

**NOVA**

**IMS**

Information  
Management  
School

# *WIZARDRY SCHOOL ENROLLMENT*

MACHINE LEARNING I  
Group Project Handout

# 1. Introduction

In this mystical world, young wizards and witches, each possessing their own unique potential, yearn to attend the most prestigious of institutions – the esteemed wizardry schools. These schools are known not only for their grandeur but also for their ability to unlock the hidden powers within each student.

In the heart of this enchanted realm, we find ourselves at a crossroads. The doors to these hallowed halls are open, but only to those deemed worthy.

Our mission is clear: to discern, with precision and insight, who among these aspiring enchanter shall earn the privilege of stepping into the world of wizardry. It is a quest that delves deep into their qualifications, their backgrounds, and their unique magical potential.

Join us as we embark on a journey into the mystical world of wizardry, where spells are cast, potions brewed, and destinies forged. Together, we shall determine who shall walk through the gates of wisdom and power, and who shall remain in the realm of the mundane.

Let the magical assessment begin!

## 2. Project Objective

You will be provided with a representative sample of data that includes information about students who have previously been accepted or not accepted into the wizardry school. The data is divided between a training set and a testing set.

Your objective is to develop a predictive model using the training set, leveraging the qualifications and magical potential of aspiring students, and then use this model to make accurate predictions on the testing set. This predictive model will play a crucial role in determining which students are suitable for enrollment in our wizardry schools, ensuring that only the most deserving candidates gain admission to the realm of magical education.

## 3. Datasets

You have access to two different datasets:

1. The training set should be used to build your machine learning models and assess their performance if needed. In this set, you also have the ground truth associated with the student admission, i.e., if the student was admitted to the school or not.
2. In the test set, you will see the same features presented in the training set. However, you will not have access to the ground truth of the test set. Your goal will be to

predict the ground truth value (“0” or “1”) by using the model you created using the training set.

The available data contains the following features:

Variable Name	Description
Student ID	Unique identifier
Program	The program to which the student is applying, the admission process is the same for all the programs
Student Gender	Gender of the student
Experience Level	Experience level of the students, indicating their skill and knowledge within the wizardry school
Student Siblings	Number of siblings
Student Family	Number of family members attending the same wizardry school
Financial Background	Financial status or wealth of the student's family
School Dormitory	Type of dormitory accommodation in the wizardry school the student is applying
School of Origin	School of origin for the aspiring students
Student Social Influence	Number between 1-25 representing the student's popularity or social influence within the school, based on factors like the number of friends and social activities
Study Element	Student's favorite element
Admitted in School	Outcome variable worth 1 if the student was admitted to the school

## 4. Deliverables

Upon the project's deadline, you will be required to submit a zip file containing:

1. A report that describes the analytical processes and the conclusions obtained with, at most, 15 pages (excluding cover, but including annexes). The file naming format should follow *ML1\_GroupXX\_Report.pdf*, where *GroupXX* should be your group number. Use the template provided in the file *Report\_Template.docx* with the following settings:

- Heading 1: Calibri, Size 14 pt, in bold
  - Heading 2 (if needed): Calibri, Size 13 pt, in bold
  - Text: Calibri, Size 11 pt, line spacing of 1.15 pt and paragraph spacing of 6 pt
2. A Jupyter notebook with your code implementation. The file naming format should be ***ML1\_GroupXX\_Notebook.ipynb***, where “GroupXX” should be your group number
  3. A pdf version of the Jupyter Notebook
  4. A csv file containing your test set's “StudentID” and your predictions for column “Admitted in School”. The file should only contain 2 columns.
  5. Your final predictions should be submitted to Kaggle.

## 5. Evaluation

The grading will be based on 2 main criteria:

1. The report (50%)
2. The defense (50%)

## 6. Notes

- 1) We will disregard steps and results that are not mentioned in your report
- 2) Report is expected to have analysis of the results.
- 3) When in doubt, we will run your Jupyter Notebooks. Therefore, make sure we can run the notebook from start to finish in one go. Notebooks that do not fulfil this condition will be penalized.
- 4) All the code that is not needed to obtain your final solution should be commented.
- 5) The report and code will pass through a process of plagiarism checking.
- 6) Your predictions will be evaluated using the metric F1-Score.
- 7) Attending the defense is Mandatory, The defense has a group component and an individual component.