IntelliSys 2022

1-2 September 2022

How to Improve the Teaching of Computational Machine Learning Applied to Large-Scale Data Science

Sergio Rogelio Tinoco-Martínez, Heberto Ferreira-Medina, José Luis Cendejas-Valdez, Froylan Hernández Rendón, Mariana Michell Flores Monroy, Bruce Hiram Ginori Rodríguez

What is the Problem?

In public universities in Mexico, specifically in ENES Morelia UNAM, Machine Learning (ML), Deep Learning (DL), and Big Data subjects are taught in the curriculum plan.

These subjects are taken from the 4th semester up to the 8th semester. Nevertheless, there are some cases where the students cannot identify and solve AI, ML, DL, and Big Data problems

Solution Proposal

Modification and improvement to the courses that adhere to the preferred requirements of the students. The solution to these problems is to give higher priority to practice with real data and use cases than to abstract theory

How to Improve Teaching

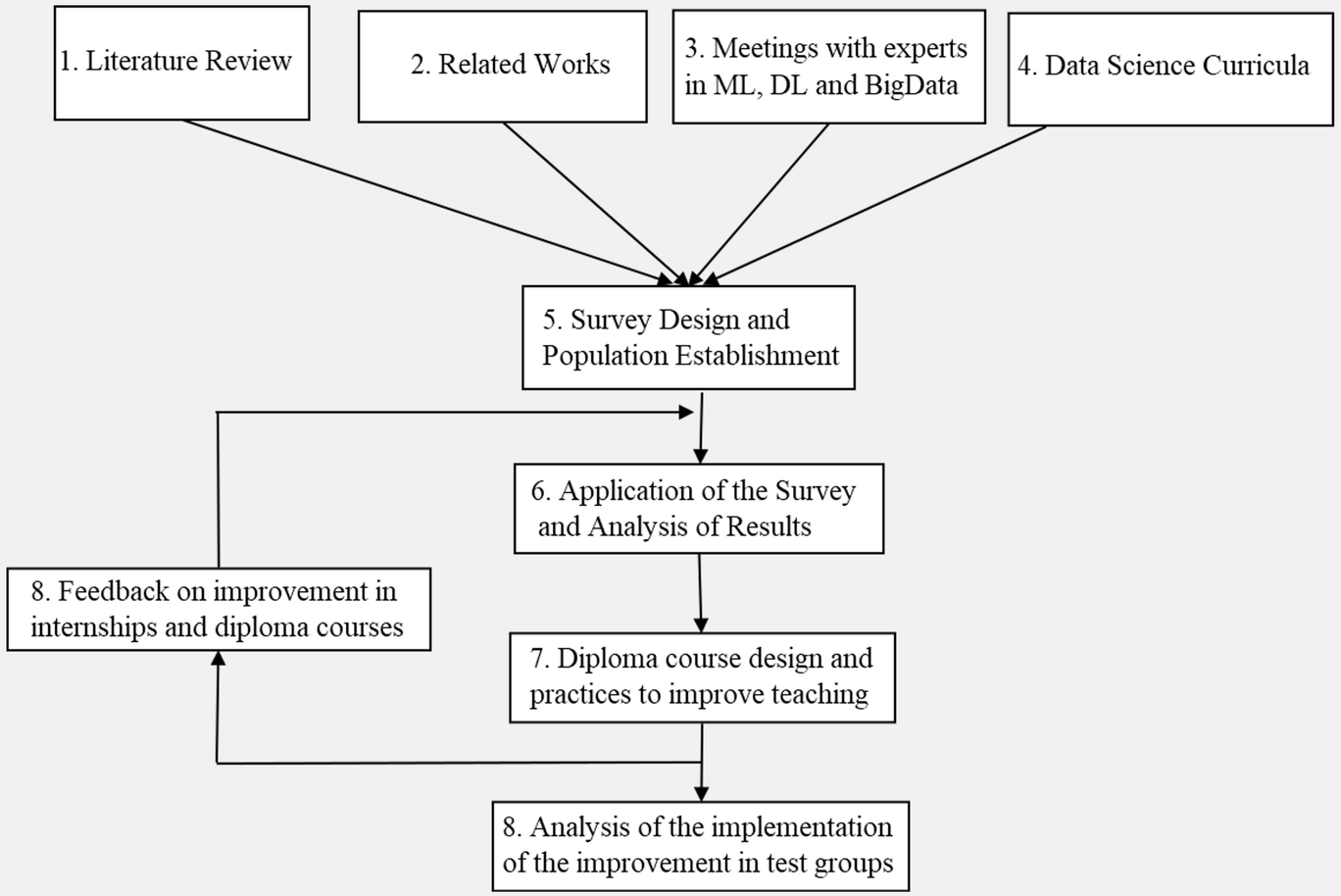
We implemented a previous knowledge survey, where the students and teachers select the topics of interest.

