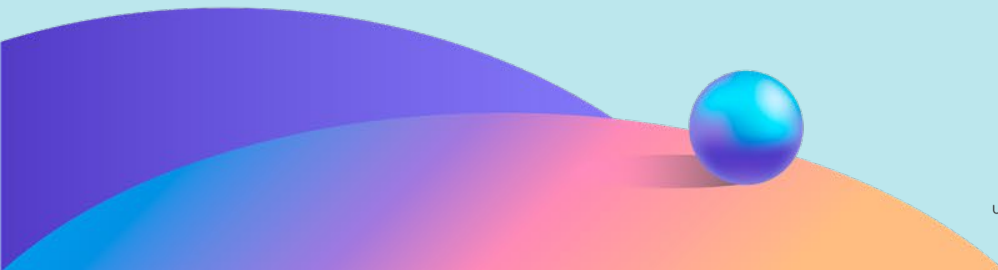




# Making Data FAIR and AI/ML Applications for Cancer Prevention and Control (AI/ML-CAPAC) Research among Hispanics

**NCI Grant No. R25CA240120-03S1**

Principal Investigator: Ana Patricia Ortiz, PhD  
Course Professor: Abiel Roche-Lima, PhD  
Coordinator: Sandra García-Camacho, MPH



**This project aims to expand the scope of the parent CAPAC Research Training Program (1R25CA240120) and prepare research workforce on:**

**01**

**The techniques and approaches to manipulate and pre-process Hispanics cancer datasets to make them FAIR and AI/ML ready, and on...**

**02**

**The available methods for developing ML-based models to analyze these data and create predictive models for cancer diagnosis and treatments with a focus on datasets from Hispanic populations.**

## Course Logo



## Tagline

**Preparing a workforce to  
apply AI/ML techniques to  
datasets derived from  
Hispanic populations to  
advance cancer prevention  
and control research.**

## New Online Course offered by the CAPAC Research Training Program!



### Artificial Intelligence & Machine Learning in Cancer Prevention and Control (AI/ML-CAPAC) Research



The CAPAC Program invites interested individuals to apply to this introductory course focused on preparing a workforce with the competencies and skills needed to make Hispanic datasets FAIR (Findable, Accessible, Interoperable, and Reusable) and AI/ML-ready, and create AI/ML - based predictive models for cancer diagnosis and treatments.

- Modality: Online
- Contact hours: 24 contact hours during 12 weeks
- Cost: Free for selected participants
- Dates: May 16, 2022 to August 5, 2022
- Primary professor: Abiel Roche-Lima, PhD, UPR Medical Sciences Campus
- Additional participating faculty from University of Puerto Rico Comprehensive Cancer Center and collaborating institutions



**New  
Application  
Deadline:  
May 3, 2022**

**To participate in the course  
you need to complete an  
application form through the  
following link:  
[https://crcweb.rcm.upr.edu/re  
dcap/surveys/?  
s=ECAHYFYM4](https://crcweb.rcm.upr.edu/re-dcap/surveys?s=ECAHYFYM4)**

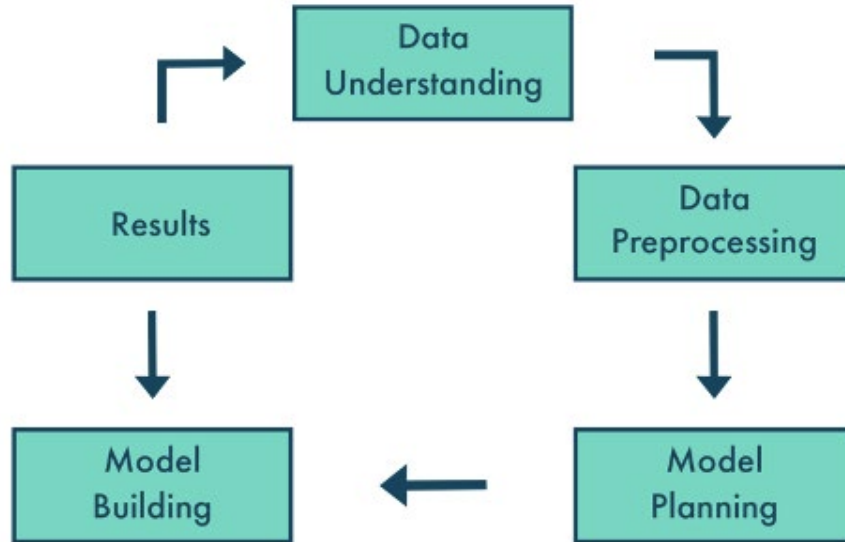


# Course Flyer

Course flyer was sent to CAPAC participants (alumni) and applicants, CAPAC mentors, as well as trainees and research staff from collaborating grants/institutions.

