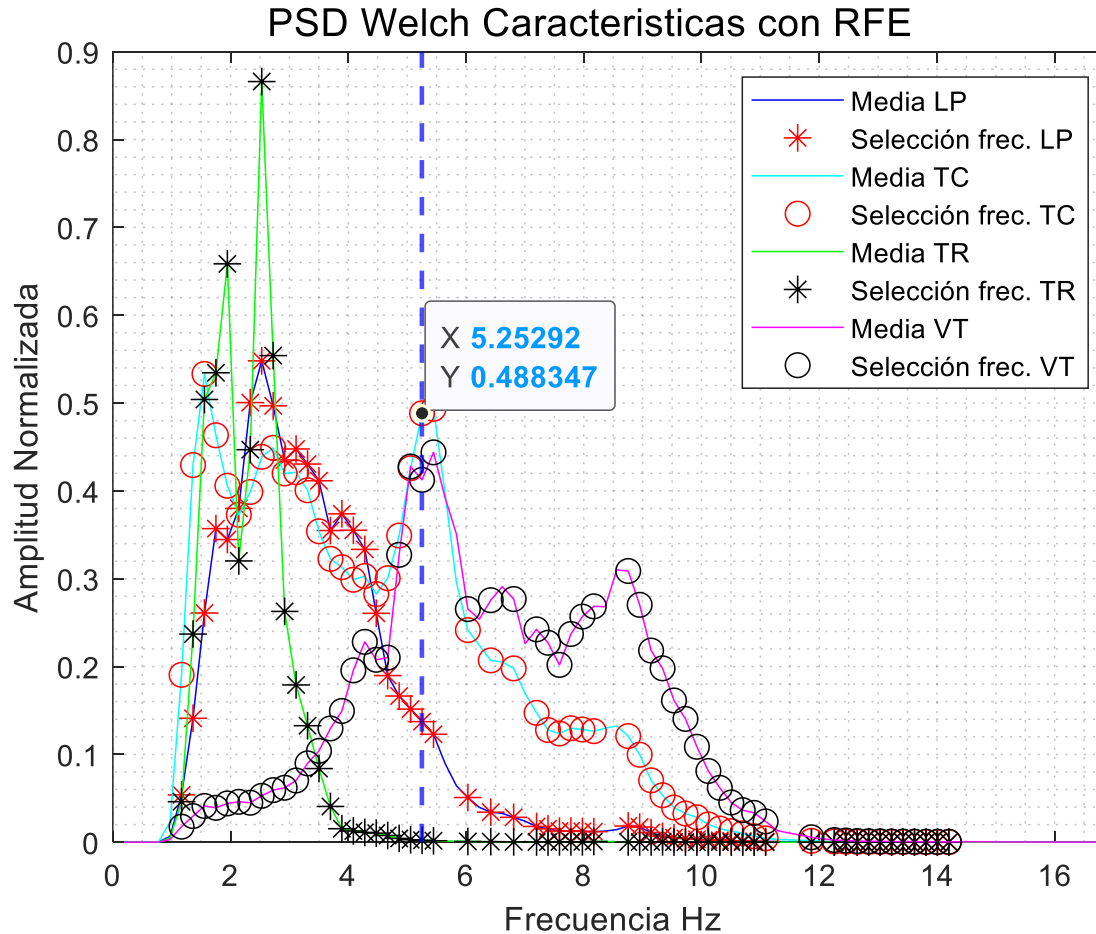
- RFE: Wrapper hacia adelante obtuvo 57 características.

Nro. Caract.	Frecuencia Hz	Nro. Caract.	Frecuencia Hz	Nro. Caract.	Frecuencia Hz
6	1,16	25	4,86	51	9,92
7	1,36	26	5,05	52	10,11
8	1,55	27	5,25	53	10,31
9	1,75	28	5,44	54	10,50
10	1,94	31	6,03	55	10,70
11	2,14	33	6,42	56	10,89
12	2,33	35	6,80	57	11,08
13	2,52	37	7,19	61	11,86
14	2,72	38	7,39	63	12,25
15	2,91	39	7,58	64	12,45
16	3,11	40	7,78	65	12,64
17	3,30	41	7,97	66	12,84
18	3,50	42	8,17	67	13,03
19	3,69	45	8,75	68	13,22
20	3,89	46	8,94	69	13,42
21	4,08	47	9,14	70	13,61
22	4,28	48	9,33	71	13,81
23	4,47	49	9,53	72	14,00
24	4,66	50	9,72	73	14,20

3. Pruebas y Resultados

Procesamiento

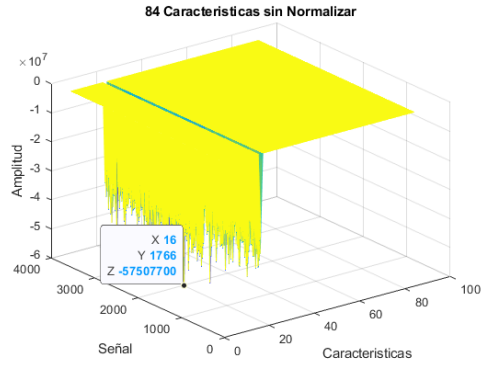
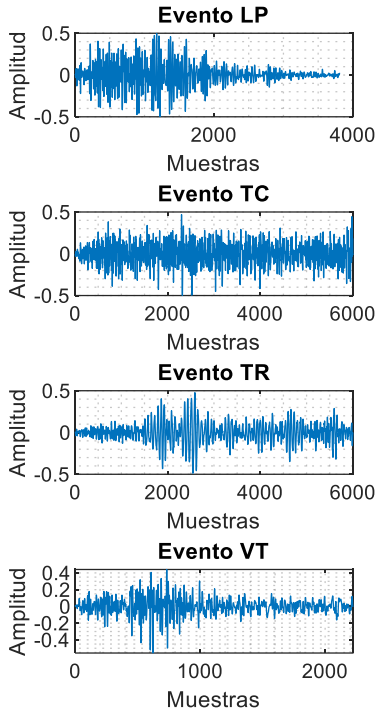
- Obtención de 57 característica mediante Wrapper hacia adelante