To solve the learning deficit, we propose a ML course based on the more unfamiliar and requested topics.

Send the survey to qualified students and teachers.

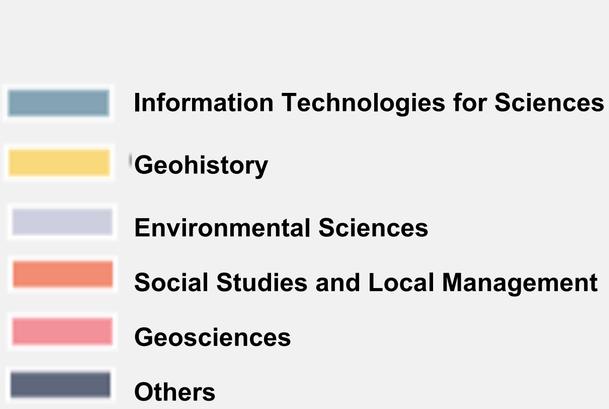
Collect the data and analyze them.

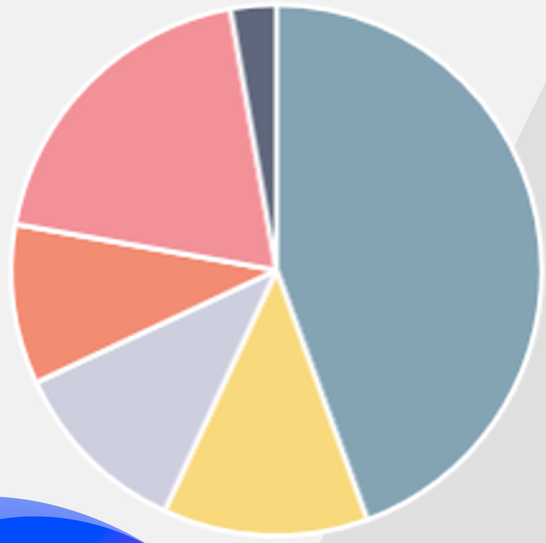
Select the appropriate topics to implement practices in Python.



Methodology

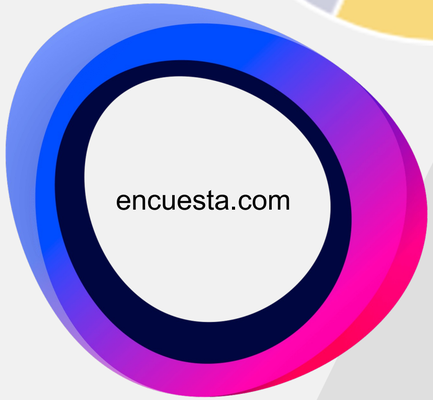
Developing the Survey





The base population size of 600 students and teachers from careers related to ML, DL, and Big Data

Online survey to compile and process all submissions



Random students and academic staff were selected across the UNAM campus in Morelia

