

Datathon Report – MC

Datathon 2025

*This document presents the detailed work and solutions proposed by Chicas de datoss members during the **MC-Datathon 2025**. It includes explanations, models used, evaluation strategies, and insights for each challenge.*

Challenge 1: Churn Detection

Solution: `churn-detection` by Asma

Type: Binary Classification

Evaluation Metric: `F1 Score`

Problem Statement:

The goal was to predict whether a customer is likely to **churn or not**, based on various features provided in a tabular dataset.

Approach:

- **Data Exploration:**
 - Handled missing values, categorical encoding, and distribution analysis.
 - Visualized correlations between features and churn label.
- **Preprocessing:**
 - Normalization/standardization where needed.
 - Split into training/validation/test sets.
- **Modeling:**
 - Tried multiple classification models: (e.g. Random Forest, XGBoost, Logistic Regression)
 - Used `GridSearchCV` or cross-validation for hyperparameter tuning.
- **Evaluation:**
 - Final model evaluated using **F1 Score**, to handle imbalanced classes.
- **Results:**
 - Achieved a score of **[insert score here]** on the test set.

Notes:

- ☐ Insert exact model name and best hyperparameters.
- ☐ Add insights about features or feature importance.

Challenge 2: Football Players Detection

Solution: `football-players-detection` by Asma

Type: Object Detection

Evaluation Metric: `mean Average Precision (mAP)`

Problem Statement:

Develop an object detection model to identify and localize all players within a football field from static images.

Approach:

- **Dataset Handling:**
 - Used bounding box annotations provided.
 - Split dataset for training/validation.
- **Modeling:**
 - Used a YOLO-based or SSD-based deep learning model.
 - Trained the model with bounding boxes and image augmentation techniques.
- **Submission Format:**
 - Predictions returned as: `filename, xmin, ymin, xmax, ymax, confidence, label`
- **Evaluation:**
 - Submissions scored using **mean Average Precision (mAP)** for object detection accuracy.

Notes:

- ☐ Add mAP score you obtained.
- ☐ Mention architecture (e.g. YOLOv8, Faster R-CNN).
- ☐ Challenges: Small objects? Occlusion? Background noise?

Challenge 3: Arabic Manuscripts Digitization

Solution: `chicas2` by Malak

Type: Handwriting Transcription (OCR)

Evaluation Metric: `Character Error Rate (CER)`

Problem Statement:

Automatically transcribe Arabic handwritten manuscripts, particularly in the **Maghribi** style, into machine-readable text.

pproach:

- **Preprocessing:**
 - Normalized image sizes, binarization, and noise removal.
 - Cropped/padded inputs for consistency.
- **Modeling:**
 - Built a **CRNN (CNN + BiLSTM + CTC)** architecture.
 - Defined custom Arabic vocabulary to improve accuracy.
- **Training:**
 - Trained on labeled dataset with image-text pairs.
- **Evaluation:**
 - Measured **Character Error Rate (CER)** on test set.
 - Lower CER = better performance.

Notes:

- ☐ Insert architecture details (e.g., VGG + BiLSTM).
- ☐ Mention final CER score.
- ☐ Add samples of predicted vs ground truth text.

Challenge 4: Sentiment Anaylisys challenge

Solution: `becca's solution`

Final Remarks:

This report outlines the datathon journey from a hands-on, technical perspective. Each challenge provided a unique problem domain and evaluation metric:

- **Churn Detection:** Binary classification using tabular data
- **Football Detection:** Real-time object localization
- **Arabic Manuscripts:** Sequence recognition for OCR

Version Control:

All notebooks and code are versioned and available via the GitHub repository.