

Implementación de una feed forward network para predecir si una persona es propensa a padecer de problemas del corazón

José Alejandro López Quel
Universidad Galileo



Introducción

En este paper, se presenta un caso de predicción de la posibilidad de llegar a tener problemas del corazón. Al determinar si una persona es propensa a padecer de problemas del corazón hay factores importantes como la presión arterial, uso de sustancias nocivas como alcohol y tabaco, obesidad, entre otras, los cuales pueden ayudar a predecir la existencia de futuras enfermedades. Por lo que en este trabajo, se analiza utilizar los datos antes mencionados para predecir estos comportamientos empleando una red neuronal del tipo feed forward.

Metodología

La utilidad de un buen modelo de predicción de problemas del corazón depende en gran medida de su precisión y la estabilidad. Para lograrlo, se divide la investigación en tres experimentos, en cada uno se emplea distintas capas para obtener las predicciones más adecuadas y se evalúa tanto para el dataset de entrenamiento como para el dataset de validación.

Model: "Experimento_1"		
Layer (type)	Output Shape	Param #
=====		
dense_96 (Dense)	(None, 12)	120
dense_97 (Dense)	(None, 9)	117
dense_98 (Dense)	(None, 1)	10
=====		
Total params: 247		
Trainable params: 247		
Non-trainable params: 0		

