Tecnológico de Monterrey - CCM

TC3006C
Advanced Artificial Intelligence for Data Science I

Module 2: Statistics for data science

Introduction to the module

Professor Profile

Jesús Manuel Vázquez Nicolás, PhD

Professor at Tecnológico de Monterrey.

Senior Computer Vision Engineer

Professor Profile

Bachelor's degree in Mechatronics

UPIITA-Instituto Politécnico Nacional

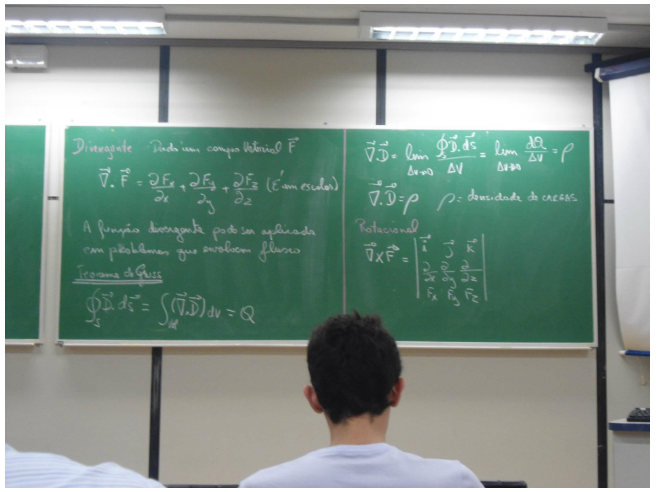
Thesis: 'Kinesthetic Teleoperation System for a Mobile Manipulator'

Professor Profile

Professor Profile

One semester of the degree (Brazil) at Universidade Estadual de Campinas

UNICAMP



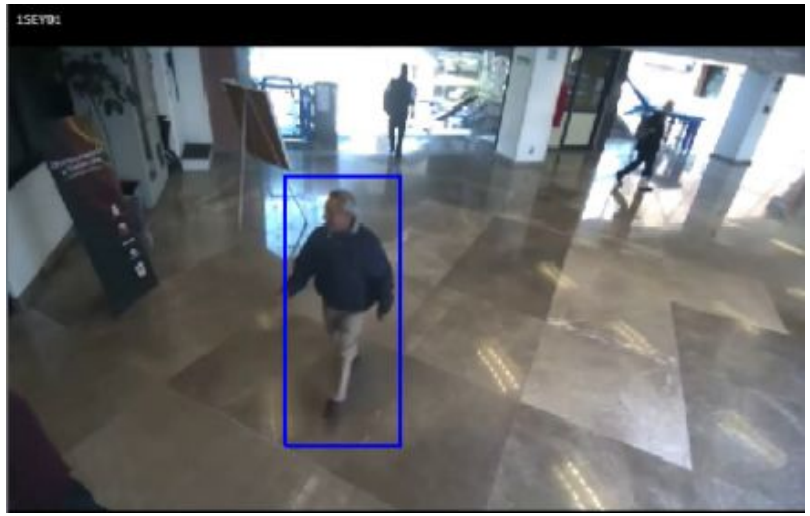
Professor Profile

Master of Science in Computer Engineering

Centro de Investigación en Computación - IPN

Thesis: 'Monitoring and Tracking System for People using Computer Vision with a focus on security.'

