

PROYECTO FINAL INTELIGENCIA ARTIFICIAL Y APRENDIZAJE

MARÍA GONZÁLEZ GARCÍA
ANA SAN ROMÁN GAITERO

Curso 2020/21

Grado en Ingeniería Biomédica

Presentación del problema

- * LBE: valor de FHS determinado por el experto (latidos por minuto)
- * LB: valor de FHS determinado por SisPorto (latidos por minuto)
- * AC: número de aceleraciones por segundo
- * FM: número de movimientos fetales por segundo
- * UC: número de contracciones uterinas por segundo
- * ASTV: porcentaje de tiempo con variabilidad anormal a corto plazo
- * MSTV: valor medio de la variabilidad a corto plazo
- * ALTV: porcentaje de tiempo con variabilidad anormal a largo plazo
- * MLTV: valor medio de la variabilidad a largo plazo
- * DL: número de pequeñas desaceleraciones por segundo
- * DS: número de desaceleraciones severas por segundo
- * DP: número de desaceleraciones prolongadas por segundo
- * DR: número de desaceleraciones repetitivas
- * Width: ancho de FHR del histograma
- * Min: FHR mínima del histograma
- * Max: FHR máxima del histograma
- * Nmax: número de picos en el histograma
- * Nzeros: número de ceros en el histograma
- * Mode: moda del histograma
- * Mean: media del histograma
- * Median: mediana del histograma
- * Variance: varianza del histograma
- * Tendency: tendencia del histograma (sesgado a la izquierda: -1; simétrico: 0; sesgado a la derecha: 1)
- * NSP: feto en estado normal: 1; en estado sospechoso: 2; en estado patológico: 3

Mediciones:

- Frecuencia cardíaca fetal .
- Contracciones uterinas

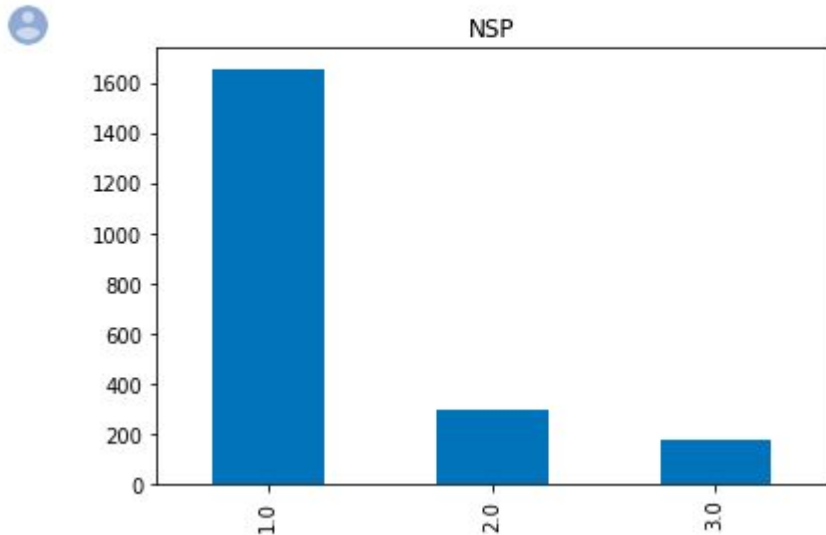
¿Objetivo?

- Estado fetal normal
- Estado fetal sospechoso
- Estado fetal patológico

Análisis Descriptivo de la base de datos

- Variable de salida NSP

```
data['NSP'].value_counts().plot.bar(title="NSP");
```



Categoría 1

77,84% del total de datos

Categoría 2

13,88% del total de datos

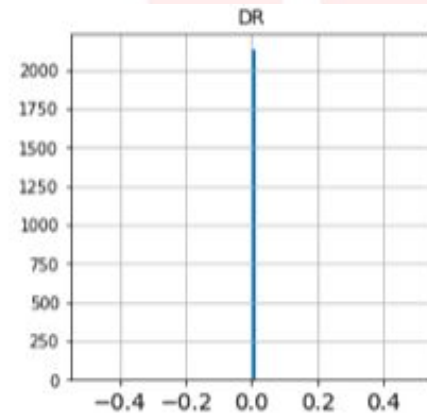
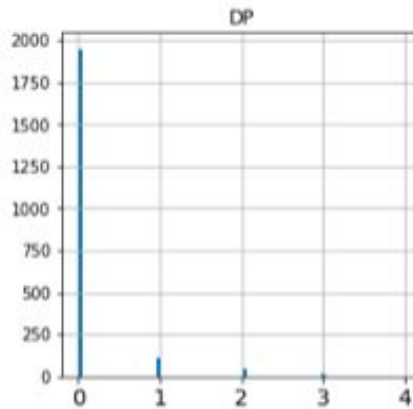
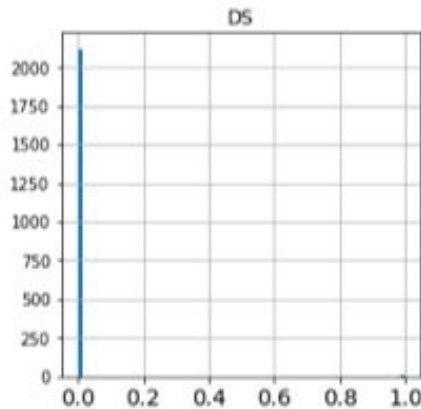
Categoría 3