3. Pruebas y Resultados

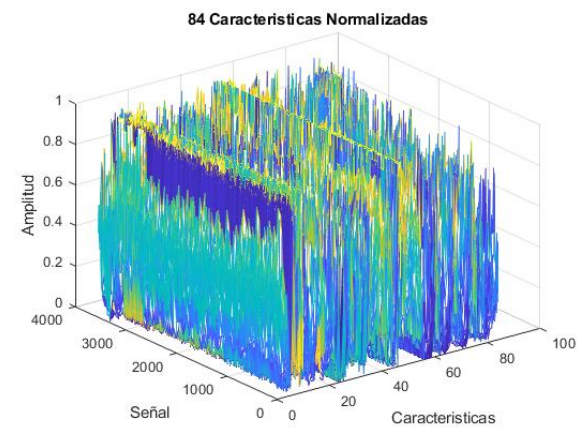
Procesamiento

- Extracción 84 Características



Extracción 84 Características

13 → Tiempo
21 → Frecuencia
50 → Escala



Normalizo
Para evitar la redundancia y anomalías de los datos

	1	2	3	...	83	84
	t_mean	t_std	t_var	...	w_t_peak2rms_D6	w_t_meanEnergyCoeff
1	-2.7427e-05	0.2683	0.0720	...	4.4975	0.0702
2	-8.2069e-05	0.1956	0.0383	...	5.4219	0.0375
3	3.3139e-05	0.2101	0.0442	...	4.0367	0.0435

3. Pruebas y Resultados

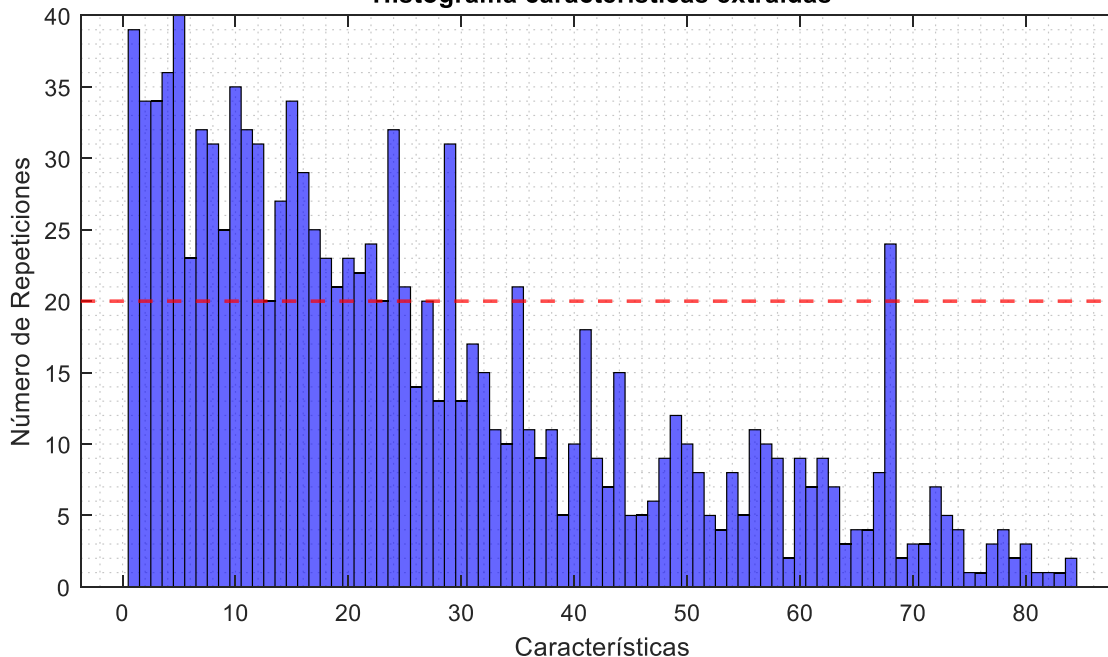
Procesamiento

RFE: Wrapper hacia adelante para 84 Características

- Conjunto de datos: 84 Características
- Algoritmo ML: SVM (multiclase)
- Métricas evaluación: A%, P%, S%, R% y BER.



Histograma características extraídas



ID	Dominio
f1 al f13	Tiempo
f14 al f25	Frecuencia
f27, f29	
f35, f68	Escala
Total de 29 características	

3. Pruebas y Resultados

Procesamiento

- Obtención de 29 característica mediante Wrapper hacia adelante

Tiempo	
ID	Característica
f1	Media
f2	Desviación Estándar
f3	Varianza
f4	Entropía
f5	Kurtosis
f6	Entropía Multiescala (MSE).
f7	Tiempo hasta máximo pico
f8	Valor RMS
f9	Valor pico a pico.
f10	Pico a RMS porción
f11	Energía
f12	Ratio del cero
f13	Density de picos above RMS