Factors to Improve Teaching

Diploma course

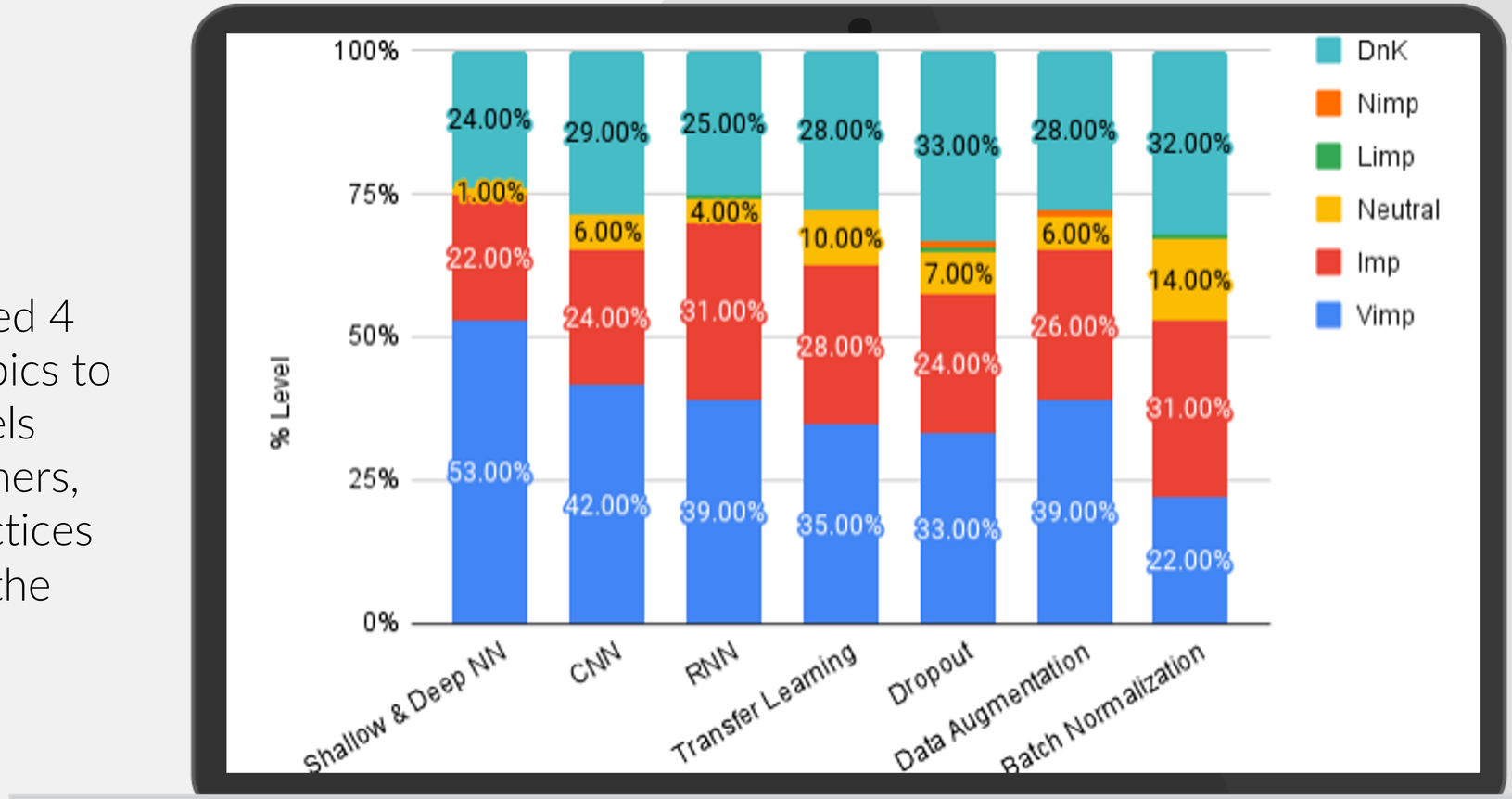
MODULE 1- Machine Learning (ML). “Theory and Practice for the Improvement of the Teaching of ML Applied to Data Science”