8,28% del total de datos.

BASE DE DATOS DESBALANCEADA

Análisis Descriptivo de la base de datos

- Variables de entrada - Variables numéricas



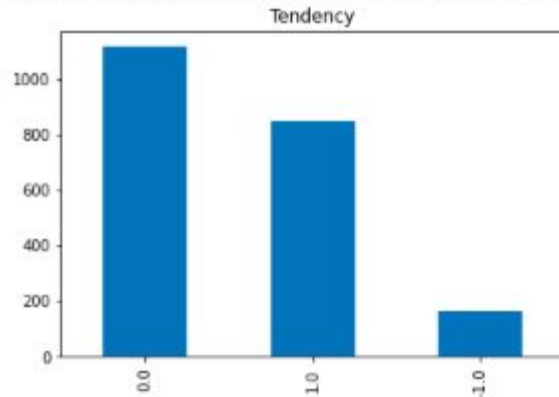
Aunque son de tipo numérico, apenas toman unos pocos valores y la gran mayoría de observaciones pertenecen a solo dos, cuatro o uno de ellos, respectivamente.

Análisis Descriptivo de la base de datos

- Variables de entrada - Variable cualitativa

```
data['Tendency'].value_counts().plot.bar(title="Tendency")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fc2c3ccb1d0>
```



```
[ ] data.Tendency.value_counts()
```

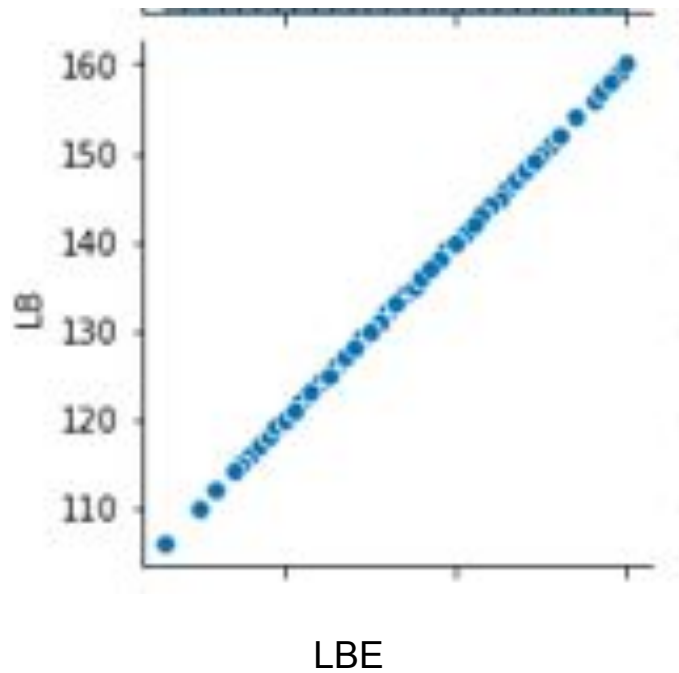
```
0.0    1115
1.0     846
-1.0    165
Name: Tendency, dtype: int64
```

Categoría predominante

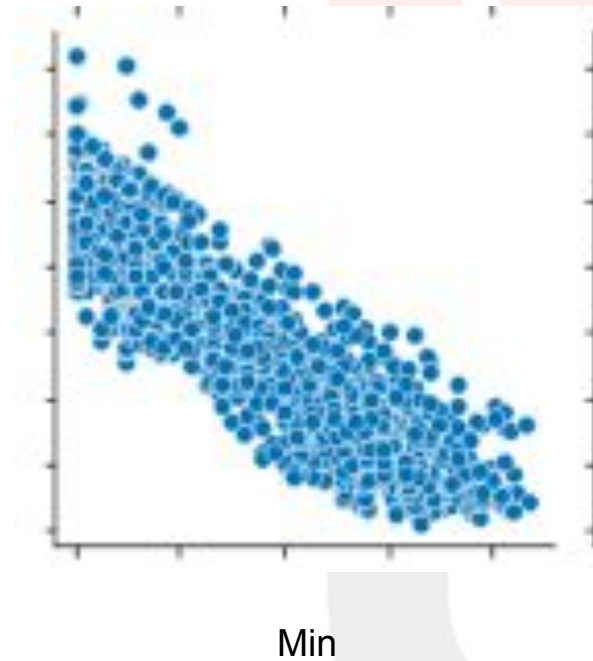
Histogramas de FHR simétricos.