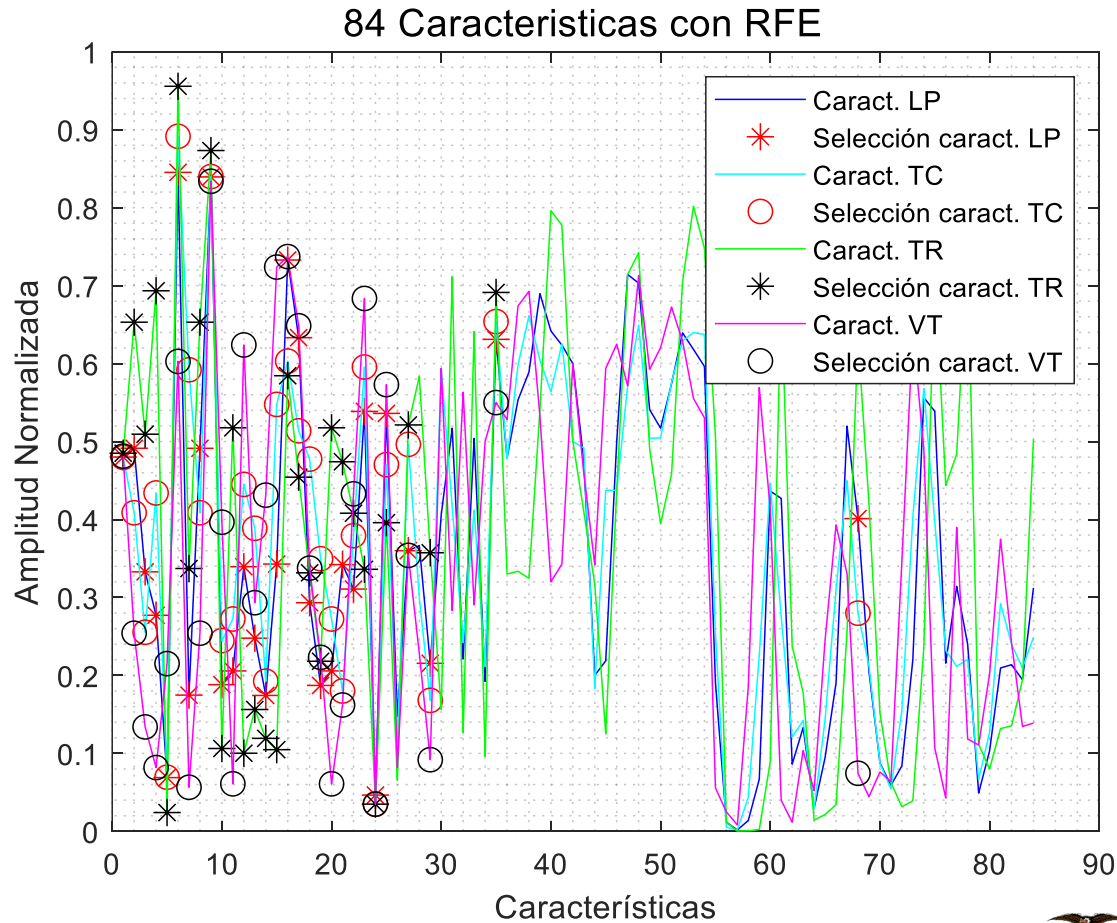
Frecuencia	
ID	Característica
f14	Frecuencia de máximo pico.
f15	Bandwidth de 90% energía
f16	Entropía
f17	Media
f18	Desviación Estándar
f19	Varianza
f20	Energía
f21	Kurtosis
f22	Entropía Multiescala
f23	Máximo pico en 10-20 Hz banda
f24	Frec. De máx. Pico en 10-20 Hz banda
f25	Máximo pico en 20-30 Hz banda
f27	Valor RMS
f29	Potencia RMS

Escala	
ID	Característica
f35	A6 Max. pico en frec-dominio
f68	5 RMS en tiempo-dominio

