Detección de anomalías en Series Temporales mediante uso de técnicas Deep Learning

Ignacio Aguilera Martos

10 Septiembre 2020

Trabajo Fin de Máster
Código disponible en GitHub

Contenidos

Contenido teórico

Contenido teórico de la memoria

Contenido teórico de la memoria

Contenido teórico tratado

1. Machine Learning: aprendizaje, regularización y cotas del aprendizaje.

Contenido teórico de la memoria

Contenido teórico tratado

1. Machine Learning: aprendizaje, regularización y cotas del aprendizaje.
2. Estadística Multivariante: independencia, probabilidad y esperanza condicionadas, desigualdades famosas.

Contenido teórico de la memoria

Contenido teórico tratado

1. Machine Learning: aprendizaje, regularización y cotas del aprendizaje.
2. Estadística Multivariante: independencia, probabilidad y esperanza condicionadas, desigualdades famosas.
3. Redes Neuronales y Deep Learning: aprendizaje, capas empleadas y Autoencoders.

Concepto de anomalía

Concepto de anomalía

Tukey's Fences

Valores fuera del rango $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ con k =1.5, 3, 5

No podemos aplicarlo directamente a varias dimensiones.

Concepto de anomalía

Tukey's Fences

Valores fuera del rango $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ con k =1.5, 3, 5

No podemos aplicarlo directamente a varias dimensiones.

Criterio de clusters

1. Agrupamos los datos por clusters.

Concepto de anomalía

Tukey's Fences

Valores fuera del rango $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ con k =1.5, 3, 5

No podemos aplicarlo directamente a varias dimensiones.

Criterio de clusters

1. Agrupamos los datos por clusters.
2. Encontramos el cluster más cercano para cada instancia.

Concepto de anomalía

Tukey's Fences

Valores fuera del rango $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ con k =1.5, 3, 5

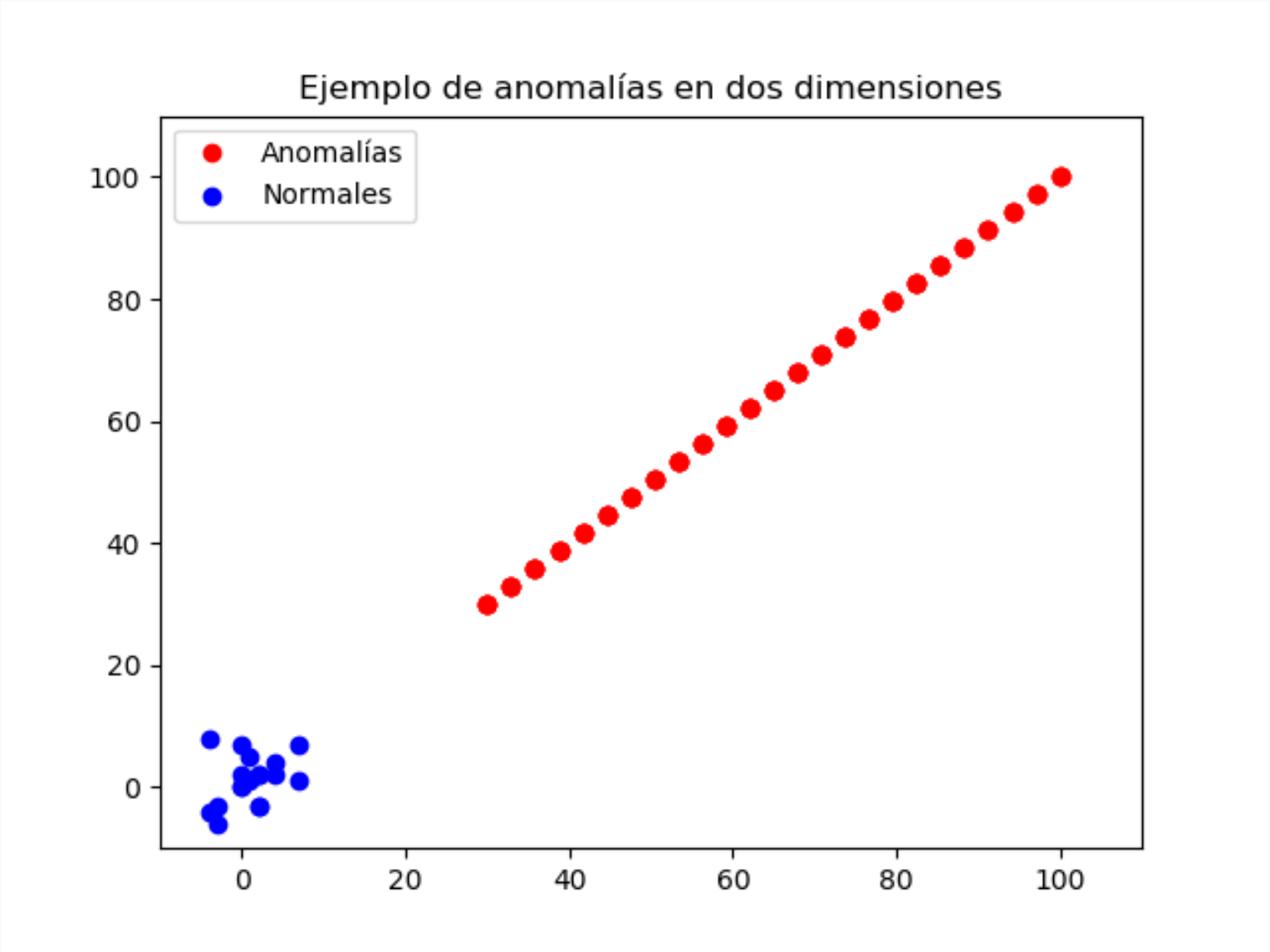
No podemos aplicarlo directamente a varias dimensiones.

Criterio de clusters

1. Agrupamos los datos por clusters.
2. Encontramos el cluster más cercano para cada instancia.
3. Si la distancia del objeto al centroide del cluster es mayor que 1,5 veces la mayor distancia intercluster entonces es una anomalía.