Análisis Descriptivo de la base de datos

- Relación entre variables

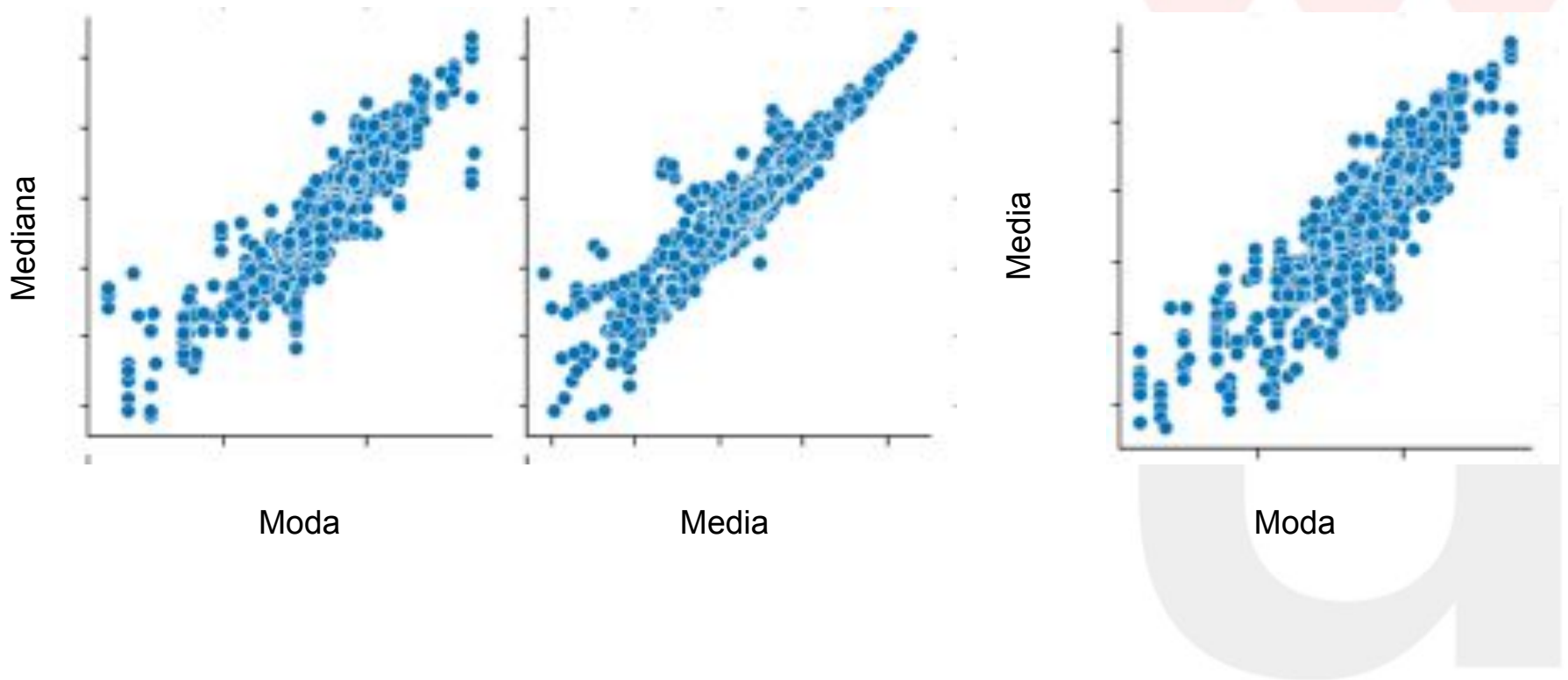


Width



Análisis Descriptivo de la base de datos

- Relación entre variables



Análisis Descriptivo de la base de datos



Análisis Descriptivo de la base de datos



Preprocesado de la base de datos

X_diseño

LBE 0
LB 0
AC 0
FM 0
UC 0
ASTV 0
MSTV 0
ALTV 0
MLTV 0
DL 0
DS 0
DP 0
DR 0
Width 0
Min 0
Max 0
Nmax 0
Nzeros 0
Mode 0
Mean 0
Median 0
Variance 0
Tendency 0
dtype: int64