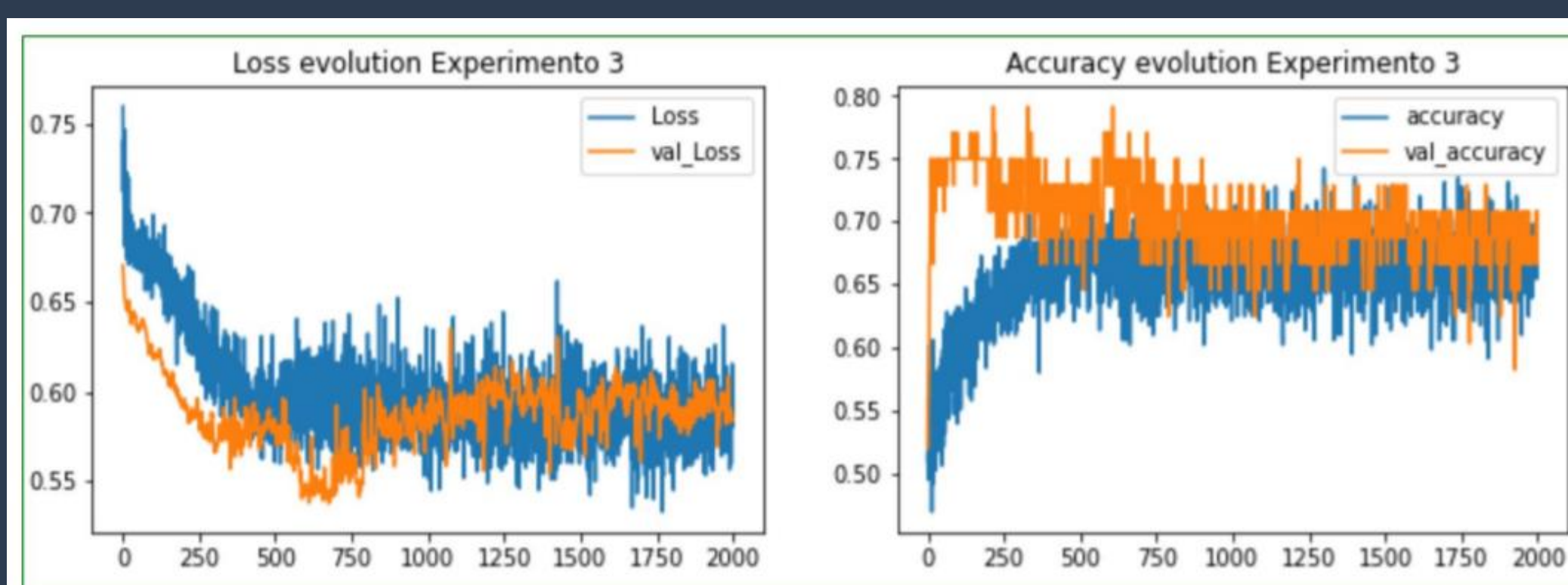
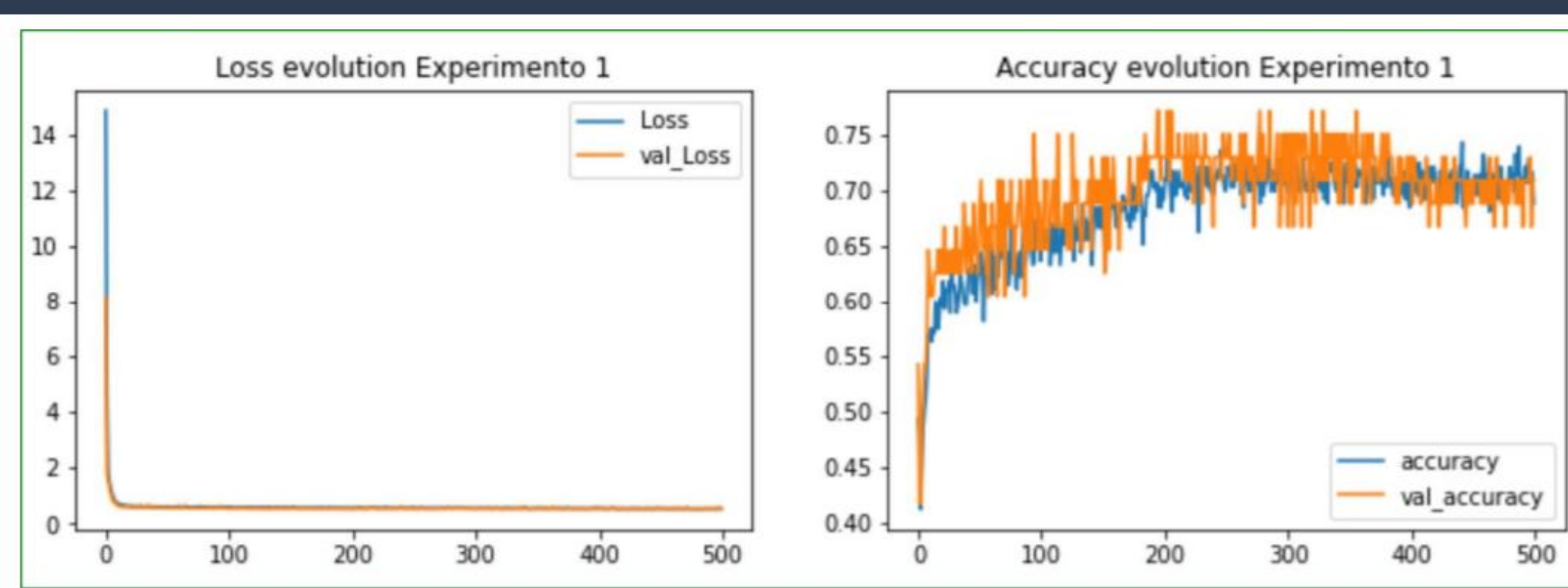
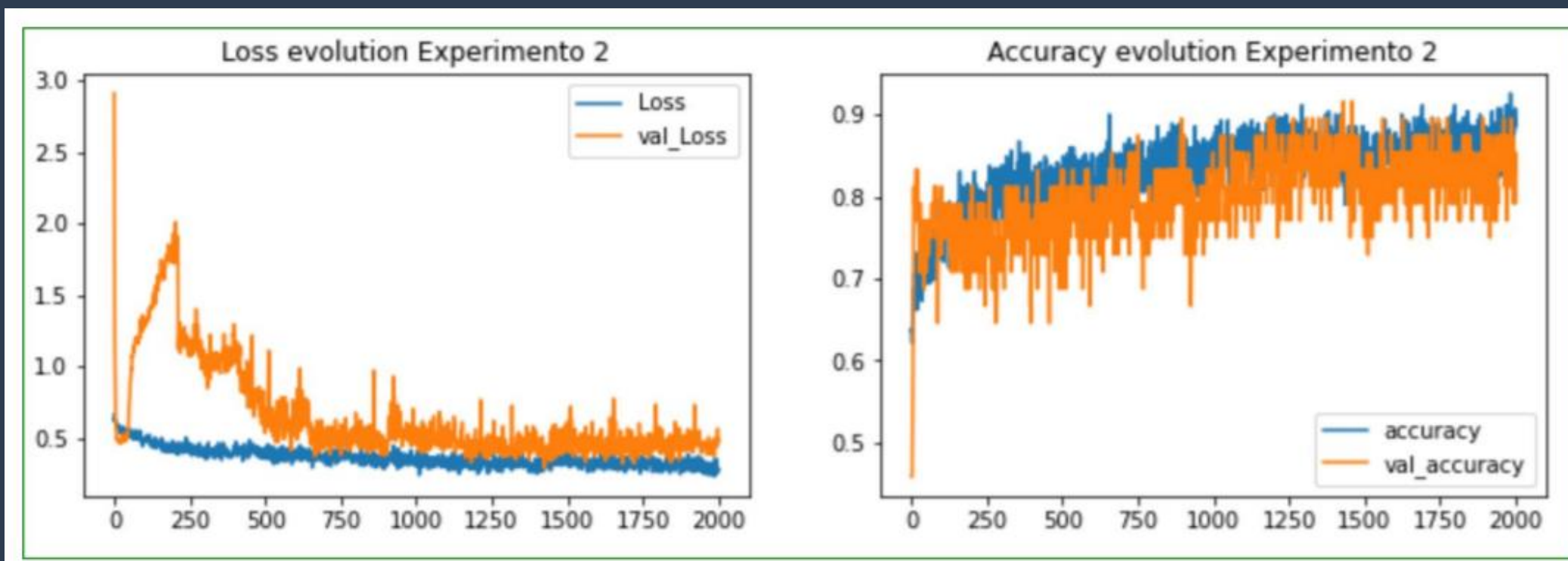
Model: "Experimento_2"		
Layer (type)	Output Shape	Param #
=====		
dense_99 (Dense)	(None, 64)	640
batch_normalization_32 (Batch Normalization)	(None, 64)	256
dense_100 (Dense)	(None, 32)	2080
dropout_30 (Dropout)	(None, 32)	0
dense_101 (Dense)	(None, 1)	33
=====		
Total params: 3,009		
Trainable params: 2,881		
Non-trainable params: 128		