MODULE 2- Deep Learning (DL). "Theory and Practice for the Improvement of the Teaching of DL Applied to Data Science"

Survey Results

Quantifying and analyzing results

Based on the survey, we identified 4 important axes for classifying topics to cover, and based on the skill levels shown by the students and teachers, we began to design the first practices to fit the average knowledge of the respondents



DL axis and related topics

Designing the ML Module

Practices that were planned and designed to cover ML needs for all students and teachers

Hours in an intensive inter-semester diploma course

#	Name	Dataset	Evaluation Metric
1	Classification using decision trees	Titanic passengers	Accuracy and/or Fbeta Metrics
2	Housing cost prediction	California Housing	RMSE and/or MAE
3	k-Nearest Neighbors	Water wells	Precision Score Fbeta
4	k-Means	Online Retail k-means & Hierarchical Clustering	Not applicable
:	:	:	:

Designing the DL Module

6

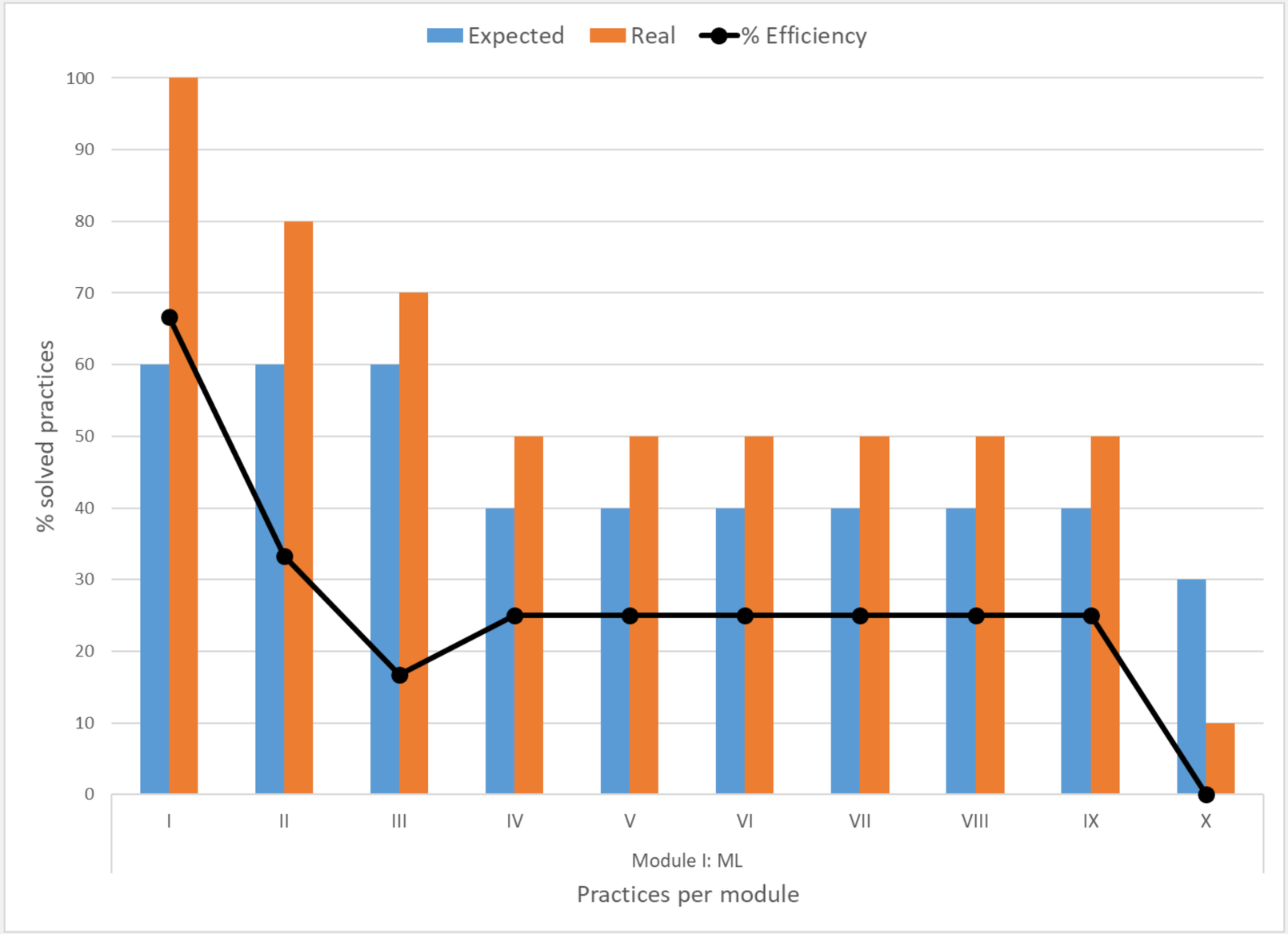
Practices that were planned and designed to cover DL needs for all students and teachers

