Final Project

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

INTER NATIONAL SUMMER SCHOOL

ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO

UNIVERSIDAD

Final Project: Data Mining & Neural Networks

- **Objective:** Apply Data Mining and Neural Networks techniques.
- **Format:** Practical project with student-chosen dataset and theme.
- **Duration:** 9-day course, submission on the last day.

Suggested Project Themes

- **TOPIC 1:** Text Analysis & Natural Language Processing (NLP)
- **TOPIC 2:** Computer Vision
- **TOPIC 3:** Tabular Data Analysis (Num & Categ)
- **TOPIC 4:** Time Series Analysis

ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO

UNIVERSIDAD

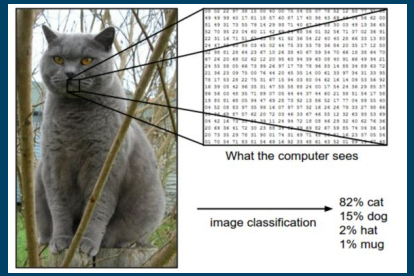
TOPIC 1: Text Analysis & Natural Language Processing (NLP)

Description: Ideal for combining text mining with neural networks.

Examples:

- **Sentiment Classification**: Reviews, social media comments.
- **Spam/Fraud Detection:** Emails, messages.
- **Document Classification:** News, articles.
- **Text Summarization:** Generating concise summaries.

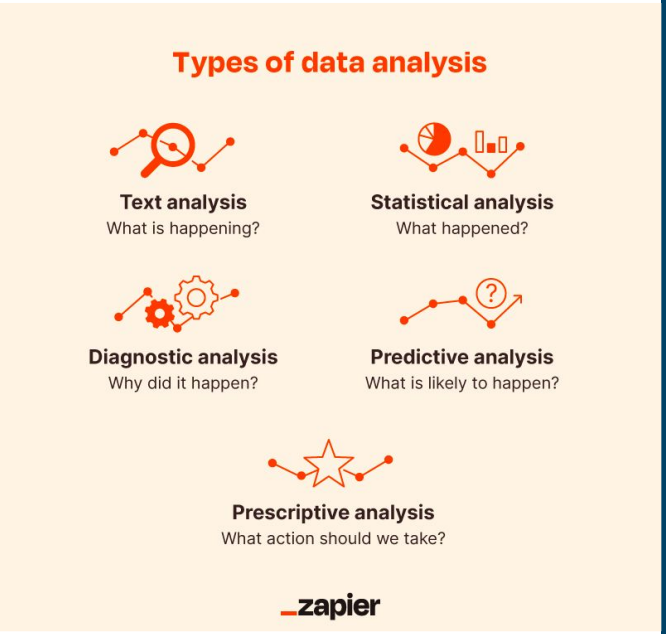
TOPIC 2: Computer Vision



Description: Preparing and analyzing large volumes of image data.

Examples:

- **Image Classification**: Objects, animals, scenes, medical diagnoses.
- **Facial/Object Recognition**: Detection and recognition.
- **Anomaly Detection**: Industrial quality control, security.



TOPIC 3: Tabular Data Analysis (Num & Categ)

- **Description:** Basis for many business and scientific applications.
- **Examples:**
  - **Fraud Prediction:** Financial transactions.
  - **Customer Churn Prediction:** User retention.
  - **Product/Service Recommendation:** Purchase history.
  - **Price Prediction:** Goods, stocks.
  - **Medical Diagnosis:** Disease prediction.