Centro de Investigación en Computación
Instituto Politécnico Nacional

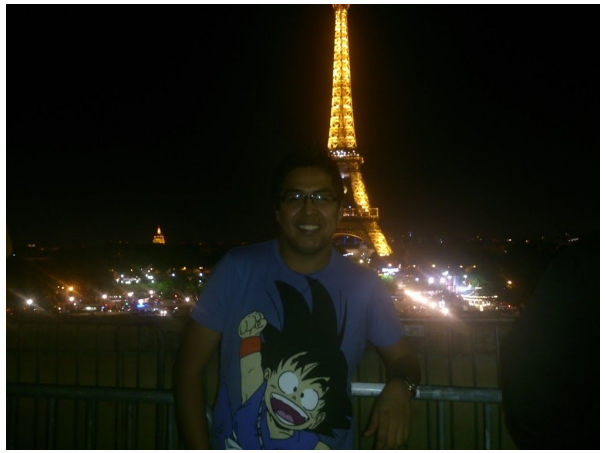


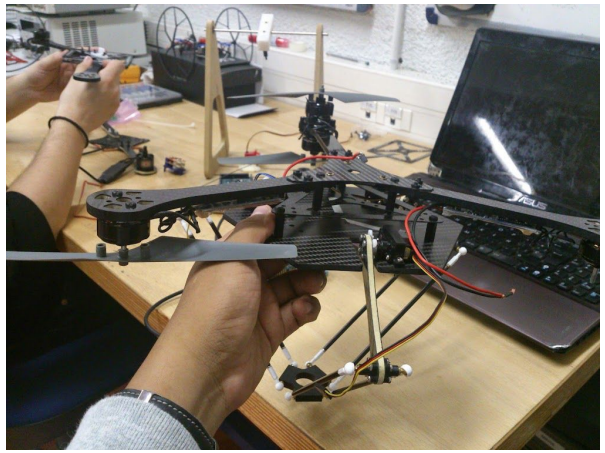
Professor Profile

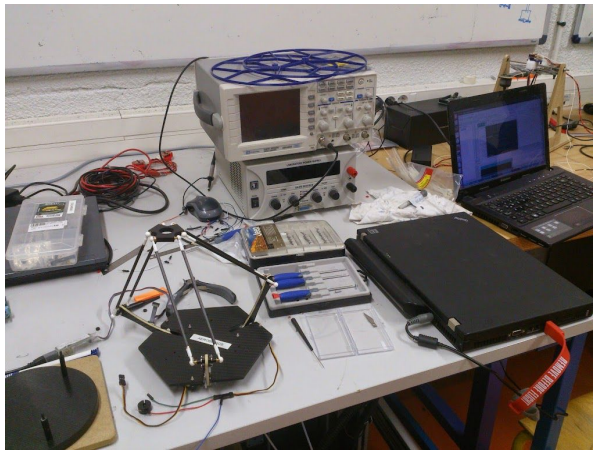
Research stay (France)