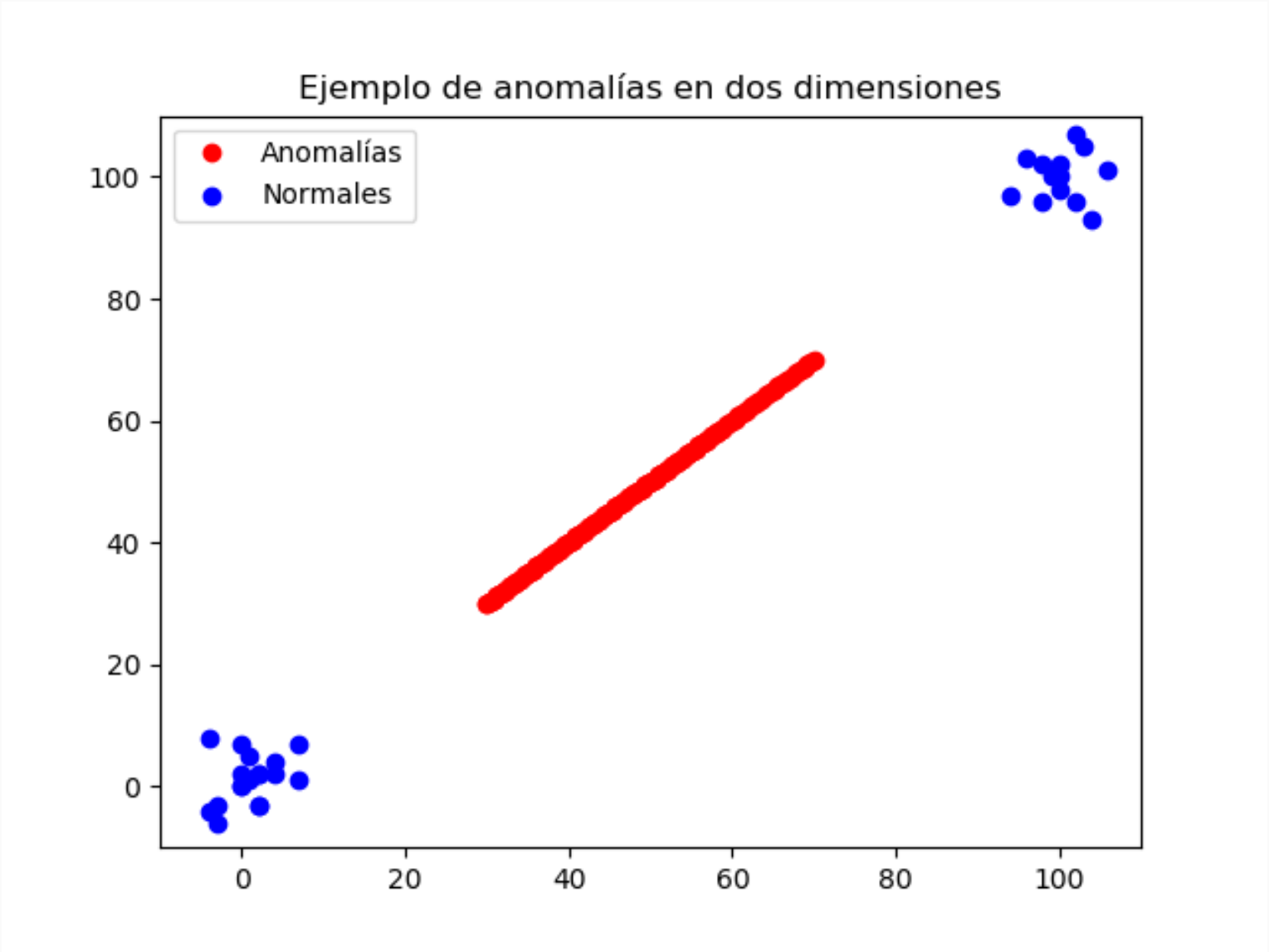
Ejemplo 1

Ejemplo 1


Ejemplo de anomalías en dos dimensiones
Anomalías
Normales
100
80
60
40
20
0
0
20
40
60
80
100

Ejemplo 2

Ejemplo 2


Ejemplo de anomalías en dos dimensiones
Anomalías
Normales
100
80
60
40
20
0
0
20
40
60
80
100

Notación

Notación

Notación usada

$$X = \{x_1, ..., x_n\},\ x_i = (x_{s_1}, ..., x_{s_d})$$

Notación

Notación usada

$$X = \{x_1, ..., x_n\},\ x_i = (x_{s_1}, ..., x_{s_d})$$

$$S = \{s_i | s_i \in \{s_1, ..., s_d\}\ con\ i \in \Delta\}$$

Notación

Notación usada

$$X = \{x_1, ..., x_n\},\ x_i = (x_{s_1}, ..., x_{s_d})$$

$$S = \{s_i | s_i \in \{s_1, ..., s_d\}\ con\ i \in \Delta\}$$

X_S *proyección de los datos en el subespacio* S

Notación

Notación usada

$$X = \{x_1, ..., x_n\},\ x_i = (x_{s_1}, ..., x_{s_d})$$

$$S = \{s_i | s_i \in \{s_1, ..., s_d\}\ con\ i \in \Delta\}$$

X_S *proyección de los datos en el subespacio* S

$$p_{s_1,...,s_p}(x_{s_1}, ..., x_{s_p})$$

Notación

Notación usada

$$X = \{x_1, ..., x_n\},\ x_i = (x_{s_1}, ..., x_{s_d})$$

$$S = \{s_i | s_i \in \{s_1, ..., s_d\}\ con\ i \in \Delta\}$$

X_S *proyección de los datos en el subespacio* S

$$p_{s_1,...,s_p}(x_{s_1}, ..., x_{s_p})$$

$$p_{s_i}(x_{s_i})$$

Definiciones

Definiciones

Definición subespacio incorrelado

Decimos que un subespacio S es un subespacio incorrelado si y sólo si:

$$p_{s_1,\ldots,s_p}(x_{s_1}, \ldots, x_{s_p}) = \prod_{i=1}^{p} p_{s_i}(x_{s_i})$$

Definiciones

Definición subespacio incorrelado

Decimos que un subespacio S es un subespacio incorrelado si y sólo si:

$$p_{s_1,\ldots,s_p}(x_{s_1},\ldots,x_{s_p}) = \prod_{i=1}^{p} p_{s_i}(x_{s_i})$$

Definición anomalía no trivial

Decimos que un objeto x_S es una anomalía no trivial respecto del subespacio S si:

$$p_{s_1,\ldots,s_p}(x_{s_1},\ldots,x_{s_p}) \ll p_{esp}(x_{s_1},\ldots,x_{s_p})$$

Definiciones

Definición subespacio incorrelado

Decimos que un subespacio S es un subespacio incorrelado si y sólo si:

$$p_{s_1,\ldots,s_p}(x_{s_1},\ldots,x_{s_p}) = \prod_{i=1}^{p} p_{s_i}(x_{s_i})$$

Definición anomalía no trivial

Decimos que un objeto x_S es una anomalía no trivial respecto del subespacio S si:

