# Predictive and Descriptive Learning and Machine Learning Lab

Final Project

Title: *Predicting/Classifying ….. from ……*

Author(s): Author A, Author B,…

(This may be a Team work by a **maximum of three team members**)

January 2018

* **In addition to this document you must also prepare a short presentation (10 minutes)**
* **Don’t forget to include detail information in ANEXES I, II & III**

## Abstract

This document focuses on building prediction/classification models to …. Exploratory data analysis and visualization techniques are used for…

We present the development and performance evaluation of different models…..

## Introduction

…..…

## Related Work

…..…

## Dataset(s)

…..…

## Features: Visualization & Exploratory Data Analysis

…..…

## Models’ development

…..…

## Results and discussion

…..…

## Conclusions and Future Work

…..…

## ANEX I: Tools and development environments (MLlab)

Include a short description of **tools you have used and for what purposes**…

Languages R, Python…

Libraries: as Sklearn… XGBoost… TensorFlow, Keras

Environments:

Google Colaboratory, Google Cloud, RSTudio, Jupyter Notebooks, IBM Workbench, Spark

## ANEX II: Summary of Theoretical Knowledge (Predictive & Descriptive Learning)

* + Statistical Analysis for exploratory data analysis: Correlation Matrix, Statistical Contrasts (Hypothesis tests)…
  + Unsupervised learning for exploratory data analysis: PCA, tSNE, Clustering
  + Feature selection….
  + Linear and non-linear models: MLR, LR,… GLM…
  + Tree-based models…. RF, Bagging, Boosting,…
  + SVM…
  + Deep Learning Models: DNN, CNN, LSTM, … Transfer Learning, Attention, …
  + Model assessment, selection, Crossvalidation,…

## ANEX III: Team activities (only for Projects with several authors)

## Include BOTH practical activities (such as developing a Python Notebook to implement a LSTM model) as well as theoretical studies (such as to study what model should be theoretically be better to address your problems)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Author A** | **Author B** | **Author V** |
| Statistical Analysis  in R | 33% | 33$ | 33% |
| Statistical Analysis  in Python | 60% | 20% | 20% |
| PCA | - | - | 100% |
| tSNE | 50% | 50% |  |
| RF development in  R | 80% | 10% | 10% |
| RF development in  Sklearn | 10% | 80% | 10% |
| RF development in  Spark | 10% | 10% | 80% |
| XGBoost R | 100% | - | - |
| Pipelines in Spark to compare Boosting and RF | - | 50% | 50% |
| Deep Learnin Models assessments using Keras | - | 100% | - |
|  |  |  |  |
|  |  |  |  |