Resultados

Se obtiene que el modelo que mejor se desempeña es el del Experimento 2, el cual es el que obtiene un mayor valor de accuracy tanto para los datos de entrenamiento como con los datos de validación.

Se puede observar en las gráficas siguientes el comportamiento del aprendizaje para cada uno de los modelos.

Model: "Experimento_3"		
Layer (type)	Output Shape	Param #
=====		
dense_102 (Dense)	(None, 16)	160
batch_normalization_33 (Batch Normalization)	(None, 16)	64
dropout_31 (Dropout)	(None, 16)	0
dense_103 (Dense)	(None, 8)	136
dropout_32 (Dropout)	(None, 8)	0
dense_104 (Dense)	(None, 2)	18
batch_normalization_34 (Batch Normalization)	(None, 2)	8
dense_105 (Dense)	(None, 1)	3
=====		
Total params: 389		
Trainable params: 353		
Non-trainable params: 36		



Conclusiones

Los resultados demuestran que, con datos suficientes y características clínicas seleccionadas, las técnicas de aprendizaje automático son capaces de predecir la ocurrencia de eventos de problemas del corazón con un porcentaje de accuracy elevado.

El éxito del aprendizaje automático depende en gran medida de la riqueza de los datos que representan el fenómeno considerado. Aunque el conjunto de datos seleccionado tiene las características y los factores de riesgo más conocidos para predecir enfermedades relacionadas con el corazón, con un conjunto bastante rico de características, más datos y más variables pueden ayudar a mejorar los resultados de la predicción.

Si se dispusiera de otros conjuntos de datos externos con las mismas características y procedentes de distintas regiones, no solo limitado a un país en específico, se podrían llegar a utilizar para validar los resultados obtenidos en este trabajo.

Bibliografía

Feurer M., Klein A., Eggensperger K., Springenberg J.T., Blum M., Hutter F. (2019). Auto-sklearn:Efficient and Robust Automated Machine Learning. In: Hutter F., Kotthoff L., Vanschoren J. (eds) Automated Machine Learning. The Springer Series on Challenges in Machine Learning. Springer, Cham.



	sbp	tobacco	ldl	adiposity	famhist	typea	obesity	alcohol	age	chd
0	160	12.00	5.73	23.11	1	49	25.30	97.20	52	1
1	144	0.01	4.41	28.61	0	55	28.87	2.06	63	1
2	118	0.08	3.48	32.28	1	52	29.14	3.81	46	0
3	170	7.50	6.41	38.03	1	51	31.99	24.26	58	1
4	134	13.60	3.50	27.78	1	60	25.99	57.34	49	1