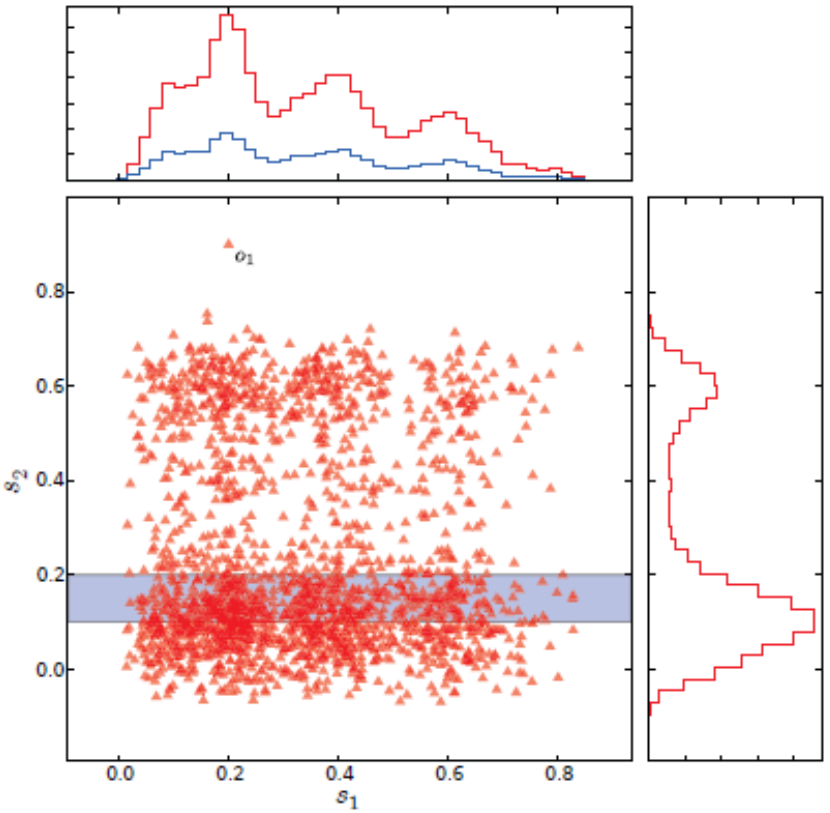
$$p_{s_1,\ldots,s_p}(x_{s_1},\ldots,x_{s_p}) \ll p_{esp}(x_{s_1},\ldots,x_{s_p})$$

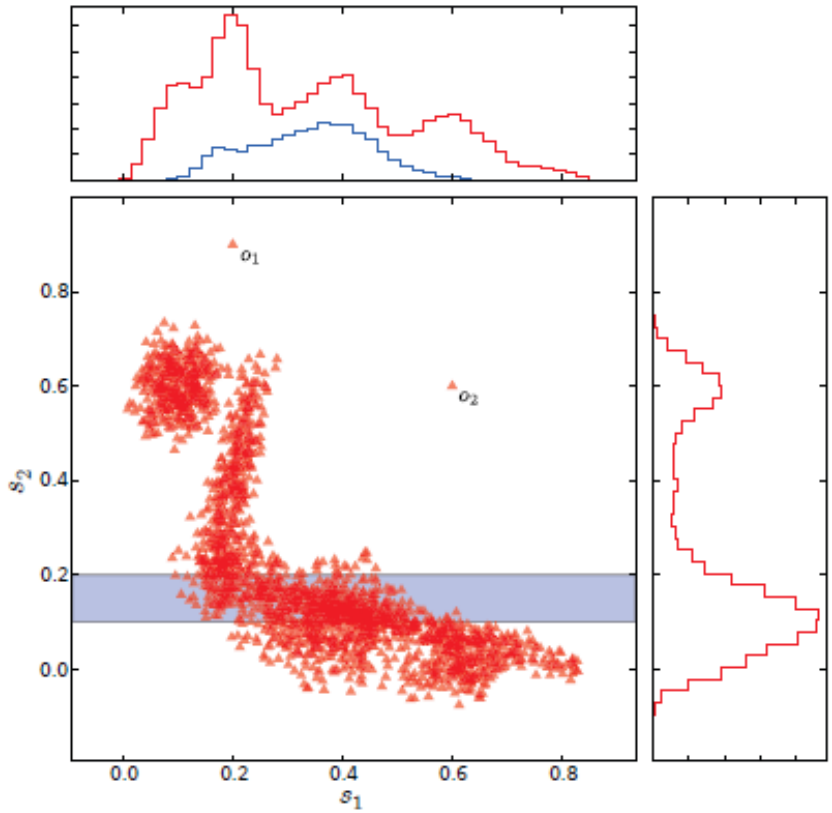
Relación entre conceptos de anomalía

Este concepto de anomalía es complementario.

Ejemplo de anomalía probabilística

Ejemplo de anomalía probabilística





Explicación del problema

Explicación del problema

Explicación del problema

Definición del problema

La empresa ArcelorMittal tiene una máquina sobre la que quiere detectar los mantenimientos antes de que ocurran.

Tenemos una serie temporal de 468 días con medidas cada un segundo y 106 variables en total.

Empleamos 438 como entrenamiento, 30 para hallar cotas y 30 para test, dejando más de 2.5 millones de instancias en test.

Explicación del problema

Definición del problema

La empresa ArcelorMittal tiene una máquina sobre la que quiere detectar los mantenimientos antes de que ocurran.

Tenemos una serie temporal de 468 días con medidas cada un segundo y 106 variables en total.

Empleamos 438 como entrenamiento, 30 para hallar cotas y 30 para test, dejando más de 2.5 millones de instancias en test.

Metodología

Calculamos las puntuaciones con un método de detección de anomalías.

Explicación del problema

Definición del problema

La empresa ArcelorMittal tiene una máquina sobre la que quiere detectar los mantenimientos antes de que ocurran.

Tenemos una serie temporal de 468 días con medidas cada un segundo y 106 variables en total.

Empleamos 438 como entrenamiento, 30 para hallar cotas y 30 para test, dejando más de 2.5 millones de instancias en test.

Metodología

Calculamos las puntuaciones con un método de detección de anomalías.

Pasamos una ventana deslizante y calculamos una cota local, hacemos la media con la cota general y comprobamos si la puntuación es mayor que la combinación de las cotas.

Explicación del problema

Definición del problema

La empresa ArcelorMittal tiene una máquina sobre la que quiere detectar los mantenimientos antes de que ocurran.

Tenemos una serie temporal de 468 días con medidas cada un segundo y 106 variables en total.

Empleamos 438 como entrenamiento, 30 para hallar cotas y 30 para test, dejando más de 2.5 millones de instancias en test.

Metodología

Calculamos las puntuaciones con un método de detección de anomalías.

Pasamos una ventana deslizante y calculamos una cota local, hacemos la media con la cota general y comprobamos si la puntuación es mayor que la combinación de las cotas.

Sobre los datos clasificados pasamos ventanas deslizantes y comprobamos si hay más de un 5 % de anomalías, entonces damos alerta.

Modelos implementados

Modelos de detección de anomalías

Modelos de detección de anomalías

Tipos de modelos

Modelos Autoencoder: codifican y reconstruyen las instancias. A peor reconstrucción más anómalo es el dato.