# We developed the online course based on the data science project lifecycle.

## *Lifecycle of Machine Learning Project Phases*



Following this lifecycle a ML project can be developed starting from the data understanding and preprocessing, including the model planning and building and finally, obtaining the results

# About the Course

- **Modality:** Online at self-paced
- **Contact hours:** 24 contact hours during 12 weeks
- **Language:** English
- The course was divided in two components with modules. After each module, there were readings and short quizzes.

## **Component I:** Making the cancer datasets FAIR and AI/ML-ready

- Fundamentals of Cancer Data Types and Cancer Datasets 'Omics data in Cancer Prevention Research
- Principles of Artificial Intelligence and Machine Learning
- Applications of AI, ML in Cancer

## **Component II:** Creating predictive AI/ML models for cancer diagnosis and treatments

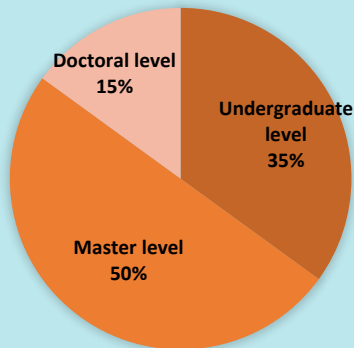
- Introduction to Machine Learning Projects
- Programming with Python
- Python for Understanding and data pre-processing in Cancer Datasets
- Supervised Learning Algorithms to Create Predictive Models for Cancer
- Final Project Orientation and Example

- The course **culminates with a project** in which students applied what they have learned.

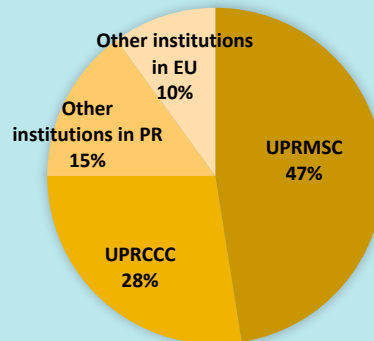
# Students:

- To apply to the course, they filled out a form that included questions regarding their background and contact information. In addition, they filled out a section indicating why they were interested in taking the course.
- Forty-one people applied to the course.
- All the available 40 slots were covered.
- **Female: 67.5%, Male: 32.5%**

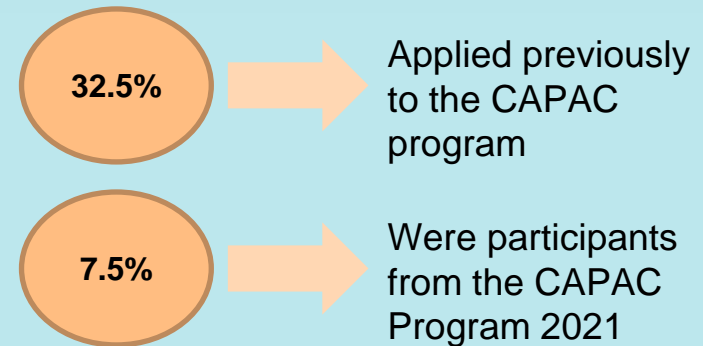
## Education Level



## Affiliations



## Capac Program





# Course Professors



**Abiel Roche-Lima, PhD**

Course Lead Professor  
Director of Integrated  
Informatics Services,  
CCHRD-UPRMSC



**Ana Patricia Ortiz, PhD**

Program Director  
CAPAC Research Training  
Program, Division of Cancer  
Control & Population Sciences,  
UPRCCC



**Josúe Pérez-Santiago, PhD**

Course Professor  
Associate Investigator of  
Computational Biology and  
Bioinformatics UPRCCC



**Diego Zavala-Zegarra, PhD**

Course Professor  
Co-Director, Puerto Rico Central  
Cancer Registry



**Harry Hochheiser, PhD**

Course Professor  
Associate Professor of  
Biomedical Informatics in the  
University of Pittsburgh



# RCM Online Platform



- The course was provided through an online platform from the UPR-MSC.
- In this platform the students had all the course materials available.