3. Pruebas y Resultados

Procesamiento

- Obtención de 29 característica mediante Wrapper hacia adelante



3. Pruebas y Resultados

Procesamiento

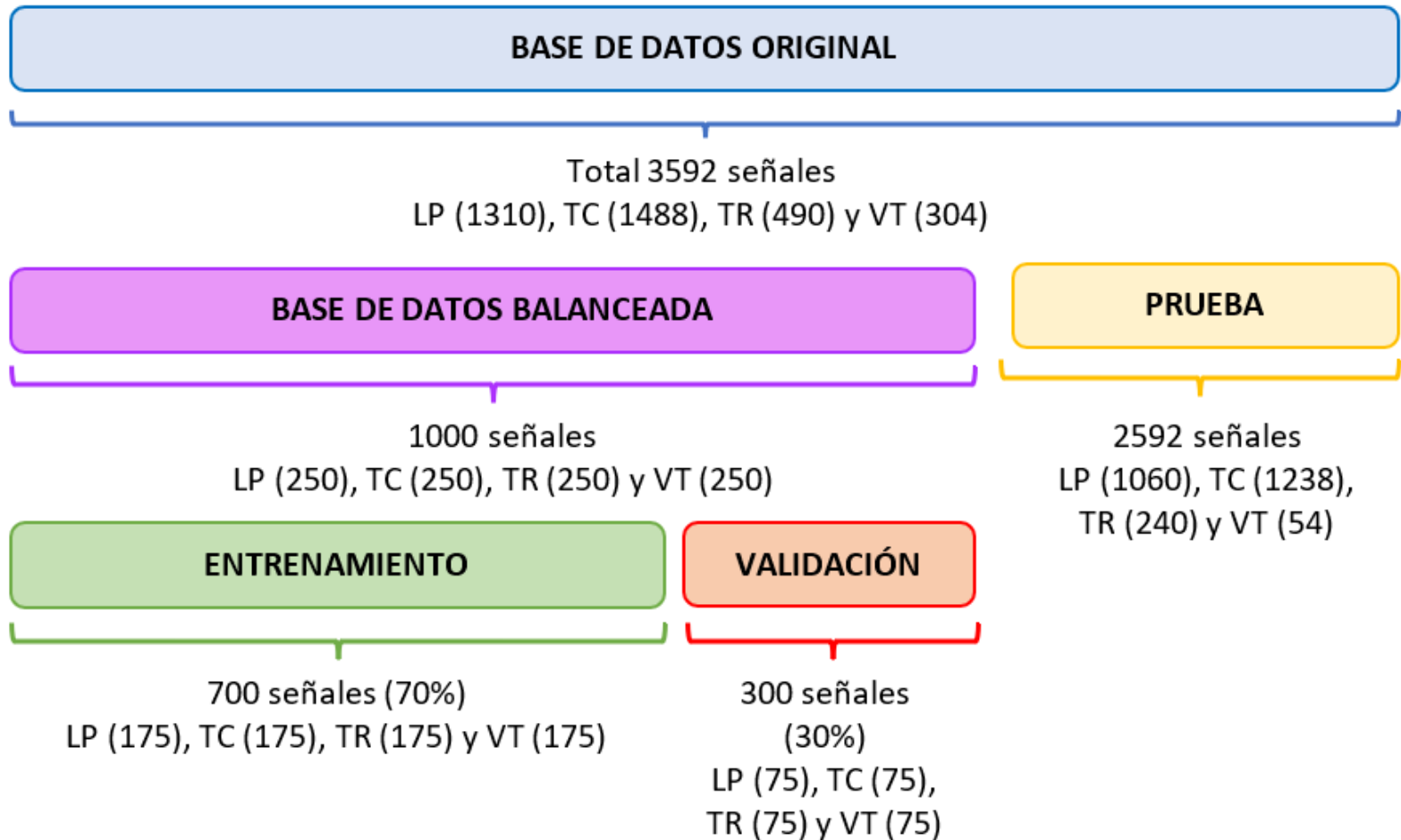
- Base de datos para entrenamiento

Base de datos	Nro. Caract.			Total
	Tiempo	Frecuencia	Escala	
PSD Welch	-	257	-	257
PSD Welch con RFE	-	57	-	57
84 Características	13	21	50	84
29 Características RFE	13	14	2	29
PSD Welch + 84 caract. (Tiempo, Frecuencia, Escala)	13	278	-	341
PSD Welch con RFE + 29 caract. extraídas de las 84 con RFE	13	71	2	286

3. Pruebas y Resultados

Procesamiento

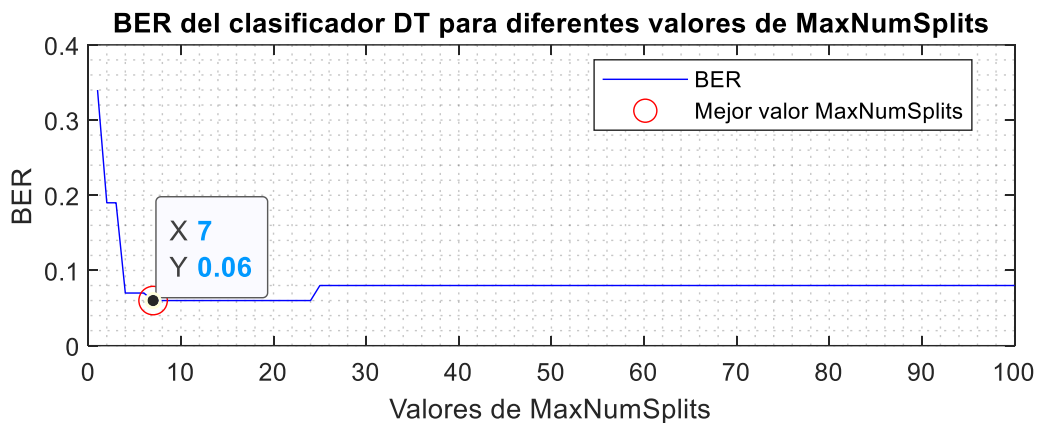
- Balanceo de datos



3. Pruebas y Resultados

Modelo DT

Resultados con datos de prueba



- MaxNumSplits (Nro. Max, de Divisiones): N
- MinLeafSize (Tamaño de hoja): 1
- MinParentSize (Tamaño Min Padre): 10
- SplitCriterion (Criterio de división): gdi
- PruneCriterion (Criterio de Poda): error
- PruneAlpha (Factor de poda): 0.05
- MergeLeaves (Fusionar hojas): off

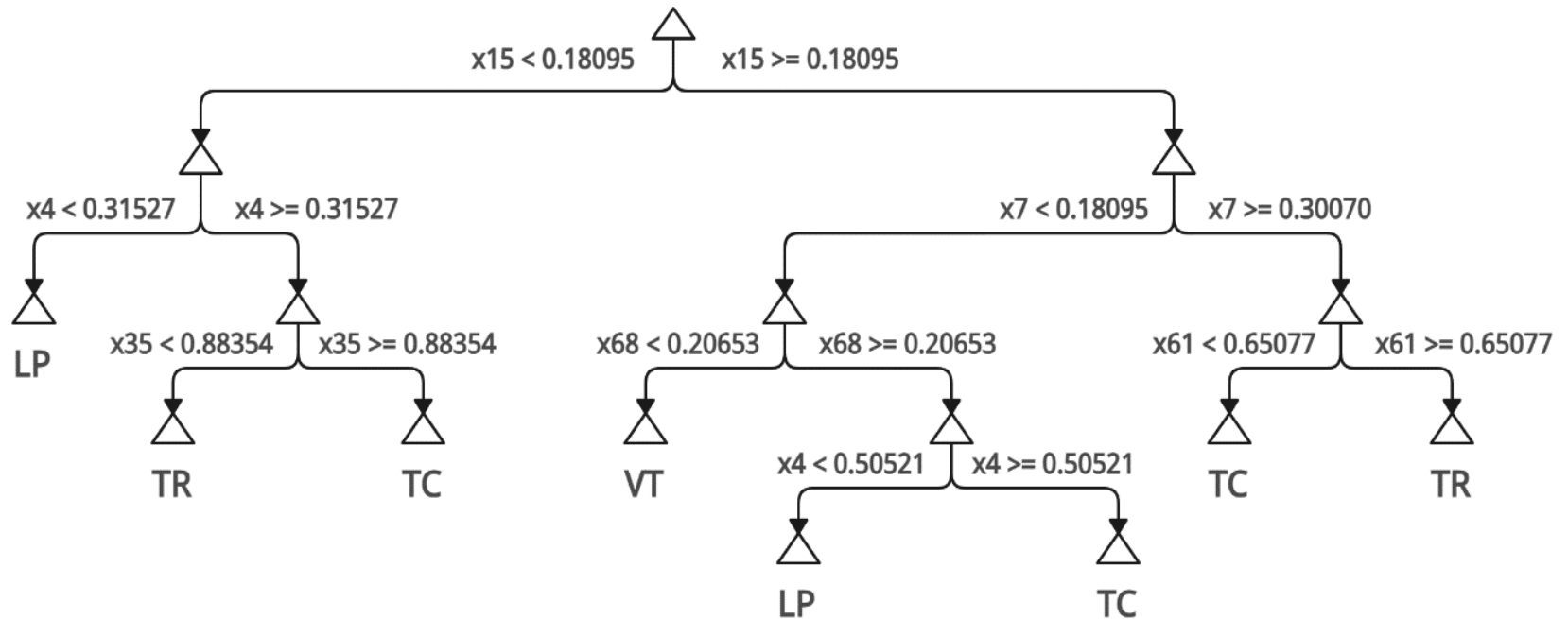
Base de Datos	MaxNumSplits	A %	P %	S %	R %	BER
PSD Welch (257 caract)	20	80	70	92	83	0,12
PSD Welch RFE (57 caract.)	5	84	73	93	82	0,12
84 Características	7	92	79	97	92	0,05
84 Características RFE (29 caract)	8	91	77	97	93	0,05
PSD Welch + 84 Caract.	9	92	79	97	93	0,05
PSD Welch RFE + 84 Caract. RFE	9	90	75	97	92	0,05



3. Pruebas y Resultados

Modelo DT

Resultado del mejor modelo con datos de entrenamiento.