Modelos de Predicción: aprenden a predecir en condiciones de normalidad. Cuanto peor lo hagan más anómalo será el dato.

Modelos de detección de anomalías

Tipos de modelos

Modelos Autoencoder: codifican y reconstruyen las instancias. A peor reconstrucción más anómalo es el dato.

Modelos de Predicción: aprenden a predecir en condiciones de normalidad. Cuanto peor lo hagan más anómalo será el dato.

Modelos

- Autoencoder totalmente conectado simétrico.

Modelos de detección de anomalías

Tipos de modelos

Modelos Autoencoder: codifican y reconstruyen las instancias. A peor reconstrucción más anómalo es el dato.

Modelos de Predicción: aprenden a predecir en condiciones de normalidad. Cuanto peor lo hagan más anómalo será el dato.

Modelos

- Autoencoder totalmente conectado simétrico.
- Autoencoder totalmente conectado simétrico por lotes.

Modelos de detección de anomalías

Tipos de modelos

Modelos Autoencoder: codifican y reconstruyen las instancias. A peor reconstrucción más anómalo es el dato.

Modelos de Predicción: aprenden a predecir en condiciones de normalidad. Cuanto peor lo hagan más anómalo será el dato.

Modelos

- Autoencoder totalmente conectado simétrico.
- Autoencoder totalmente conectado simétrico por lotes.
- Autoencoder con capas LSTM simétrico.

Modelos de detección de anomalías

Tipos de modelos

Modelos Autoencoder: codifican y reconstruyen las instancias. A peor reconstrucción más anómalo es el dato.

Modelos de Predicción: aprenden a predecir en condiciones de normalidad. Cuanto peor lo hagan más anómalo será el dato.

Modelos

- Autoencoder totalmente conectado simétrico.
- Autoencoder totalmente conectado simétrico por lotes.
- Autoencoder con capas LSTM simétrico.
- Predicción con capas LSTM.

Modelos de detección de anomalías

Tipos de modelos

Modelos Autoencoder: codifican y reconstruyen las instancias. A peor reconstrucción más anómalo es el dato.

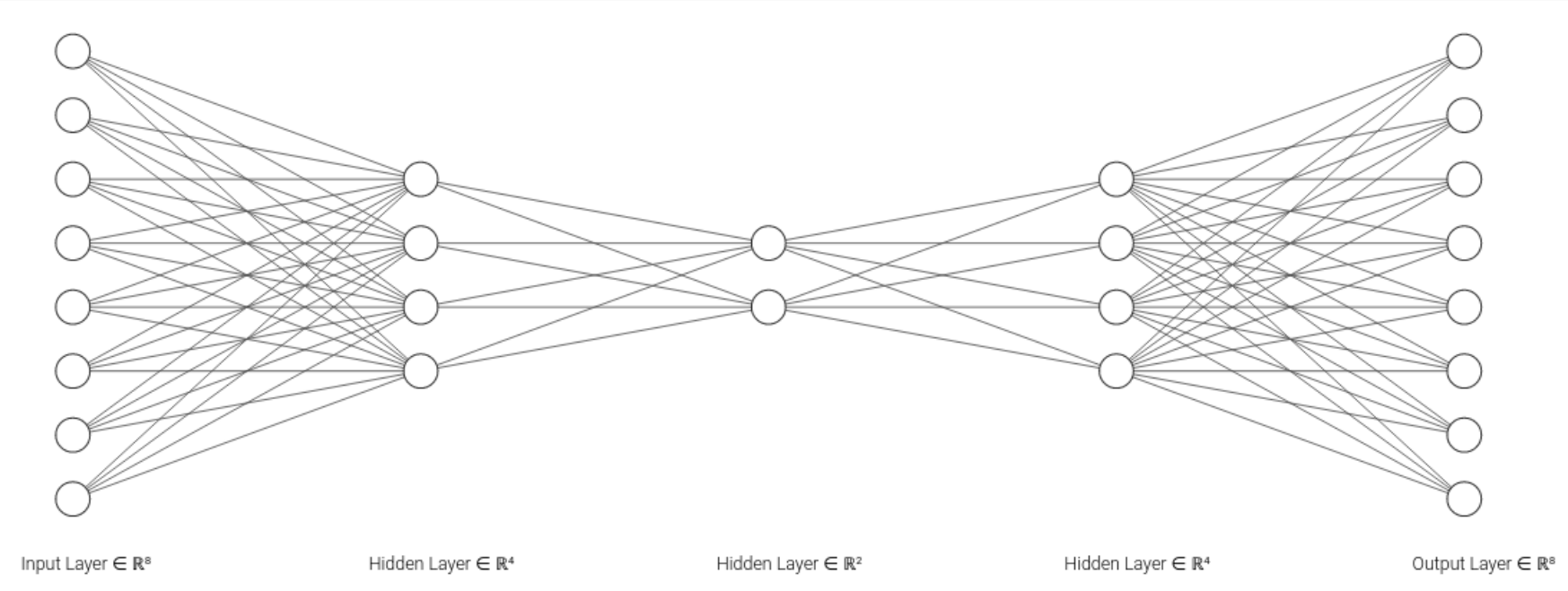
Modelos de Predicción: aprenden a predecir en condiciones de normalidad. Cuanto peor lo hagan más anómalo será el dato.

Modelos

- Autoencoder totalmente conectado simétrico.
- Autoencoder totalmente conectado simétrico por lotes.
- Autoencoder con capas LSTM simétrico.
- Predicción con capas LSTM.
- Predicción con capas CNN-LSTM.

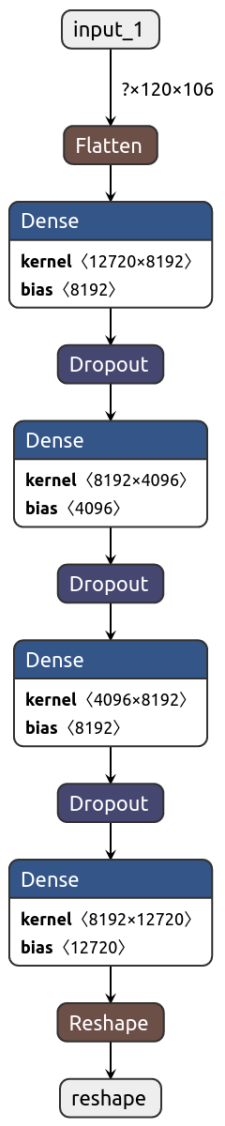
Autoencoder totalmente conectado

Autoencoder totalmente conectado



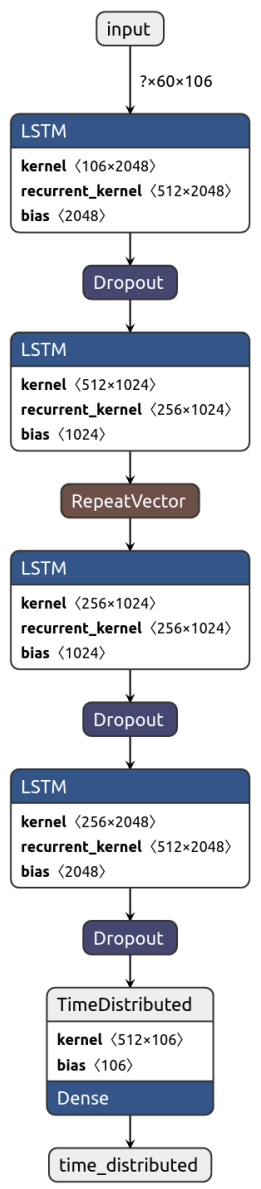
Autoencoder totalmente conectado por lotes