Missing values

	LBE	LB	AC	FM	UC	ASTV	MSTV	ALTV	MLTV
count	1488.000000	1488.000000	1488.000000	1488.000000	1488.000000	1488.000000	1488.000000	1488.000000	1488.000000
mean	133.531586	133.531586	2.711022	7.172715	3.618952	47.029570	1.327755	9.920699	8.128293
std	9.812106	9.812106	3.524507	37.483828	2.836375	17.075699	0.894481	18.527151	5.593993
min	106.000000	106.000000	0.000000	0.000000	0.000000	12.000000	0.200000	0.000000	0.000000
25%	127.000000	127.000000	0.000000	0.000000	1.000000	32.000000	0.700000	0.000000	4.600000
50%	133.000000	133.000000	1.000000	0.000000	3.000000	49.000000	1.200000	0.000000	7.400000
75%	141.000000	141.000000	4.000000	2.000000	5.000000	61.000000	1.700000	11.000000	10.700000
max	160.000000	160.000000	26.000000	564.000000	23.000000	86.000000	7.000000	91.000000	50.700000

y_diseño 0

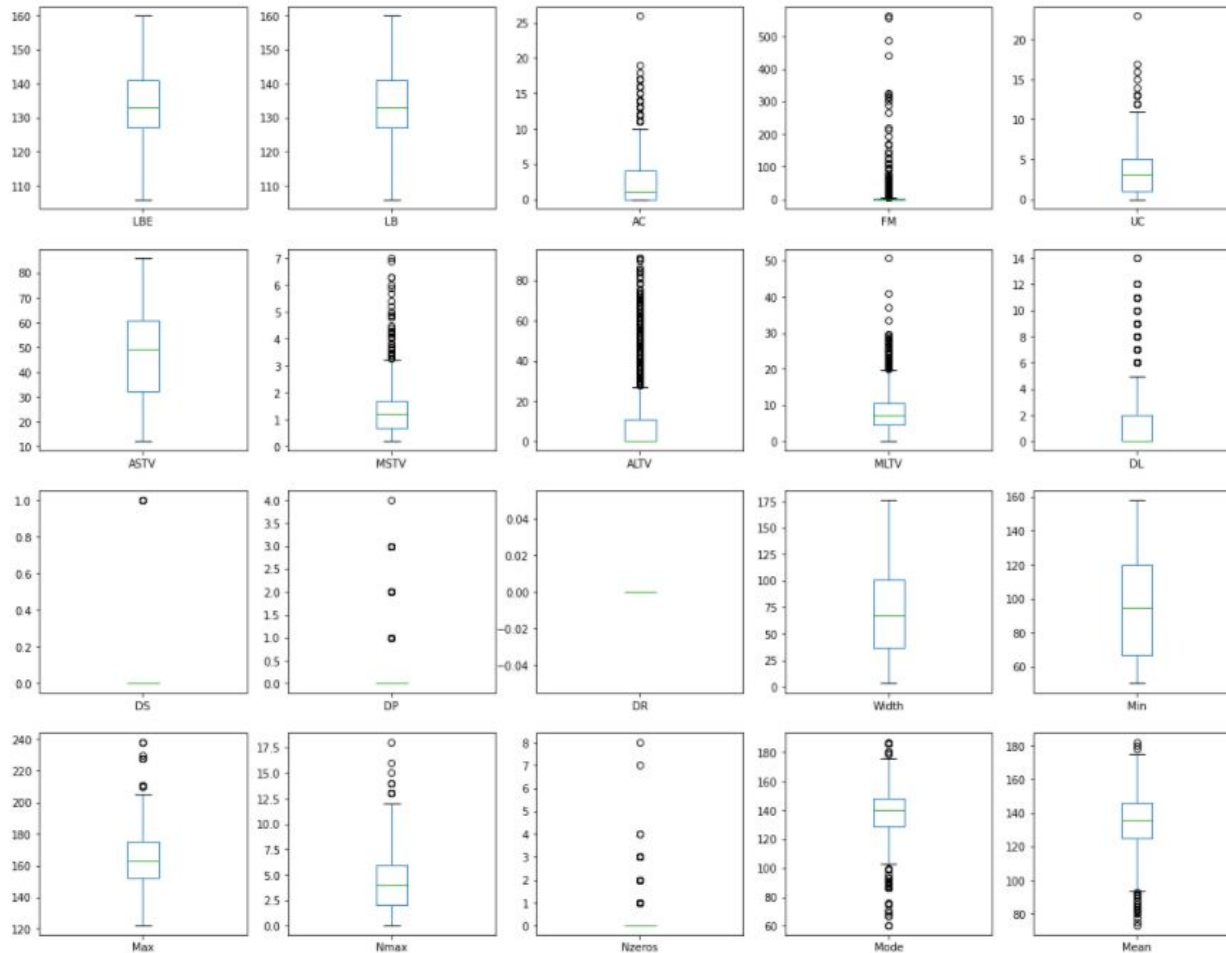
y_test 0

X_test

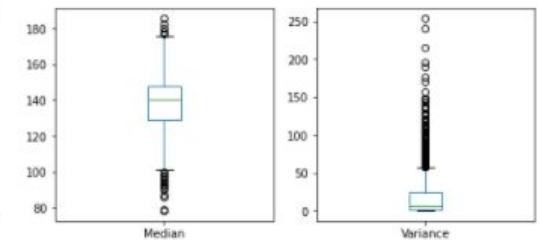
LBE 0
LB 0
AC 0
FM 0
UC 0
ASTV 0
MSTV 0
ALTV 0
MLTV 0
DL 0
DS 0
DP 0
DR 0
Width 0
Min 0
Max 0
Nmax 0
Nzeros 0
Mode 0
Mean 0
Median 0
Variance 0
Tendency 0
dtype: int64

Preprocesado de la base de datos

Outliers



Inspección visual
para localizar
valores fuera del
rango normal de
valores



Preprocesado de la base de datos

Outliers

- 1.- Detección outliers mediante z-score
- 2.- Reemplazamiento por Nan
- 3.- Cálculo estadístico mediana
- 4.- Reemplazamiento por estadístico mediana

X_diseño

LBE	17
LB	17
AC	48
FM	23
UC	25
ASTV	0
MSTV	33
ALTV	74
MLTV	33
DL	57
DS	5
DP	46
DR	0
Width	3
Min	0
Max	15
Nmax	33
Nzeros	19
Mode	50
Mean	32
Median	34
Variance	57
Tendency	0

X_test

LBE	7