- Institut Polytechnique Des Sciences Avancées
- Université d'Evry-Val-d'Essonne

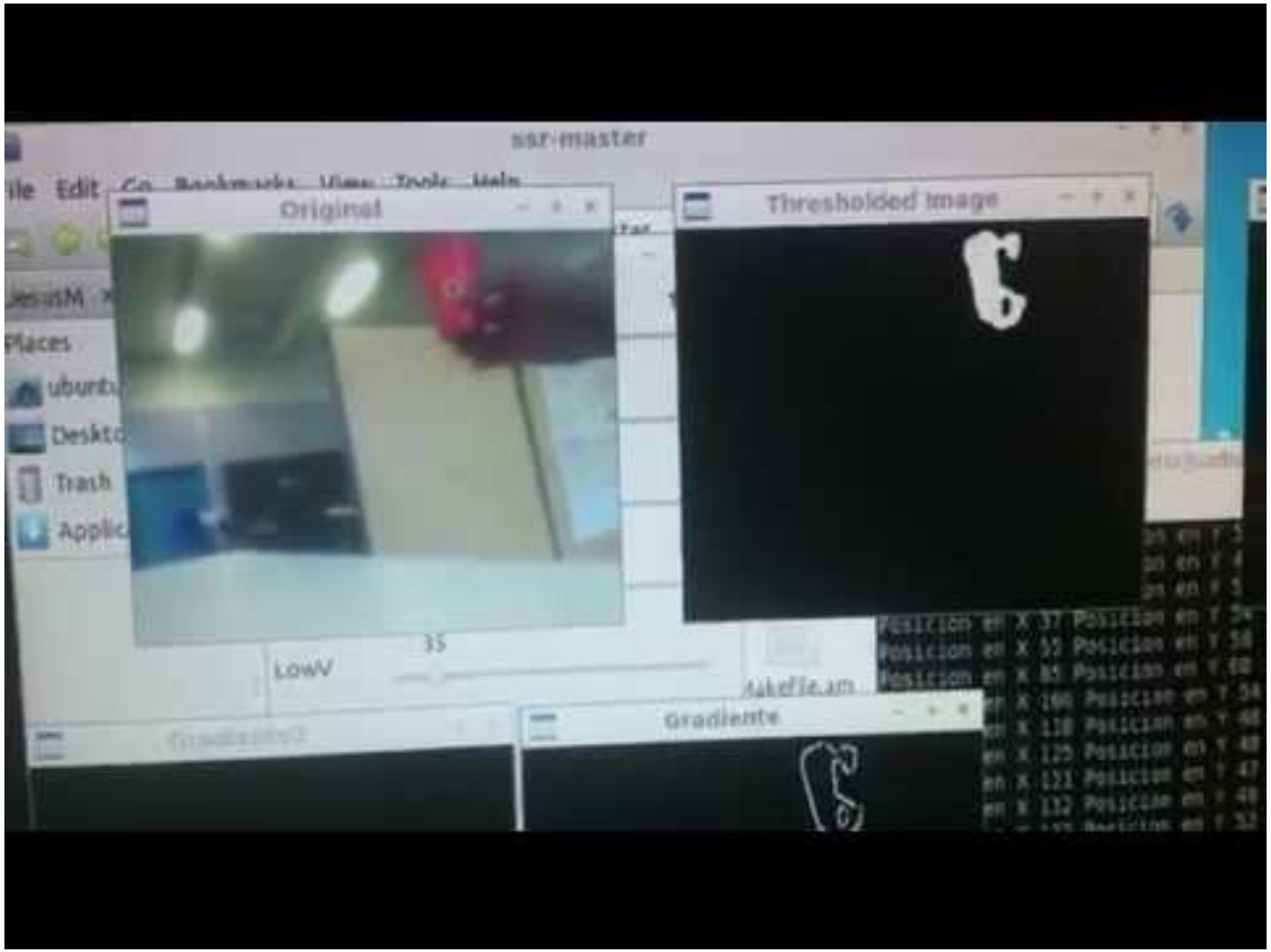
Trabajo: Optimal grasping points identification for a rotational four-fingered aerogripper