Tiempo

ID	Característica
f4	Entropía
f7	Tiempo hasta máximo pico

Frecuencia

ID	Característica
f15	Bandwidth de 90% energía

Escala

ID	Característica
f35	A6 Max. Pico en frec-dom.
f61	% de energía para D5
f68	5 RMS en tiempo-dom.



3. Pruebas y Resultados

Modelo DT

Resultado del mejor modelo con datos de validación.

Modelo DT validación : 84 características					
Evento	A %	P %	S %	R %	BER
LP	93	84	94	89	0,08
TC	96	92	97	91	0,06
TR	98	97	99	96	0,02
VT	94	90	97	87	0,08
General	91	91	97	91	0,06

Confusion Matrix

Output Class	Target Class				
	LP	TC	TR	VT	
LP	67 22.3%	3 1.0%	1 0.3%	9 3.0%	83.8% 16.2%
TC	3 1.0%	68 22.7%	2 0.7%	1 0.3%	91.9% 8.1%
TR	2 0.7%	0 0.0%	72 24.0%	0 0.0%	97.3% 2.7%
VT	3 1.0%	4 1.3%	0 0.0%	65 21.7%	90.3% 9.7%
	89.3% 10.7%	90.7% 9.3%	96.0% 4.0%	86.7% 13.3%	90.7% 9.3%

3. Pruebas y Resultados

Modelo DT

Resultado del mejor modelo con datos de prueba.

Modelo DT prueba: 84 características con RFE					
Evento	A %	P %	S %	R %	BER
LP	95	96	97	92	0,05
TC	94	96	96	91	0,06
TR	98	87	99	94	0,04
VT	97	37	97	93	0,05
General	92	79	97	92	0,05

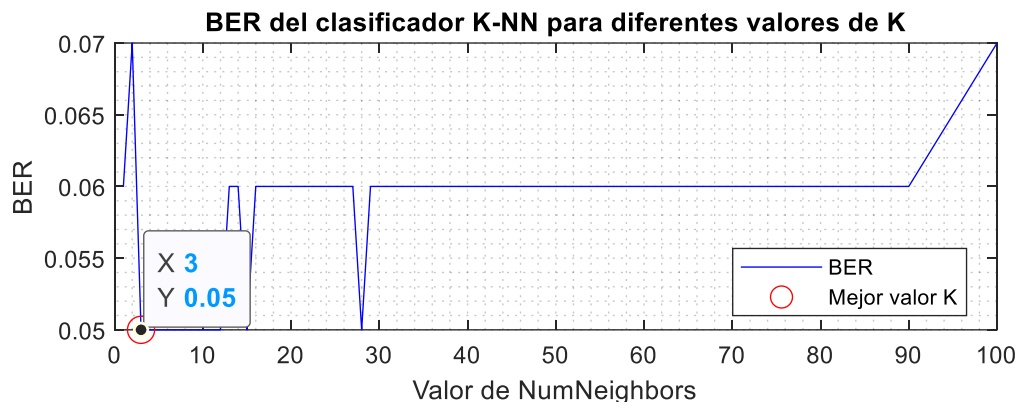
Confusion Matrix

Output Class	Target Class				
	LP	TC	TR	VT	
LP	980 37.8%	36 1.4%	2 0.1%	4 0.2%	95.9% 4.1%
TC	37 1.4%	1126 43.4%	13 0.5%	0 0.0%	95.7% 4.3%
TR	8 0.3%	27 1.0%	225 8.7%	0 0.0%	86.5% 13.5%
VT	35 1.4%	49 1.9%	0 0.0%	50 1.9%	37.3% 62.7%
	92.5% 7.5%	91.0% 9.0%	93.8% 6.2%	92.6% 7.4%	91.9% 8.1%

3. Pruebas y Resultados

Modelo k -NN

Resultados con datos de prueba.



- NumNeighbors (Nro. Vecinos): Varía
- Distance (Distancia): Euclidean
- NSMethod (Método para resolver empates): Exhaustive
- DistanceWeight (Peso de distancia): Equal
- Standardize (Estandarizar): false

Base de Datos	k	A %	P %	S %	R %	BER
PSD Welch (257 caract)	5	84	69	94	84	0,11
PSD Welch RFE (57 caract.)	9	83	68	94	86	0,10
84 Características	3	94	82	98	94	0,04
84 Características RFE (29 caract)	3	94	82	98	94	0,04
PSD Welch + 84 Caract.	4	92	80	97	93	0,05
PSD Welch RFE + 84 Caract. RFE	4	93	81	97	91	0,06



3. Pruebas y Resultados

Modelo *k*-NN (29 características)

Tiempo	
ID	Característica
f1	Media
f2	Desviación Estándar
f3	Varianza
f4	Entropía
f5	Kurtosis
f6	Entropía Multiescala (MSE).
f7	Tiempo hasta máximo pico
f8	Valor RMS
f9	Valor pico a pico.
f10	Pico a RMS porción
f11	Energía
f12	Ratio del cero
f13	Density de picos above RMS

Frecuencia	
ID	Característica
f14	Frecuencia de máximo pico.
f15	Bandwidth de 90% energía
f16	Entropía
f17	Media
f18	Desviación Estándar
f19	Varianza
f20	Energía
f21	Kurtosis
f22	Entropía Multiescala
f23	Máximo pico en 10-20 Hz banda
f24	Frec. De max. Pico en 10-20 Hz banda
f25	Máximo pico en 20-30 Hz banda
f27	Valor RMS
f29	Potencia RMS

Escala	
ID	Característica
f35	A6 Max. pico en frec-dominio
f68	5 RMS en tiempo-dominio

3. Pruebas y Resultados

Modelo *k*-NN

Resultado del mejor modelo con datos de validación.