Autoencoder totalmente conectado por lotes



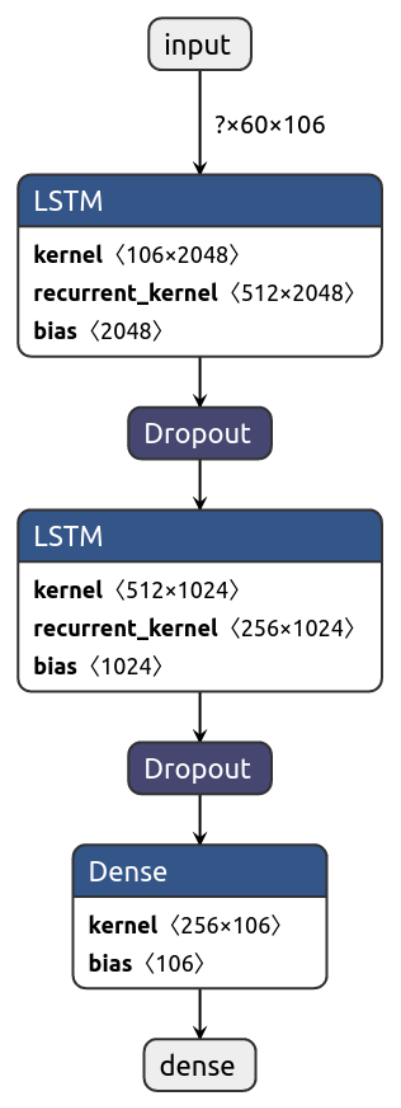
Autoencoder LSTM

Autoencoder LSTM



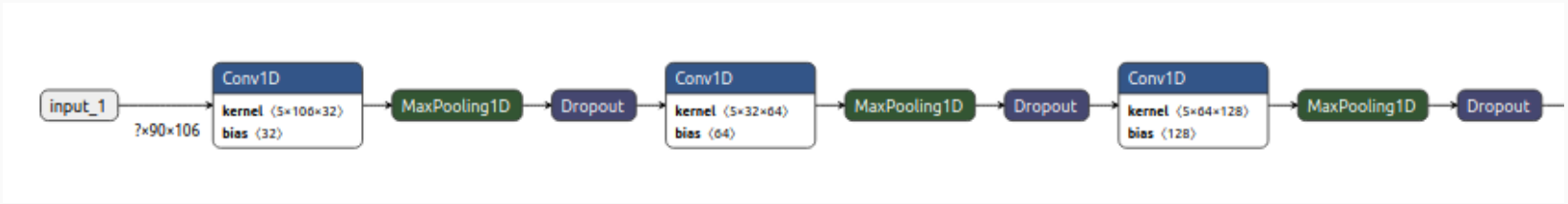
Predictor LSTM

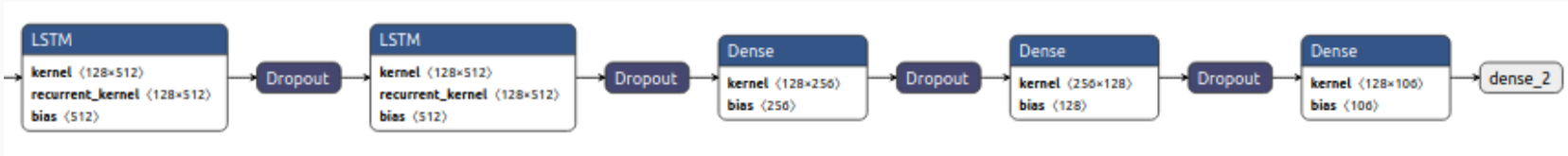
Predictor LSTM



Predictor CNN-LSTM

Predictor CNN-LSTM





Resultados obtenidos

Resultados iniciales

Resultados iniciales

Modelo	TP	FP	TN	FN	M.D.	M.T.
AE FCC	33	441	75	8	20	28
AE FCC lotes	33	393	123	8	20	28
AE LSTM	33	447	69	8	20	28
Predictor LSTM	33	393	123	8	20	28
Predictor CNN-LSTM	35	213	301	8	22	28
FB LOF	15	**56**	**340**	20	8	28
HBOS	**44**	505	0	**5**	**23**	28
IForest	28	344	177	9	19	28
KNN	42	505	1	**5**	**23**	28
LODA	36	341	172	6	22	28
LOF	40	471	38	6	22	28
PCA	26	256	267	12	16	28

Porcentajes de acierto

Porcentajes de acierto

Modelo	Tasa mantenimientos detectados	Acierto sobre ventanas
AE FCC	0.7143	0.1939
AE FCC lotes	0.7143	0.2801
AE LSTM	0.7143	0.1831
Predictor LSTM	0.7143	0.2801
Predictor CNN-LSTM	0.7857	0.6054
FB LOF	0.2857	**0.8237**
HBOS	**0.8214**	0.0794
IForest	0.6786	0.3674
KNN	**0.8214**	0.0794
LODA	0.7857	0.3748
LOF	0.7857	0.1405
PCA	0.5714	0.5223