TOPIC 4: Time Series Analysis

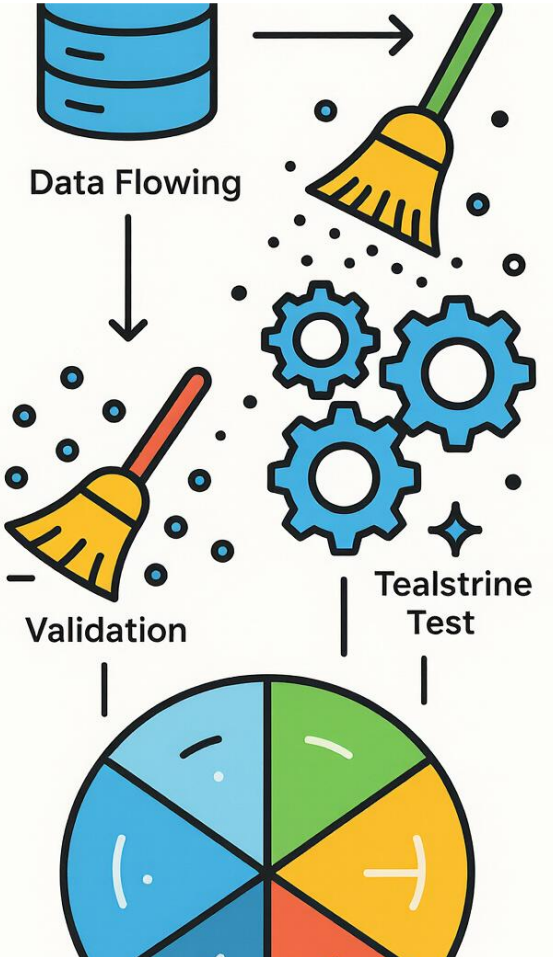
- **Description:** Data that changes over time, using RNNs/LSTMs.
- **Examples:**
  - **Stock/Price Prediction:** Price movements.
  - **Demand Forecasting:** Inventory optimization.
  - **Weather Prediction:** Weather patterns.
  - **Signal Analysis:** Audio, vibration, sensor data.

Project Phases

- **Phase 1:** Ideation & Proposal (Day 3 - Day 4)
- **Phase 2:** Data Exploration & Preprocessing (Day 4 - Day 5)
- **Phase 3:** Neural Network Design & Training (Day 5 - Day 7)
- **Phase 4:** Evaluation, Analysis, & Conclusions (Day 8)
- **Phase 5:** Project Presentation (Day 9 - Final Presentation)

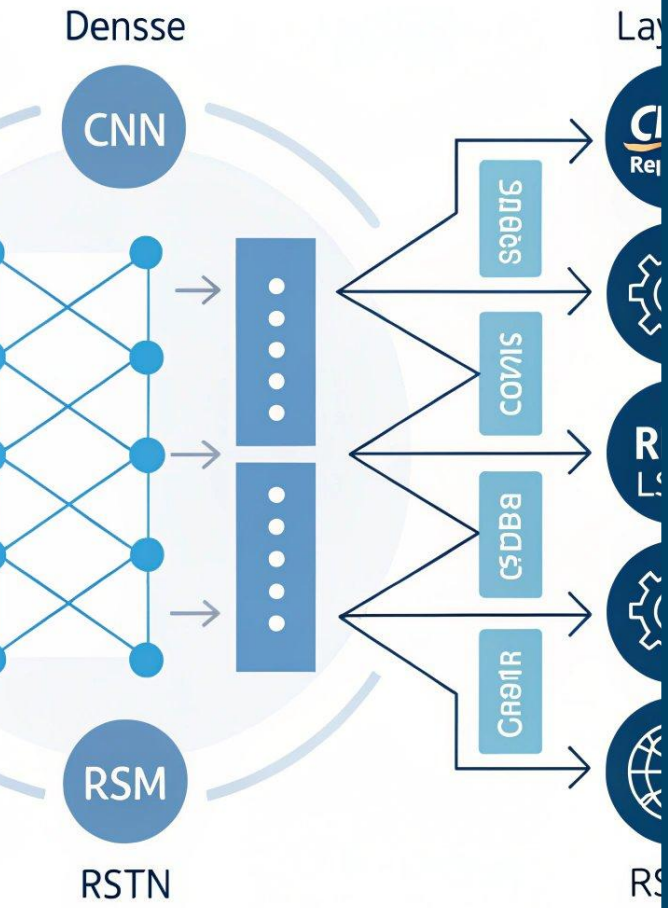
Phase 1: Ideation & Proposal (Day 3 - Day 4)

- **1.1. Theme Exploration & Problem Definition:**
  - Choose a specific problem from the suggested themes.
  - Key question: What problem do I want to solve?
- **1.2. Dataset Search & Selection:**
  - Find a relevant public dataset (Kaggle, UCI ML, open data).
- **1.3. Proposal Formulation (Written):**
  - Project Title.
  - Problem to Solve.
  - Dataset Description (name, source, size, characteristics).
  - Project Objectives.