## CAPAC Virtual Training Program (CAPAC - VTP)



FAIR+AIML+CAPAC

Title: FAIR Data and Artificial Intelligence / Machine Learning applied to Cancer Prevention  
Professor: Dr Abiel Roche-Lima.  
Modality: Self-paced (learn on your own time and schedule)  
Time of Completion: 24 hours

- The course was published in the course catalog of the RCM platform. Through this catalog students accessed the course.

# Course Completion

- Students that completed 70% or more of the course requirements passed the course.

## Course Completion

- **26 students (70.3%)** passed the course.
- **11 students (29.7%)** not passed the course.

\*Three students were excluded because they do not started the course

Example of the certificate  
that was sent to the  
students that passed the  
course.

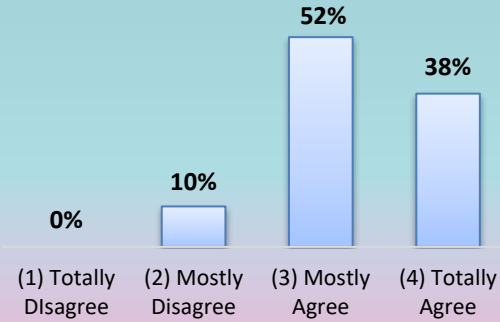




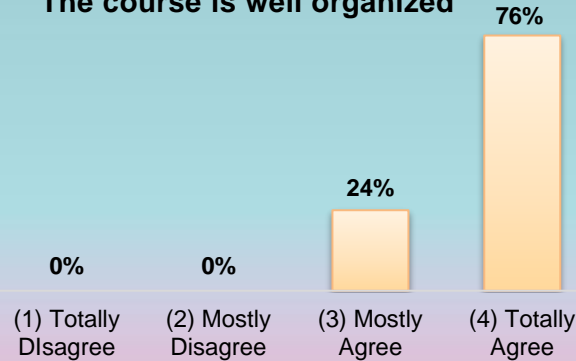
# Some important questions from the course evaluation:

(21 students answered the course evaluation)

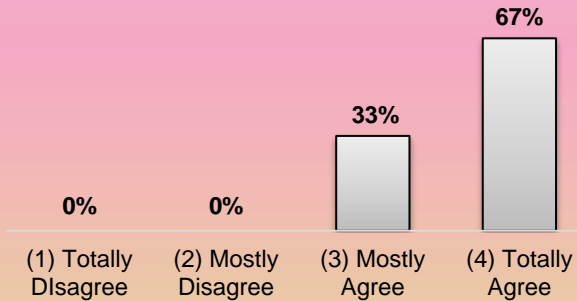
### The course is easy to handle



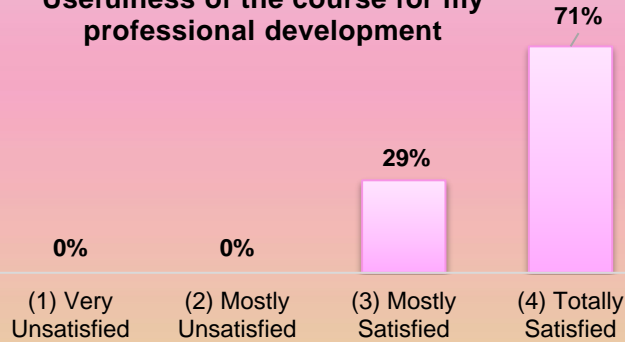
### The course is well organized



### The work plan is easy to understand and follow




### Usefulness of the course for my professional development





# Next Steps:

- 
1. Write and publish a manuscript about the development and description of the course.
  2. Make the course available in the future to additional cohorts.
- 