45

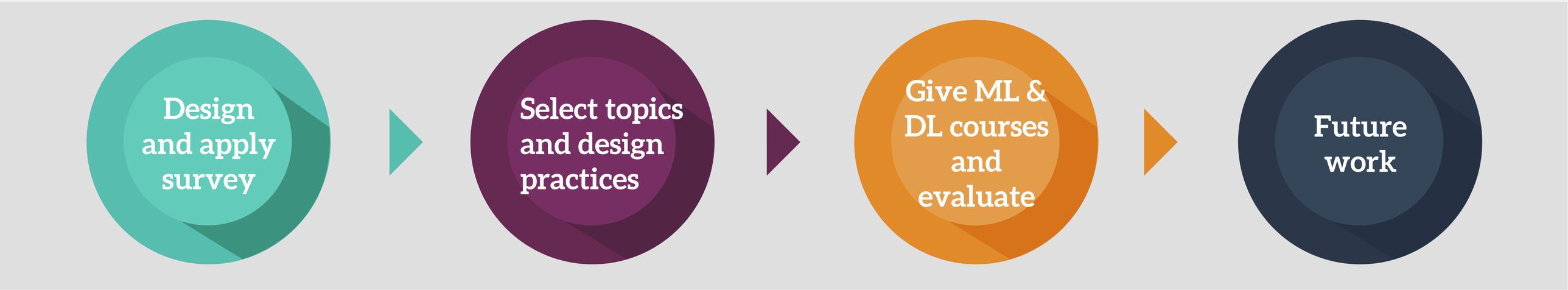
Hours in an intensive inter-semester diploma course.

#	Name	Dataset	Evaluation Metric
1	Binary classification with CNN	800 images of mosquitoes, UNAM	Accuracy
2	Binary classification with CNN	Covid-19 Pneumonia Screening	Precision and Confusion Matrix
3	CNN & Data Augmentation	Ship Classification	Precision and Confusion Matrix
4	RNN	Sarcasm Detection	Accuracy and Precision
:	:	:	:

ML Teaching Effectiveness



Results and Conclusions



ACKNOWLEDGMENT

We are grateful for the support of the Escuela Nacional de Estudios Superiores (ENES) and Instituto de Investigaciones en Ecosistemas y Sustentabilidad (IIES) from UNAM Campus Morelia, CA TRATEC-PRODEP of the Universidad Tecnológica de Morelia (UTM), the TecNM campus Morelia, and DGAPA UNAM PAPIME PE106021.

Contact

Sergio Rogelio Tinoco-Martínez
stinoco@enesmorelia.unam.mx

Heberto Ferreira-Medina,
hferreira@iies.unam.mx

José Luis Cendejas-Valdez
luis.cendejas@ut-morelia.edu.mx

Froylán Hernández Rendón
fhernandez@enesmorelia.unam.mx

Mariana Michell Flores Monroy
michellmonroy18@gmail.com

Bruce Hiram Ginori Rodriguez
bruginrod@gmail.com

ESCUELA NACIONAL DE ESTUDIOS SUPERIORES
UNIDAD MORELIA

UTM