Modelo <i>k</i> -NN validación : 84 características					
Evento	A %	P %	S %	R %	BER
LP	95	86	95	96	0,05
TC	98	99	100	92	0,04
TR	99	99	100	96	0,02
VT	95	90	97	88	0,08
General	93	93	98	93	0,05

Confusion Matrix

Output Class	Target Class				
	LP	TC	TR	VT	
LP	72 24.0%	1 0.3%	2 0.7%	9 3.0%	85.7% 14.3%
TC	0 0.0%	69 23.0%	1 0.3%	0 0.0%	98.6% 1.4%
TR	0 0.0%	1 0.3%	72 24.0%	0 0.0%	98.6% 1.4%
VT	3 1.0%	4 1.3%	0 0.0%	66 22.0%	90.4% 9.6%
	96.0% 4.0%	92.0% 8.0%	96.0% 4.0%	88.0% 12.0%	93.0% 7.0%

3. Pruebas y Resultados

Modelo k -NN

Resultado del mejor modelo con datos de prueba.

Modelo k -NN prueba: 84 características con RFE

<i>Evento</i>	<i>A %</i>	<i>P %</i>	<i>S %</i>	<i>R %</i>	<i>BER</i>
LP	97	96	97	95	0,04
TC	96	98	98	93	0,05
TR	98	86	98	94	0,04
VT	98	48	98	94	0,04
General	94	82	98	94	0,04

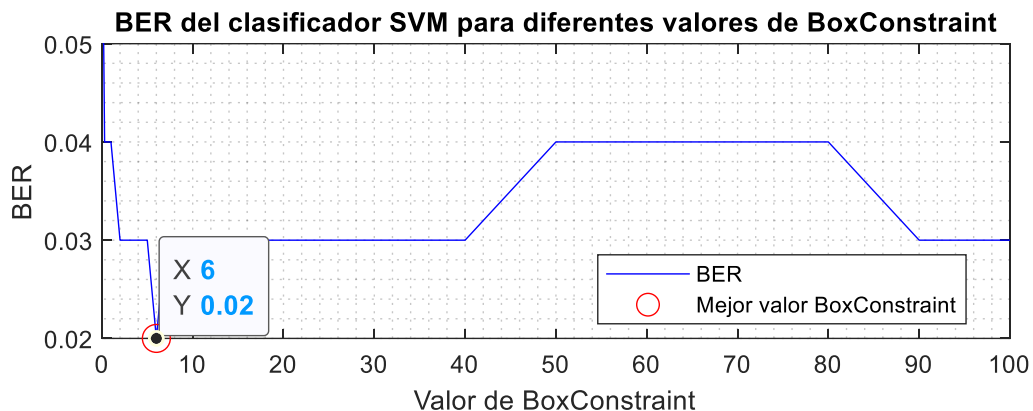
Confusion Matrix

	LP	TC	TR	VT	
LP	1010 39.0%	33 1.3%	4 0.2%	3 0.1%	96.2% 3.8%
TC	15 0.6%	1147 44.3%	10 0.4%	0 0.0%	97.9% 2.1%
TR	7 0.3%	30 1.2%	226 8.7%	0 0.0%	85.9% 14.1%
VT	28 1.1%	28 1.1%	0 0.0%	51 2.0%	47.7% 52.3%
	95.3% 4.7%	92.6% 7.4%	94.2% 5.8%	94.4% 5.6%	93.9% 6.1%
	LP	TC	TR	VT	

3. Pruebas y Resultados

Modelo SVM

Resultados con datos de prueba.



- BoxConstraint (Restricción de Caja): Varia
- Coding (Codificación): onevsone
- Cost (Costo): vacío.
- FitPosterior (Ajustar Probabilidades a Posteriori): false.
- Options (Opciones): Vacio

Base de Datos	BoxConstraint	A %	P %	S %	R %	BER
PSD Welch (257 caract.)	0,4	82	69	94	87	0,1
PSD Welch RFE (57 caract.)	2	85	72	95	88	0,08
84 Características	6	95	84	98	95	0,03
84 Características RFE (29 caract.)	6	96	85	99	96	0,03
PSD Welch + 84 caract.	0,5	86	76	96	89	0,07
PSD Welch RFE + 84 caract. RFE	0,5	92	79	97	93	0,05



3. Pruebas y Resultados

Modelo SVM

Resultado del mejor modelo con datos de validación.

Modelo SVM entrenamiento : 84 características					
Evento	A %	P %	S %	R %	BER
LP	97	91	97	97	0,03
TC	99	100	100	96	0,02
TR	99	100	100	97	0,01
VT	97	95	98	95	0,04
General	96	96	99	96	0,02

Confusion Matrix

Output Class	Target Class				
	LP	TC	TR	VT	
LP	73 24.3%	1 0.3%	2 0.7%	4 1.3%	91.2% 8.8%
TC	0 0.0%	72 24.0%	0 0.0%	0 0.0%	100% 0.0%
TR	0 0.0%	0 0.0%	73 24.3%	0 0.0%	100% 0.0%
VT	2 0.7%	2 0.7%	0 0.0%	71 23.7%	94.7% 5.3%
	97.3% 2.7%	96.0% 4.0%	97.3% 2.7%	94.7% 5.3%	96.3% 3.7%

3. Pruebas y Resultados

Modelo SVM

Resultados del mejor modelo con datos de prueba.

