

Course Overview

Data Mining & Neural Networks

Dr. Wilmer Garzón

Director, Master's Program in Data Science
Department of Computer Engineering

Escuela Colombiana de Ingeniería
Universidade da Coruña

2025



UNIVERSIDAD

Welcome and Course Overview



- Welcome to this exciting journey into **Data Mining** and **Neural Networks**!
- You'll learn how to extract valuable knowledge from large datasets and understand the foundations of neural networks.
- The course is designed for beginners in programming and will be highly practical, using **Python** and **Jupyter Notebooks**.
- Today, we will explore the course goals, structure, learning tools, and an introduction to the software we will be using.



Course Objectives

By the end of the course, you will:

- **Understand** key concepts of data mining (e.g., data cleaning, pattern discovery, clustering)
- **Gain** a strong foundation in neural networks and their applications
- **Apply** Python programming to analyze real datasets
- **Develop** skills using Jupyter Notebooks for reproducible data science
- **Work** on a project integrating data mining and basic deep learning

Introduction Purpose

GETTING TO KNOW YOU



Before we dive deeper into the course content, let's take a few minutes to introduce ourselves. This will help build a **collaborative** and **friendly learning environment**.

Introduction Prompts

Please share the following:

- Your **full name** - **Academic background** (what did you study?)
- Current level of studies (e.g., **undergraduate**, **master's**, **PhD**)
- Where are you from?
- Have you **coded** before? If yes, which **languages**?
- What are your **expectations** for this course?
- (Optional) Any **fun fact** or **personal interest** you'd like to share

Course Structure and Methodology

Course Modules

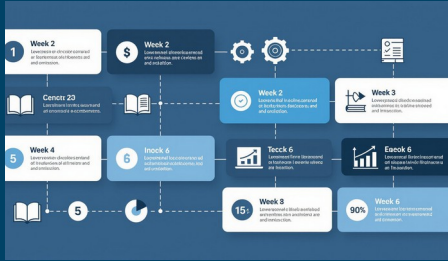
The course is structured into **daily sessions**.

Session Components

Each session includes a **theoretical introduction**, **hands-on coding exercises**, and **discussion** and **collaborative problem-solving**.

Final Project and Resources

The course culminates in a **mini-project** using **real-world data**. Resources will be shared on the **website** and **GitHub**.





Evaluation Criteria

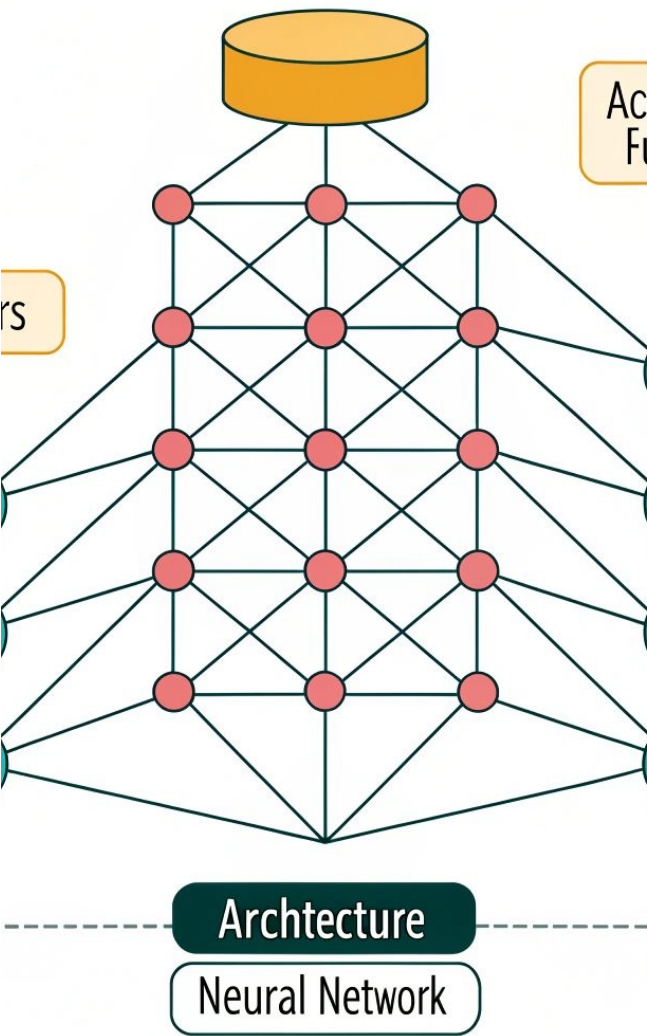
- **Weekly Assignments (30%)** – Short tasks in Python/Jupyter
- **Participation and Quizzes (20%)** – Kahoot-style assessments and code reviews
- **Final Project (50%)** – A small project using a dataset of your choice

Evaluation will focus on effort, creativity, and problem-solving rather than perfection. The goal is to learn by doing.

What is Data Mining?