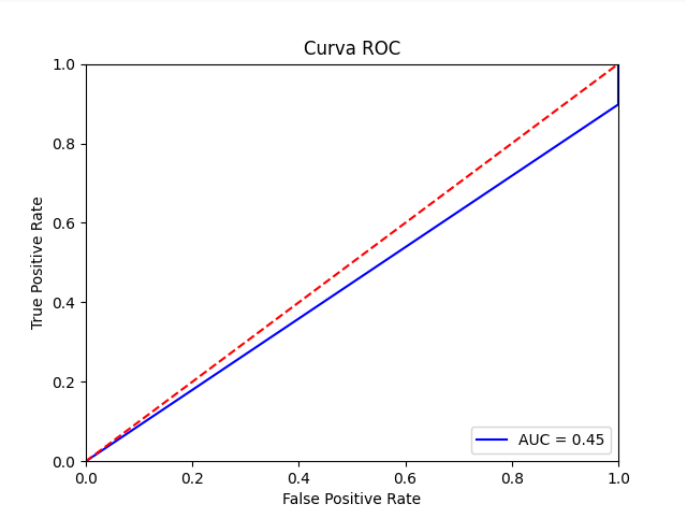
TPR, TNR, F1 y AUC

TPR, TNR, F1 y AUC

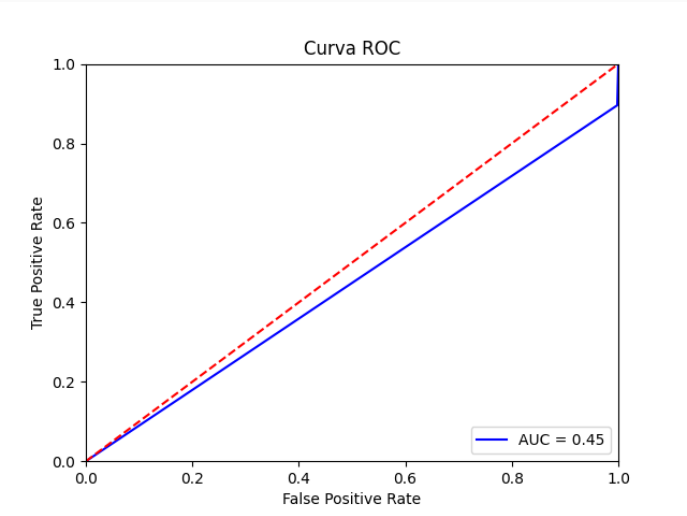
Modelo	TPR	TNR	TPRxTNR	F1	AUC
AE FCC	0.8049	0.1453	0.1170	0.1282	0.48
AE FCC lotes	0.8049	0.2384	0.1919	0.1413	0.52
AE LSTM	0.8049	0.1337	0.1076	0.1267	0.47
Pred. LSTM	0.8049	0.2384	0.1919	0.1413	0.52
Pred. CNN-LSTM	0.8537	0.5856	**0.5**	0.2422	**0.72**
FB LOF	0.4286	**0.8586**	0.3680	**0.2830**	0.64
HBOS	**0.8980**	0	0	0.1471	0.45
IForest	0.7568	0.3397	0.2571	0.1369	0.55
KNN	0.8959	0.0020	0.0018	0.1443	0.45
LODA	0.8571	0.3353	0.2874	0.1718	0.6
LOF	0.8696	0.0747	0.0649	0.1436	0.47
PCA	0.6842	0.5105	0.3493	0.1625	0.6

ROC de los mejores modelos

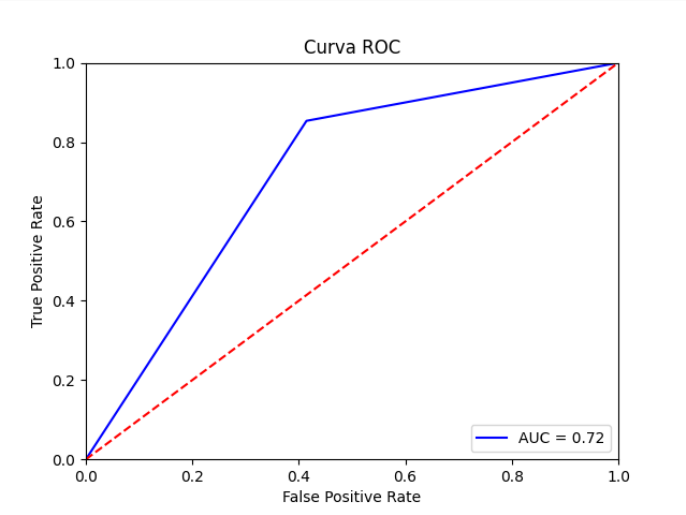
ROC de los mejores modelos



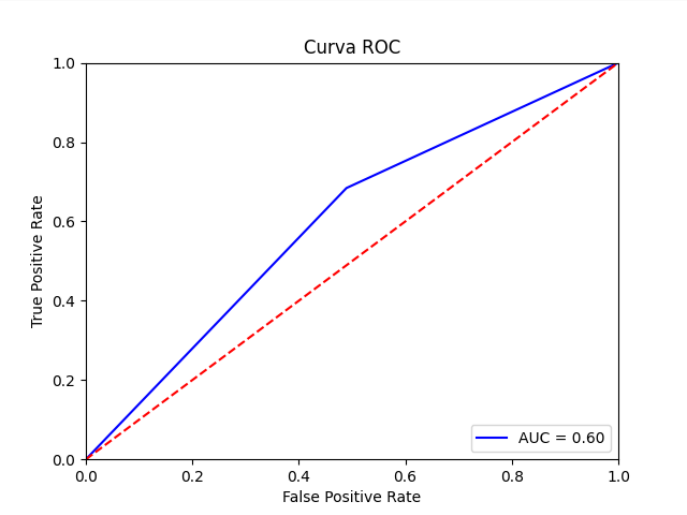
(a) Curva ROC HBOS.



(b) Curva ROC KNN.



