
Diagnóstico de diabetes con Feed Forward Neural Network

Juan Carlos Castillo López - 22008463

Abstract

1 Diabetes mellitus es un grupo de enfermedades metabólicas en las cuales se presenta
2 un elevado nivel de azúcar en sangre durante periodos largos de tiempo. De
3 acuerdo a la Organización Panamericana de Salud aproximadamente 62 millones
4 de personas en las Américas (422 millones de personas en todo el mundo) tienen
5 diabetes, y 244,084 muertes (1.5 millones en todo el mundo) se atribuyen a la
6 enfermedad cada año. Para su tratamiento es importante una detección en sus
7 primeras etapas, el presente paper presenta un modelo de red neuronal capaz de
8 predecir la enfermedad de acuerdo a factores de riesgo, la red fue construida
9 haciendo uso del dataset PIMA Indian Diabetes.

10 1 Introducción

11 La diabetes es una enfermedad prolongada (crónica) en la cual el cuerpo no puede regular la cantidad
12 de azúcar en sangre.

13 La insulina es una hormona producida por el páncreas para controlar el azúcar en la sangre. La
14 diabetes puede ser causada por muy poca producción de insulina, resistencia a la insulina o ambas.

15 2 Redes neuronales artificiales

16 Una red neuronal artificial (ANN) por sus siglas en ingles, es un modelo de aprendizaje de machine
17 learning basado en la estructura y funcionamiento de una red neural biológica. Una red ANN es
18 considerada una herramienta de modelado de datos estadísticos no lineal en la cual las relaciones
19 complejas entre entradas y salidas son modelas por medio de patrones.

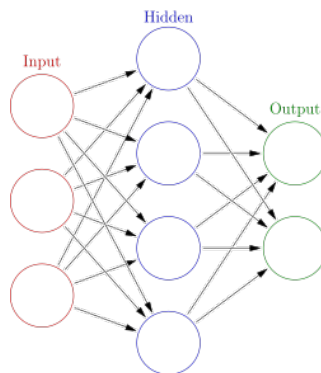


Figure 1: Ejemplo de red neuronal artificial

20 Las redes neuronales artificiales brindan numerosas ventajas, una de ellas es la habilidad para aprender
21 a partir de un conjunto de datos.

22 Las ANN utilizan muestras del conjunto de datos en la fase de entrenamiento para reducir el tiempo y
23 el costo del proceso.

24 2.1 Entrenamiento de una red neuronal

25 El entrenamiento de una red neuronal es un proceso iterativo que inicia con la recolección de los
26 datos, los datos son pre-procesados para eficientar el proceso. Durante el pre procesamiento los datos
27 son divididos en 3 subconjuntos, conjunto de entrenamiento, conjunto de validación y conjunto de
28 pruebas.

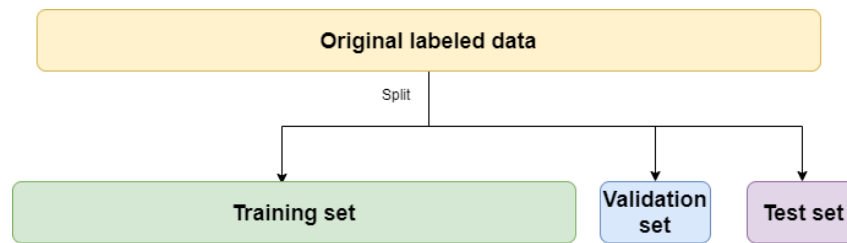


Figure 2: Proceso de división de los datos

29 3 Análisis de datos

30 El dataset utilizado cuenta con la siguiente información:

No.	Factores de riesgos		
	Atributo	Descripción	Rango
1	Pregnancies	Número de embarazos	0-17
2	Glucose	Concentración de glucosa	0-199
3	BloodPressure	Presión sanguínea distolica	0-122
4	SkinThickness	Grosor de pien en triceps	0-99
5	Insulin	Insulina serum 2-horas	0-846
6	BMI	Indicador de indice de masa corporal	0-67.1
7	DiabetesPedigree	Función de probabilidad de diabetes basado en historia familiar	0.078-2.42
8	Age	Edad	21-81
9	Outcome	Variable que indica si el paciente tiene diabetes o no	0-1

Figure 3: Tabla con descripción de atributos

31 4 Modelo

32 Para la definición del modelo se realizo previamente un análisis de la importancia de cada una de las
33 variables presentes en el conjunto de datos.

34 En base a la importancia se utilizaron las siguientes variables dependientes:

- 35 • Glucose
- 36 • BMI
- 37 • DiabetesPedigreeFunction
- 38 • Pregnancies



Figure 4: Importancia de variables

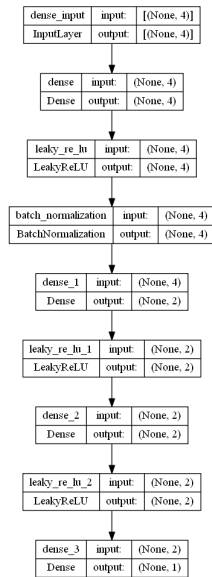


Figure 5: Modelo utilizado para el entrenamiento

5 Conclusiones

Utilizando el modelo de red neuronal artificial se obtuvo una precisión del 81 por ciento en los datos de prueba.