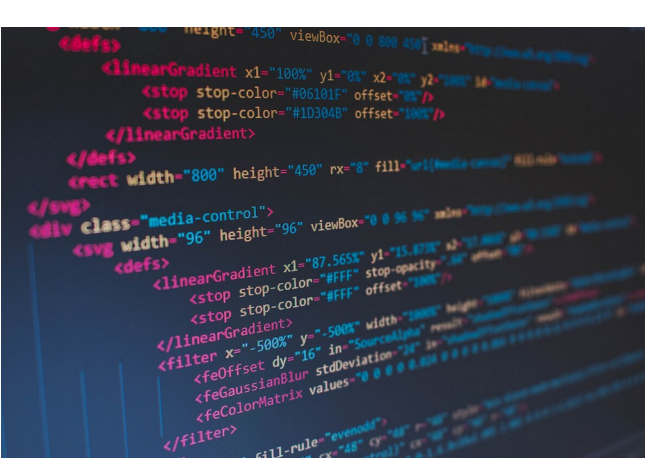
- Data Mining is the process of **discovering patterns, trends, and useful information** from datasets.
- It involves various steps such as **data collection, cleaning, transformation, modeling, and evaluation**.
- Real-world applications include **fraud detection, marketing strategies, customer segmentation,**
- We will explore **classification, clustering, association rules**, and more throughout the course.



What is a Neural Network?

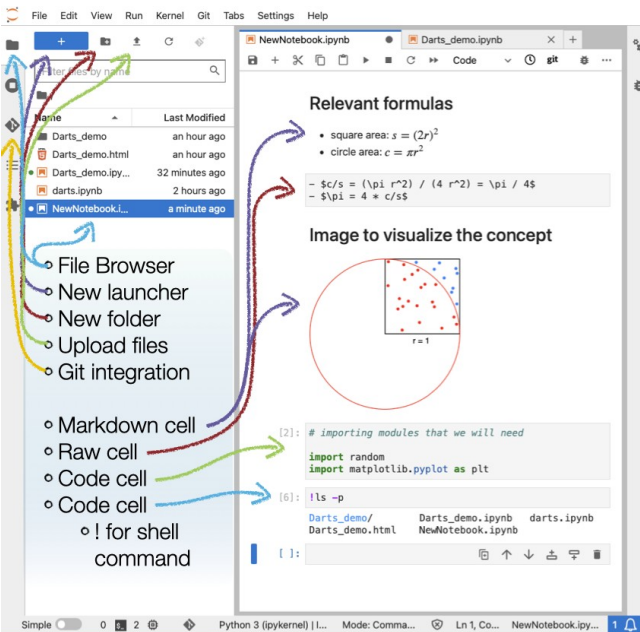
- **Neural networks are inspired by the human brain.**
- They consist of **layers of connected nodes (neurons)** that can learn patterns from data.
- We'll explore the architecture of **neural networks, activation functions, forward and backpropagation, and how neural networks are trained using data.**
- You'll build your first neural network using Python libraries like Keras and PyTorch.

Why Python for Data Mining and Neural Networks?



- **Python** is the most widely used language in data science.
- It is **easy to learn**, has a **clean syntax**, and offers **extensive libraries** for data mining and neural networks:
 - **pandas**, **numpy**, **scikit-learn**, **matplotlib**, and **tensorflow**.
- We'll use Python throughout this course to build everything from simple data pipelines to basic neural networks.

Introduction to Jupyter Notebooks



The screenshot shows the Jupyter Notebook interface with several annotations. On the left, a list of features is shown with arrows pointing to their locations in the interface:

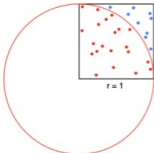
- File Browser
- New launcher
- New folder
- Upload files
- Git integration
- Markdown cell
- Raw cell
- Code cell
- Code cell
 - ! for shell command

The main notebook area displays the following content:

Relevant formulas

- square area: $s = (2r)^2$
- circle area: $c = \pi r^2$

Image to visualize the concept



Code cells:

```
[2]: # importing modules that we will need
import random
import matplotlib.pyplot as plt

[6]: !ls -p
Darts_demo/  Darts_demo.ipynb  darts.ipynb
Darts_demo.html  NewNotebook.ipynb
```

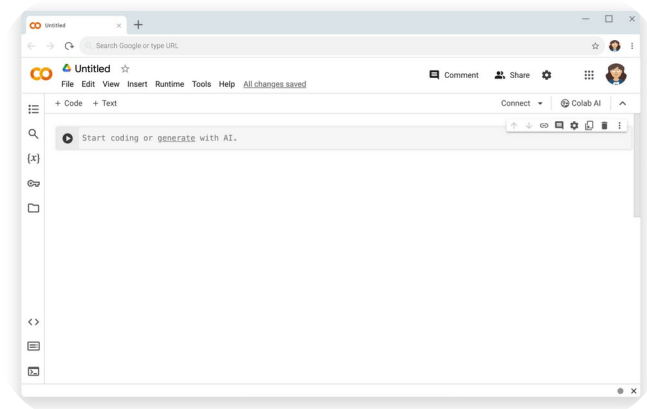
- **Jupyter is an interactive environment** where you can write, document, and run Python code.
- It supports **text, equations, charts, and interactive outputs**. We will use **Jupyter Notebooks for all coding tasks**.
- You'll learn to write **readable code, document your analysis, and share reproducible reports**.
- Let's explore how to install and use Jupyter today.