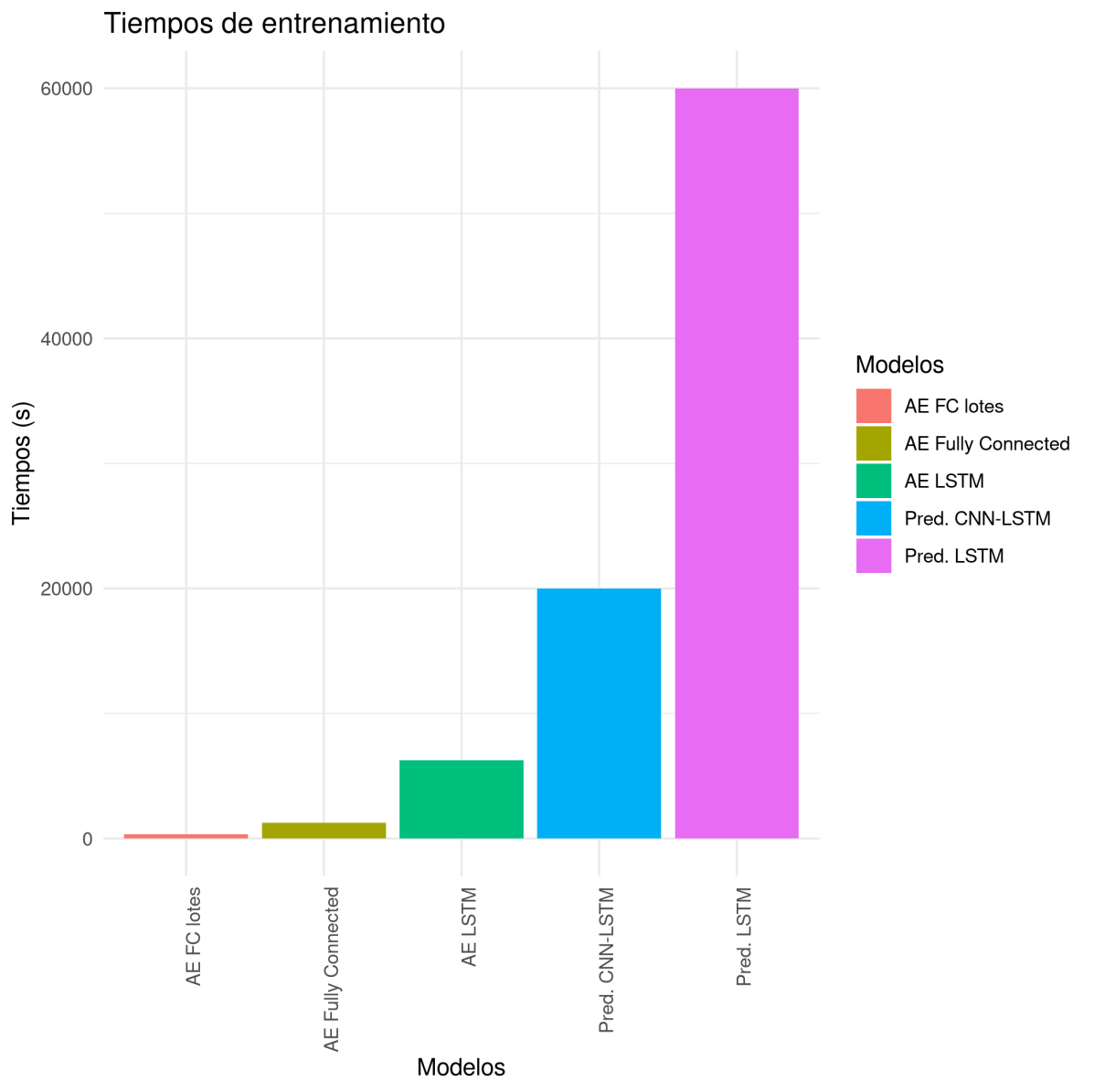
(a) Curva ROC Predictor CNN-LSTM.



(b) Curva ROC PCA.