Professor Profile

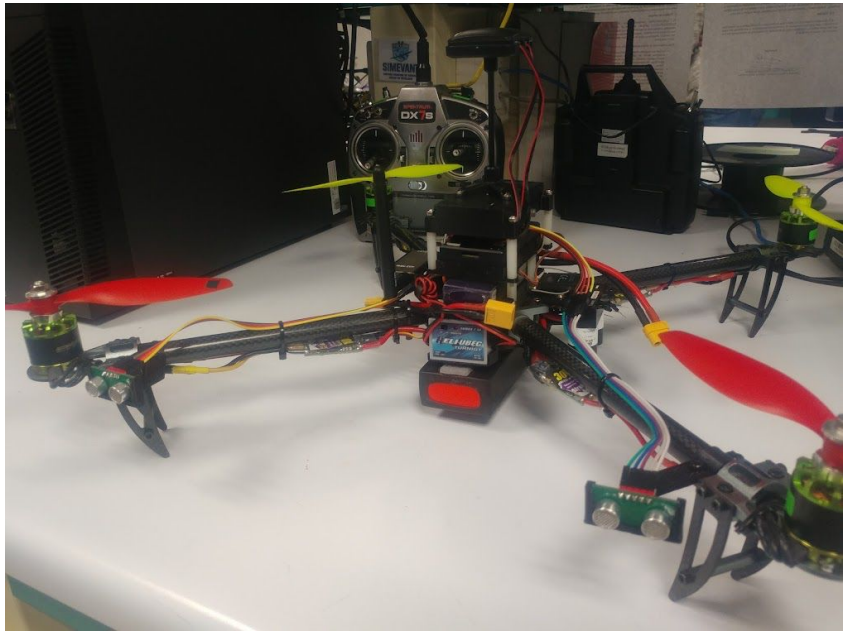


Professor Profile

PhD in Autonomous Systems for Aerial and Underwater Navigation

CINVESTAV-IPN

Thesis: 'Autonomous Navigation of an Aerial Robot for Inspection Tasks on Structures



Professor Profile

Now it's your turn...

Name

Why did you choose this course?

Hobbies

Life goal

Contact Information

E-mail:

jesus_vazquez@tec.mx
jvazq150@ford.com

Instant messaging:

Assessment Plan

Actividades	Ponderacion
Module 1. Python Libraries	10%
Module 2. Statistics for Data Science	10%
Module 3. Machine Learning	10%
Module 4 Big Data and AWS	10%
Module 5. Natural Language Processing 1	10%
Evaluation Module	10%
Total	**60%**

Evidencia	Ponderacion
Presentación argumentativa	40%

Evaluation Module	Ponderacion
Module 1. Python Libraries	2%
Module 2. Statistics for Data Science	2%
Module 3. Machine Learning	2%
Module 4 Big Data and AWS	2%
Module 5. Natural Language Processing 1	2%
Total	**10%**

Activities

During the module, we will be conducting various activities such as quizzes and coding exercises. **All activities are part of the assessment, and it is important to complete them.**

Basic Rules

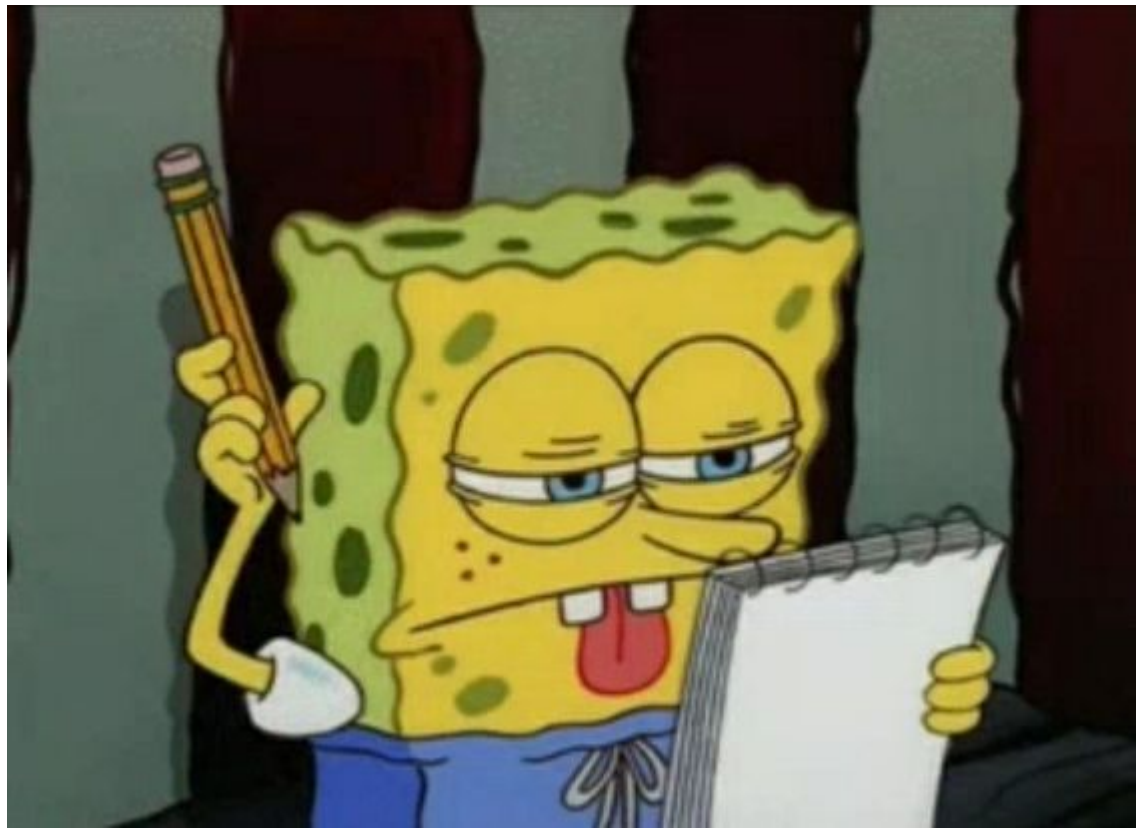
1. Class start time: 7:00 p.m.
2. Respect is fundamental among class members, both in the classroom and on Slack.
3. Activities that appear to be copied will lower your grade, potentially leading to their nullification.
4. Quizzes will be conducted during class, so if you do not attend, you will miss that evaluation