Installation and Setup

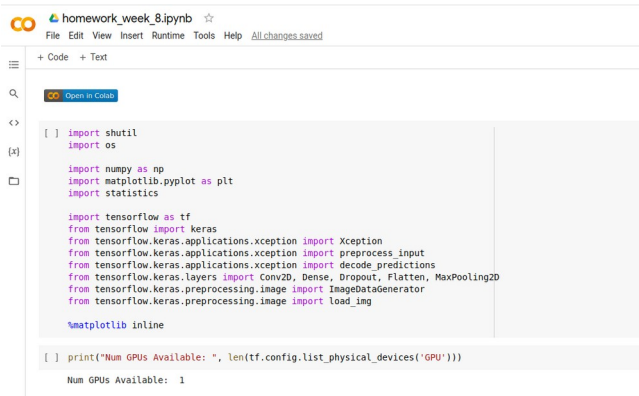
To get started, you can:

- **Use** Google Colab (no installation required)
- **Install** Anaconda (includes Python, Jupyter, and many libraries)

Today, we'll do a live demo using Google Colab. Make sure you have a Gmail account. If you prefer working locally, I'll walk you through installing Anaconda later this week.



First Code Example



```
homework_week_8.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Open in Colab

[ ] import shutil
import os

import numpy as np
import matplotlib.pyplot as plt
import statistics

import tensorflow as tf
from tensorflow.keras.applications.xception import Xception
from tensorflow.keras.applications.xception import preprocess_input
from tensorflow.keras.applications.xception import decode_predictions
from tensorflow.keras.layers import Conv2D, Dense, Dropout, Flatten, MaxPooling2D
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.preprocessing.image import load_img

%matplotlib inline

[ ] print("Num GPUs Available: ", len(tf.config.list_physical_devices('GPU'))

Num GPUs Available: 1
```

Let's open Google Colab and write a simple Python script together:

python print("Welcome to Data Mining and Neural Networks!")

Then we'll do basic operations with variables, lists, and functions. We'll explore how Python code works, how to document your notebooks, and run cells interactively.

This is your first step as a data scientist!



Expectations and Participation

- Attend and participate actively in each session.
- Practice coding regularly—even small exercises help.
- Ask questions: there are no bad questions.
- Be curious and explore data and tools beyond the class.
- Respect collaboration and feedback—it's part of learning.

This course is as much about mindset as it is about coding.



Quiz Time: Baseline Assessment

Let's start with a quick quiz to understand your background and interests.
The quiz is anonymous and covers:



Wrap-up and Next Steps

Thank you for participating in this first session! Before next class:

- Try accessing Google Colab and running the examples again.
- Review the course site and syllabus.
- Join the course GitHub (link will be shared).

Next class, we'll dive deeper into data preprocessing and explore our first dataset using Python.

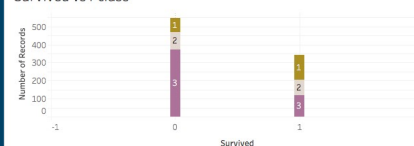


Assignment – Titanic Dataset in Google Colab

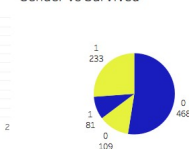
Exploring the Titanic Dataset, is a warm-up exercise to practice working with data in Python using Google Colab.

- **Steps to follow:**
- Open Google Colab (<https://colab.research.google.com/>) and create a new notebook.
- Download the Titanic dataset from Kaggle (<https://www.kaggle.com/c/titanic/data>) or use the preloaded data
- import seaborn as sns vdf = sns.load_dataset('titanic')
- Perform the following tasks:
 - Load the dataset and display the first 5 rows.
 - Check the number of rows and columns, and describe the data types.
 - Calculate the number of missing values per column.
 - Show how many passengers survived and how many did not.
 - Find the average age of passengers.
 - Create one simple plot (bar or histogram) using seaborn or matplotlib.

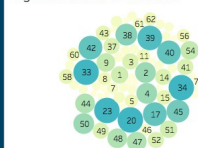
Survived vs Pclass



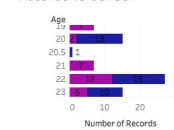
Gender vs Survived



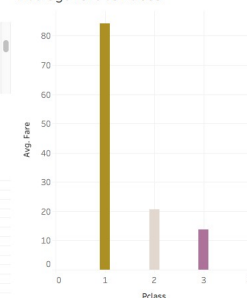
Age vs Number of Records



Age vs Number of Records vs Gender



Average Fare vs Pclass



Gender vs Pclass



Deliverable: Submission Guidelines

- Submit the **Google Colab link** (make sure it's public or shared with the instructor) before next class.
- Include **clear code comments** and a **short summary** of what you discovered.



Summary

The casistors link cole tions,
and perersiontie caile for
avest your text.