Phase 2: Data Exploration & Preprocessing (Day 4 - Day 5)

- **2.1. Initial Loading & Exploration:**
  - Load dataset, exploratory analysis: data types, nulls, distribution.
- **2.2. Data Cleaning:**
  - Handle nulls, duplicates, errors, outliers.
- **2.3. Feature Transformation & Preparation (Feature Engineering):**
  - **Text:** Tokenization, stop-words, lemmatization, TF-IDF, Word Embeddings.
  - **Images:** Resizing, normalization.
  - **Tabular:** Categorical encoding, numerical scaling.
  - **Time Series:** Lag creation, window-based features.
- **2.4. Dataset Splitting:**
  - Separate into training, validation, and test sets

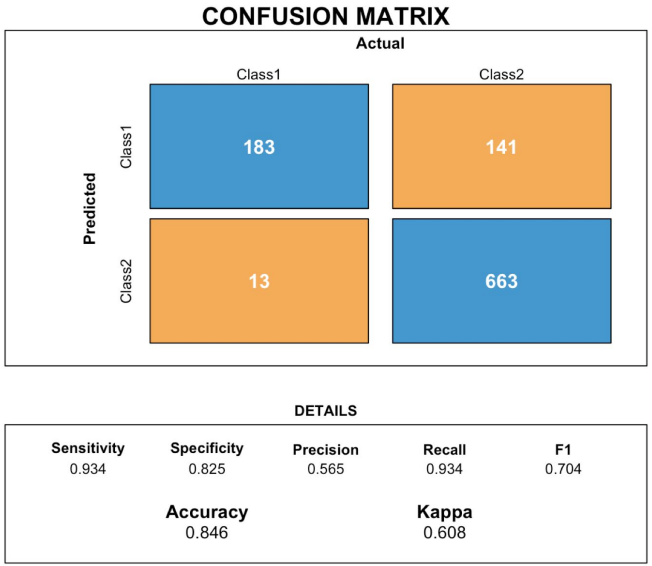


Phase 3: Neural Network Design & Training (Day 5 - Day 7)

- **3.1. Neural Network Architecture Selection:**
  - Choose and justify architecture: Dense, CNN, RNN/LSTM.
- **3.2. Model Construction:**
  - Implement in TensorFlow/Keras: layers, activations, output.
- **3.3. Compilation & Training:**
  - Configure: loss function, optimizer (Adam, RMSprop), metrics (accuracy, precision, recall, F1-score).
  - Train model and monitor validation.
- **3.4. Hyperparameter Tuning (Optional/Advanced):**
  - Experiment with learning rates, epochs, batch sizes, layers.

ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO

UNIVERSIDAD



Phase 4: Evaluation, Analysis & Conclusions (Day 8)

- **4.1. Model Evaluation:**
  - Use test set for final performance metrics.
- **4.2. Results Analysis:**
  - Interpret metrics, visualize (confusion matrix, ROC curves).
  - Analyze strengths and weaknesses.
- **4.3. Conclusions & Future Improvements:**
  - Reflect on findings, limitations, possible improvements.

Phase 5: Project Presentation (Day 9 - Final Presentation)

- **5.1. Report/Presentation Preparation:**
  - Summarize work: Introduction, methodology, results, conclusions, future work.
- **5.2. Demonstration & Defense:**
  - Brief model demonstration and Q&A session.

INTER NATIONAL SUMMER SCHOOL
UNIVERSIDADE DA CORUÑA

ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO

UNIVERSIDAD