Software tools

Syllabus

1. Data Extraction, Transformation, and Loading (ETL)
    a. Data Extraction
    b. Data Transformation
    c. Data Loading
    d. Data Imputation
2. Metrics
    a. Classification Metrics
    b. Regression Metrics
    c. Cross-Validation
    d. Confusion Matrix
3. Descriptive Statistics
    a. Measures of Central Tendency
    b. Measures of Dispersion
    c. Correlation
4. Probability Theory
    a. Basic Concepts
    b. Conditional Probability
    c. Random Variables
    d. Important Theorems
    e. Hypothesis Testing
5. Optimization
    a. Basic Concepts of Optimization
    b. Convex and Non-Convex Optimization
    c. Optimization Algorithms
    d. Optimization in Machine Learning:
    e. Principal Component Analysis (PCA)

What is AI and Data Science?



- **Artificial Intelligence**: Branch of computer science that seeks to develop algorithms, models, and techniques that enable machines to learn and make decisions in a way similar to humans.

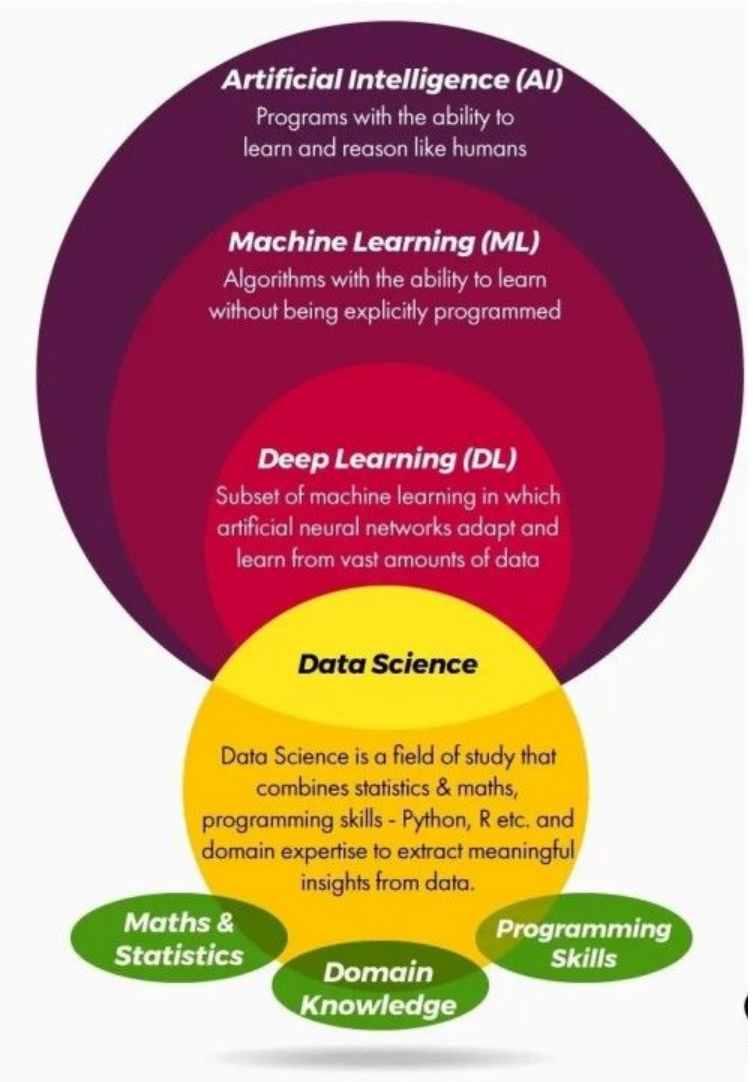
- **Data Science**: Interdisciplinary field that uses scientific methods, algorithms, and systems to extract knowledge and insights from various types of data.