Modelo SVM prueba : 84 características con RFE

Evento	A %	P %	S %	R %	BER
LP	98	99	99	96	0,02
TC	97	98	98	96	0,03
TR	99	92	99	95	0,03
VT	98	54	98	98	0,02
General	96	85	99	96	0,03

Confusion Matrix

Output Class	Target Class				
	LP	TC	TR	VT	
LP	1020 39.4%	11 0.4%	3 0.1%	1 0.0%	98.6% 1.4%
TC	16 0.6%	1184 45.7%	9 0.3%	0 0.0%	97.9% 2.1%
TR	3 0.1%	18 0.7%	228 8.8%	0 0.0%	91.6% 8.4%
VT	21 0.8%	25 1.0%	0 0.0%	53 2.0%	53.5% 46.5%
	96.2% 3.8%	95.6% 4.4%	95.0% 5.0%	98.1% 1.9%	95.9% 4.1%

3. Pruebas y Resultados

Algoritmo votación

Resultados con datos de prueba en Votación

Mejor Modelo	A %	P %	S %	R %	BER
DT	92	79	97	92	0,05
k-NN	94	82	98	94	0,04
SVM	96	85	99	96	0,03
Votación	96	85	99	96	0,03

Votación Mejores Modelos					
Evento	A %	P %	S %	R %	BER
LP	98	98	99	97	0,02
TC	97	98	98	95	0,03
TR	99	93	99	95	0,03
VT	98	50	98	96	0,03
General	96	85	99	96	0,03

Confusion Matrix

Output Class	Target Class				
	LP	TC	TR	VT	
LP	1024 39.5%	17 0.7%	3 0.1%	2 0.1%	97.9% 2.1%
TC	11 0.4%	1177 45.4%	10 0.4%	0 0.0%	98.2% 1.8%
TR	3 0.1%	15 0.6%	227 8.8%	0 0.0%	92.7% 7.3%
VT	22 0.8%	29 1.1%	0 0.0%	52 2.0%	50.5% 49.5%
	96.6% 3.4%	95.1% 4.9%	94.6% 5.4%	96.3% 3.7%	95.7% 4.3%



3. Pruebas y Resultados

Interfaz Gráfica

Detection and classification of microseismic events

Detección: Castillo Tipantuña Edwin Alexander
Ortega Noroña Oscar Ortega

Clasificación: Cachipueno Yacelga Cesar Ariel
Tuttilo Moyón Javier Alejandro

Director: Dr. Román Lara
Colaborador: Ing. Fernando Lara

© ESPE 2023. Todos los Derechos Reservados

Configuración de operación automática.

Ruta Señales
Ruta Redes
Ruta Almacenamiento

parámetros de archivos a procesar

Cotopaxi Llaima

Nombre: testLlaima

Numeración: 0, 1, 4

Formato: *.mat, *.seed

Detección Rápida: Sí, No

Formato CNN: *.mat, *.onnx

Formato ML: *.mat, *.onnx

Cambio automatico de archivo: Tiempo [min] 0 5 10 15 20 25 30

Selección el detector a utilizar: Homomorfico, ML, VAD

Coordenadas Volcán: Latitud -38.7167, Longitud -71.7167

Universidad de las Fuerzas Armadas ESPE - Fernando Lara - nando9571@gmail.com

Mapa de Volcán Llaima: 38°40'S, 38°45'S, 71°50'W, 71°45'W, 71°40'W, 71°35'W

Registro Sismico

Registro de Detección

Espectrograma Detección

Espectrograma Clasificación

Densidad Espectral Microsismo

Microsismo

Panel de Estados: Inico, Detección, Fin, Clasificación

Registro Sismico: Amplitud vs. tiempo (s)

Registro de Detección: Amplitud vs. tiempo (s) with detection boxes

Espectrograma Detección: Frecuencia (Hz) vs. Muestras

Espectrograma Clasificación: Frecuencia (Hz) vs. Muestras

Densidad Espectral Microsismo: Normalized Amplitude vs. Frequency (Hz)

Microsismo: Amplitud vs. tiempo (s)

Tipo Evento detectado: **TR**

Registro: %s Tipo LP (%)-Inicio:4900(%)-Fin:9900
%s Tipo LP (%)-Inicio:19500(%)-Fin:25500
%s Tipo TR (%)-Inicio:33000(%)-Fin:37500

