

---

# Statistical Learning II - Proyecto del curso

---

**Henry G. Barrientos G.**  
Universidad Galileo 2021  
Guatemala, C.A.  
henry.barrientos@galileo.edu

## Abstract

El proyecto del curso de Statistical Learning propone varios retos en cuanto a poner en práctica lo aprendido en el curso. El proyecto consiste en implementar tres modelos de los aprendidos en el curso, usando como base datasets de autoría propia o alguno de los que se encuentran en páginas especializadas en datasets. Los modelos propuestos son los que siguen:

- Red neuronal feed forward/MLP: datos estructurados
- Red Neuronal Convolutacional: datos espaciales
- Red Neuronal Recurrente: datos secuenciales/temporales.

## 1 Parte 1 - Multilayer Perceptron

Dataset estructurado con información sobre clientes de tarjetas de crédito. Este conjunto de datos consta de 10,000 clientes que proporcionan edad, salario, estado civil, límite de tarjeta de crédito, categoría de tarjeta de crédito, deserción de cliente, etc. Contiene casi 18 características.

### 1.1 Objetivo

En esta primera parte del proyecto el objetivo es implementar una red neuronal artificial para leer datos estructurados que contienen información acerca de tarjetas de crédito y como objetivo final, indicar si un cliente basado en la información que se tiene podría darse de baja y dejar de ser cliente de una entidad financiera.

### 1.2 Metodología

El método utilizado para lograr el objetivo fue la construcción de una red neuronal que tiene como entrada la información y como salida la indicación de si la persona dejará de ser usuario de la tarjeta de crédito.

## 2 Parte 2 - Convolutional Neural Network

Image set de perros expulsando sus excrementos, recopilación propia de varias fuentes y ampliado mediante copia y tratamiento de las mismas imágenes para generar otras y así tener más con qué trabajar. Basado en una búsqueda previa se confirma que en Internet hay varias imágenes para entrenar.

### 2.1 Objetivo

Leer y entrenar imágenes para que la CNN determine si en la imagen hay o no un perro.

## **2.2 Metodología**

Crear y entrenar una red neuronal convolucional para capturar imágenes y que la CNN determine si en la imagen hay o no un perro. Esto tiene como objetivo determinar si un perro está en posición fecal.

## **3 Parte 3 - Recurrent Neural Network**

El conjunto de datos propuesto tiene un listado de los precios de compra y venta de acciones para varias empresas que son públicas en las bolsas de valores más grandes de Estados Unidos (NYSE, NASDAQ, y NYSE MKT). Se tomará una para entrenar una red neuronal recurrente para determinar un modelo de predicción de precios.

### **3.1 Objetivo**

Con información de precios de acciones de una empresa, estimar valores a futuro de los precios estimados de las acciones.

### **3.2 Metodología**

Crear y entrenar una red neuronal recurrente basado en la arquitectura LSTM que son la mejor opción dentro del ámbito de RNN.

## **A Appendix**

Estos modelos de arquitecturas, tienen una complejidad importante que conviene atacar con un tiempo adecuado.