How do DS and AI relate to each other?

Artificial Intelligence (AI): Focuses on creating algorithms that enable machines to perform tasks that require intelligence, such as learning, reasoning, and perception.

Data Science (DS): Multidisciplinary field centered around extracting knowledge and insights from data. It uses scientific techniques, processes, and systems for data analysis.

Point of Intersection: Machine Learning - This is the area where AI and DS overlap the most, allowing for data-driven analysis and predictions.



What is the way?

Applications

- Customer Service: Chatbots and virtual assistants can handle a wide variety of customer inquiries, providing quick and accurate responses 24/7.
- Autonomous Vehicles: Autonomous cars, trucks, and drones use AI to navigate the world safely and efficiently.
- Fraud Detection: In the financial sector, AI is used to identify patterns of suspicious behavior and prevent fraud.
- Personalized Medicine: Data Science and AI are being used to develop personalized treatments based on the individual characteristics of each patient.
- Weather Prediction: Machine learning algorithms can analyze massive amounts of weather data to predict weather conditions.

Applications

- Personalized Marketing: Companies use Data Science to personalize their marketing campaigns, offering customers products and services that fit their interests and needs.
- Smart Agriculture: With the help of AI and Data Science, farmers can optimize their practices to increase efficiency and productivity.
- Inventory Control: Companies use AI to optimize inventory management, improving efficiency and minimizing waste.
- Internet Search: Search engines use AI to provide more accurate and relevant results to users.

Why are data important?

The quality of the data used to feed an artificial intelligence or data science model is fundamental to its success.

If the input data is filled with errors, inaccuracies, or biases, the results produced by the model will also be filled with errors and inaccuracies, regardless of how sophisticated the model is.

Models, no matter how advanced they are, are only as good as the data they are trained on.

Why are data important?

This is why a significant part of the work in data science and artificial intelligence is dedicated to data collection, cleaning, transformation, and analysis. This process is crucial to ensure that the data is in a format that the models can effectively use and that it accurately and fairly represents the phenomenon or process being modeled.

Therefore, it is essential for any data scientist or AI engineer to understand their data and dedicate the necessary time to preprocess and analyze them before selecting and training a model.

References

1. Grus, J. (2019). *Data science from scratch: First Principles with Python*. O'Reilly Media.

2. Bruce, P., Bruce, P. C., Bruce, A., & Gedeck, P. (2020). *Practical Statistics for data scientists: 50+ Essential Concepts Using R and Python*. O'Reilly Media.

3. Nield, T. (2022). *Essential math for data science: Take Control of Your Data with Fundamental Linear Algebra, Probability, and Statistics*. O'Reilly Media.

4. Reis, J., & Housley, M. (2022). *Fundamentals of Data Engineering: Plan and Build Robust Data Systems*. O'Reilly Media.