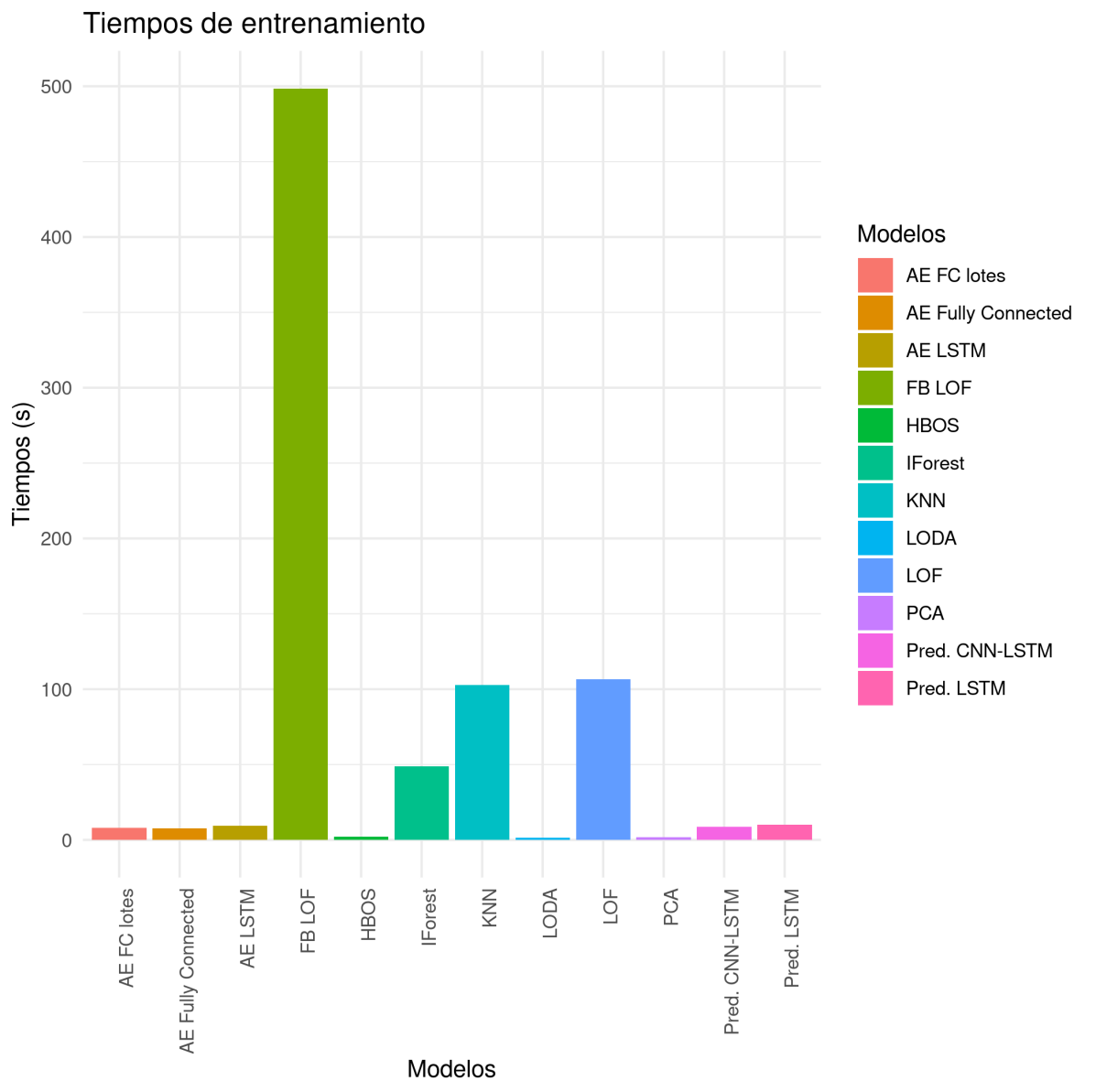
Tiempos entrenamiento

Tiempos entrenamiento



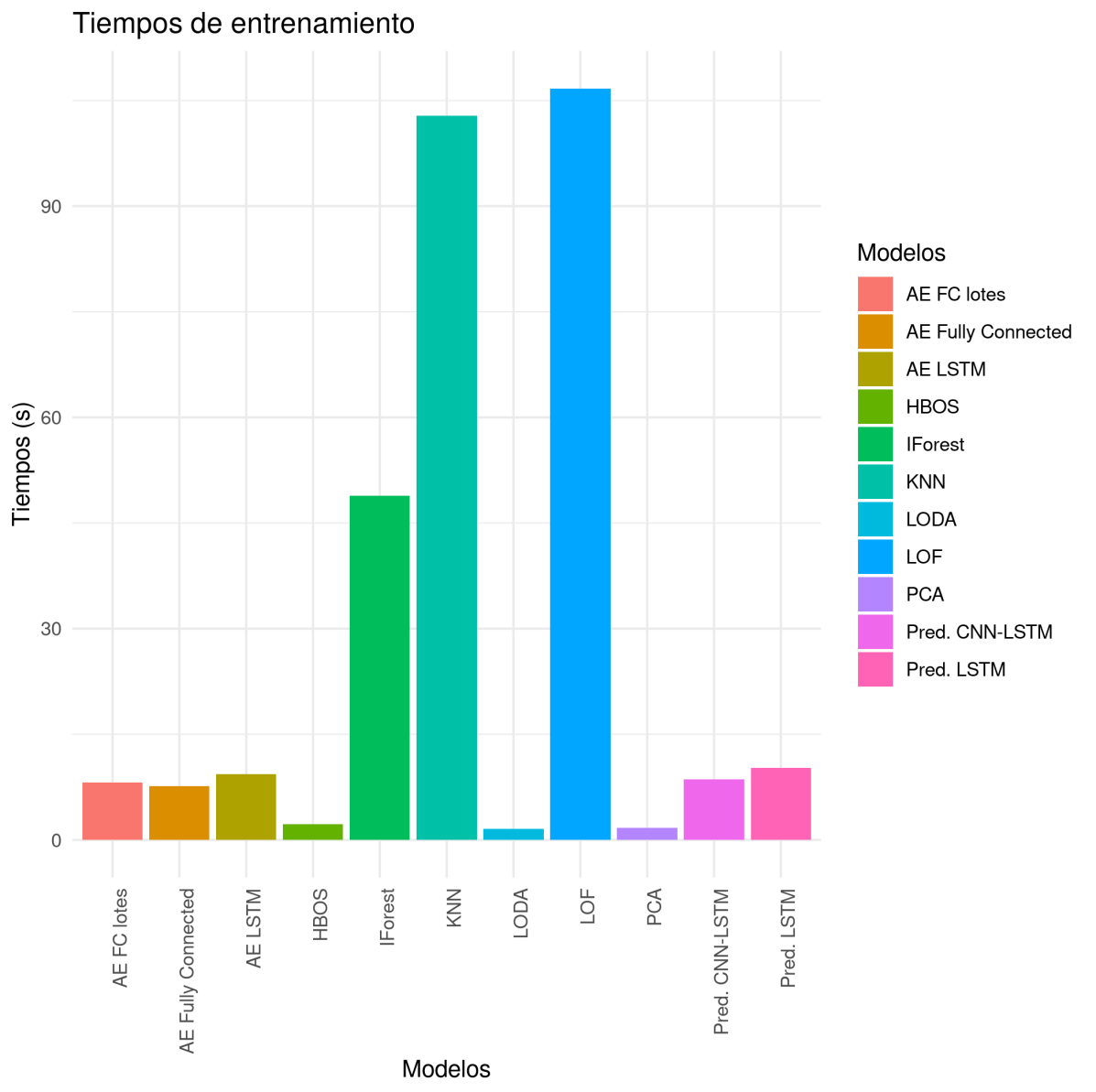
Tiempos de predicción

Tiempos de predicción



Tiempos de predicción

Tiempos de predicción



Conclusiones y Trabajo Futuro

Conclusiones

Conclusiones

- El modelo más equilibrado y el mejor es el modelo de predicción CNN-LSTM

Conclusiones

- El modelo más equilibrado y el mejor es el modelo de predicción CNN-LSTM
- Los modelos Deep Learning han demostrado gran potencial en la detección de anomalías.

Conclusiones

- El modelo más equilibrado y el mejor es el modelo de predicción CNN-LSTM
- Los modelos Deep Learning han demostrado gran potencial en la detección de anomalías.
- Los modelos HBOS y KNN han dado buenos resultados como modelos clásicos.

Conclusiones

- El modelo más equilibrado y el mejor es el modelo de predicción CNN-LSTM
- Los modelos Deep Learning han demostrado gran potencial en la detección de anomalías.
- Los modelos HBOS y KNN han dado buenos resultados como modelos clásicos.
- Los modelos de predicción resultan mejores que los Autoencoder.

Conclusiones

- El modelo más equilibrado y el mejor es el modelo de predicción CNN-LSTM
- Los modelos Deep Learning han demostrado gran potencial en la detección de anomalías.
- Los modelos HBOS y KNN han dado buenos resultados como modelos clásicos.
- Los modelos de predicción resultan mejores que los Autoencoder.
- El algoritmo de clasificación de anomalías y detección de eventos anómalos es crucial.

Conclusiones

- El modelo más equilibrado y el mejor es el modelo de predicción CNN-LSTM
- Los modelos Deep Learning han demostrado gran potencial en la detección de anomalías.
- Los modelos HBOS y KNN han dado buenos resultados como modelos clásicos.
- Los modelos de predicción resultan mejores que los Autoencoder.
- El algoritmo de clasificación de anomalías y detección de eventos anómalos es crucial.
- Utilidad del estudio en su aplicación en el mundo real.

Trabajo Futuro

Trabajo Futuro

- Empleo de técnicas de reducción de falsos positivos.

Trabajo Futuro

- Empleo de técnicas de reducción de falsos positivos.
- Búsqueda de los mejores parámetros para los modelos.

Trabajo Futuro

- Empleo de técnicas de reducción de falsos positivos.
- Búsqueda de los mejores parámetros para los modelos.
- Añadir más arquitecturas Deep Learning a la comparativa.

Trabajo Futuro

- Empleo de técnicas de reducción de falsos positivos.
- Búsqueda de los mejores parámetros para los modelos.
- Añadir más arquitecturas Deep Learning a la comparativa.
- Implementación de una arquitectura propia.

Trabajo Futuro

- Empleo de técnicas de reducción de falsos positivos.
- Búsqueda de los mejores parámetros para los modelos.
- Añadir más arquitecturas Deep Learning a la comparativa.
- Implementación de una arquitectura propia.
- Implementar la metodología Test-Then-Train

Trabajo Futuro

- Empleo de técnicas de reducción de falsos positivos.
- Búsqueda de los mejores parámetros para los modelos.
- Añadir más arquitecturas Deep Learning a la comparativa.
- Implementación de una arquitectura propia.
- Implementar la metodología Test-Then-Train
- Unir los TP, FP, TN y FN para analizarlos temporalmente.

Gracias por su atención.

¿Preguntas?