LB	7
AC	19
FM	12
UC	13
ASTV	0
MSTV	16
ALTV	33
MLTV	16
DL	24
DS	2
DP	23
DR	0
Width	4
Min	0
Max	9
Nmax	8
Nzeros	9
Mode	14
Mean	19
Median	8
Variance	12
Tendency	0

LBE	0
LB	0
AC	0
FM	0
UC	0
ASTV	0
MSTV	0
ALTV	0
MLTV	0
DL	0
DS	0
DP	0
DR	0
Width	0
Min	0
Max	0
Nmax	0
Nzeros	0
Mode	0
Mean	0
Median	0
Variance	0
Tendency	0

Aprendizaje supervisado o no supervisado

Aprendizaje supervisado

- Métodos no paramétricos

En el conjunto de validación

Accuracy	KNN	KNN Balanceado	Tree	Tree Balanceado	MLP	MLP Balanceado
Clase 1	0.951149	0.931034	0.959770	0.925287	0.885057	0.764368
Clase 2	0.693548	0.988506	0.645161	0.939655	0.419355	0.781609
Clase 3	0.729730	0.994236	0.729730	0.965418	0.216216	0.314121

Resultados

Filter vs Embedded vs PCA - Árbol de clasificación

En el conjunto de validación

Accuracy Tree	23D	Filter 17D	IM 13D	Embebbed 12D	PCA 9D
Clase 1	0.925287	0.933908	0.919540	0.922414	0.856322
Clase 2	0.939655	0.925287	0.925287	0.922414	0.910920
Clase 3	0.965418	0.968300	0.968300	0.951009	0.933718

MODELO ELEGIDO ARBOL:
INFORMACION MUTUA 13D

Resultados

Filter vs Embedded vs PCA - KNN

En el conjunto de validación

Accuracy kNN	23D	Filter 17D	IM 13D	Embebbed 12D	PCA 9D
Clase 1	0.931034	0.945402	0.942529	0.928161	0.910920
Clase 2	0.988506	0.985632	0.925287	0.985632	0.982759
Clase 3	0.994236	0.997118	0.988473	0.997118	0.994236

MODELO ELEGIDO KNN FILTER 17D

Discusión de resultados

Evaluación de la calidad del mejor modelo

Accuracy kNN	23D
Clase 1	0.951710
Clase 2	0.704545
Clase 3	0.792453

Clase 1: feto normal
Clase 2: feto sospechosos
Clase 3: feto patológico

Modelo elegido: kNN