Datos de microsismos reconocidos

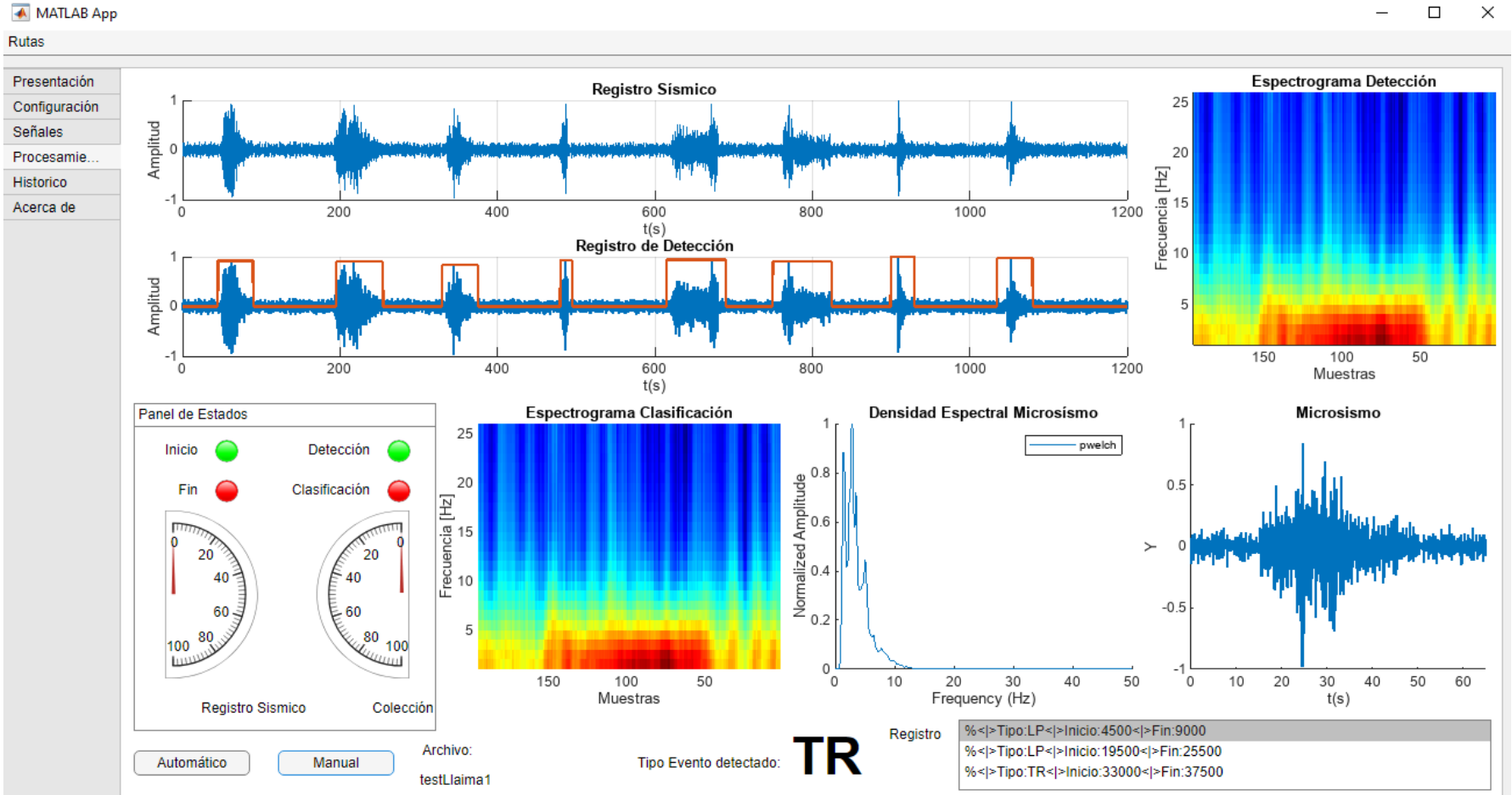
Network	Station	SampleRate	Component	Year	Month	Type	Duration(s)	StartPoint	EndPoint	Data
Chile	LAV	100	SHZ	2023	8	LP	65 4500	9000		1x0500
Chile	LAV	100	SHZ	2023	8	LP	80 19500	25500		1x8000
Chile	LAV	100	SHZ	2023	8	TR	65 33000	37500		1x6500
Chile	LAV	100	SHZ	2023	8	VT	35 48000	49500		1x3500
Chile	LAV	100	SHZ	2023	8	TR	95 61500	69000		1x9500
Chile	LAV	100	SHZ	2023	8	TR	95 75000	82500		1x9500
Chile	LAV	100	SHZ	2023	8	TR	80 90000	93000		1x5000
Chile	LAV	100	SHZ	2023	8	TR	65 103500	108000		1x6500

Estado: Almacenar



3. Pruebas y Resultados

Interfaz Gráfica



4. Conclusiones

- Se ha desarrollado un sistema de clasificación de microsismos en el volcán Llaima utilizando tres algoritmos principales: Árboles de Decisión (DT), k -Nearest Neighbors (k -NN) y Máquinas de Vectores de Soporte (SVM), junto con un algoritmo de votación para mejorar los resultados. Este sistema se basó en criterios fundamentales establecidos a través de la revisión exhaustiva de investigaciones anteriores, incluyendo la extracción y selección de características utilizando *Recursive Feature Elimination* (RFE) y métricas de rendimiento.
- La normalización de la base de datos se ha revelado como un paso crítico para mejorar significativamente las métricas de desempeño y acelerar el proceso de entrenamiento y extracción de características. Esto asegura que los modelos generados sean precisos y eficientes en la clasificación de nuevas señales sísmicas.
- La extracción y selección de 84 características en los dominios del tiempo, frecuencia y escala fueron fundamentales para el éxito de las técnicas de aprendizaje supervisado.



4. Conclusiones

- Aunque los resultados no cumplieron con el requisito del Instituto Geofísico de un BER igual a 0.01, se lograron bajos valores de la tasa de error y exactitud (BER, A%) con SVM (0.03; 96), k-NN (0.04; 94) y DT (0.05; 92).
- El algoritmo de votación tuvo un impacto neutral en las métricas de rendimiento, con un ligero aumento en Exactitud y Precisión del 0.5%. Los tiempos de clasificación se distribuyeron de la siguiente manera: 0.79 segundos para DT, 0.88 segundos para *k*-NN, 1.28 segundos para SVM y 17.84 segundos para el algoritmo de votación.
- Este estudio representa un avance importante en la clasificación de microsismos volcánicos y proporciona una base sólida para futuras investigaciones y aplicaciones en la detección temprana de actividad volcánica.



4. Trabajos Futuros

- Los trabajos futuros se enfocan en validar y mejorar el clasificador de microsismos desarrollado. Primordialmente, se busca verificar su precisión con datos en tiempo real del volcán Llaima para ajustar el modelo según las condiciones cambiantes. Además, se pretende evaluar su aplicabilidad en otros volcanes, como el Cotopaxi.
- Se considera la incorporación de otras técnicas de selección de características, como filtros o embebidos, para identificar características clave. La exploración de técnicas de *Deep Learning* y la generación de bases de datos sintéticas se plantean para capturar patrones complejos en los microsismos.
- Investigar la eficacia del algoritmo de votación y su combinación con otros modelos de ML, como *Naive Bayes* y *Extreme Gradient Boosting*. También se analiza la variación de parámetros en cada algoritmo para optimizar su rendimiento.
- En resumen, se busca mejorar la adaptabilidad, precisión y robustez del clasificador de microsismos, lo que contribuirá notablemente a la detección temprana y la gestión de riesgos volcánicos.



DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES

CARRERA DE TELECOMUNICACIONES

“Creación de algoritmos inteligentes basados en la teoría de machine learning tradicional para la clasificación de los eventos sísmicos en el volcán Llaima (multiclase)”

Autores: Cachipueno Yacelga , César Ariel y Tutillo Moyón , Javier Alejandro

Director del Proyecto: Ing. Román Alcides Lara Cueva, PhD.

06 de septiembre de 2023

