CS5056 Data Analytics
March 17, 2021
Class 6/16

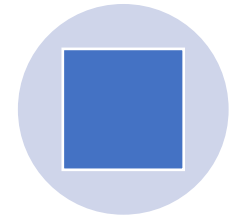Tecnológico de Monterrey

Agenda
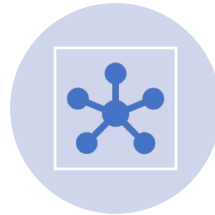
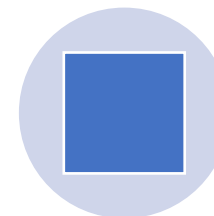**COMMENTS ON HWA3, FPA2, MIDDLETERM EXAM, CONFERENCES, AND CASE STUDIES**

HANDS-ON EXERCISE ON MODEL EVALUATION

TIME:

PANEL DATA (FC) AND TIME SERIES ANALYSIS (HC)

LOGISTIC REGRESSION

HANDS-ON EXERCISES USING PYTHON, R, WEKA

CS5056 Data Analytics
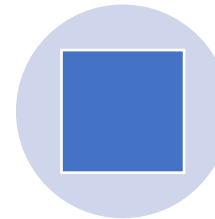## March 24, 2021
Class 7/16

Agenda

UNCERTAINTY: THE BAYESIAN APPROACH

BAYES CLASSIFIER
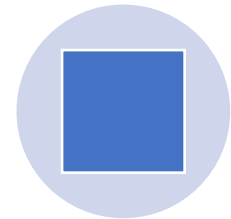
BAYESIAN NETWORKS

EXPECTATION MAXIMIZATION

HANDS-ON EXERCISES USING PYTHON, R, WEKA

Tecnológico de Monterrey
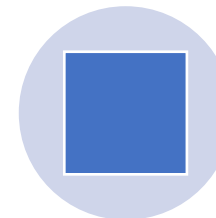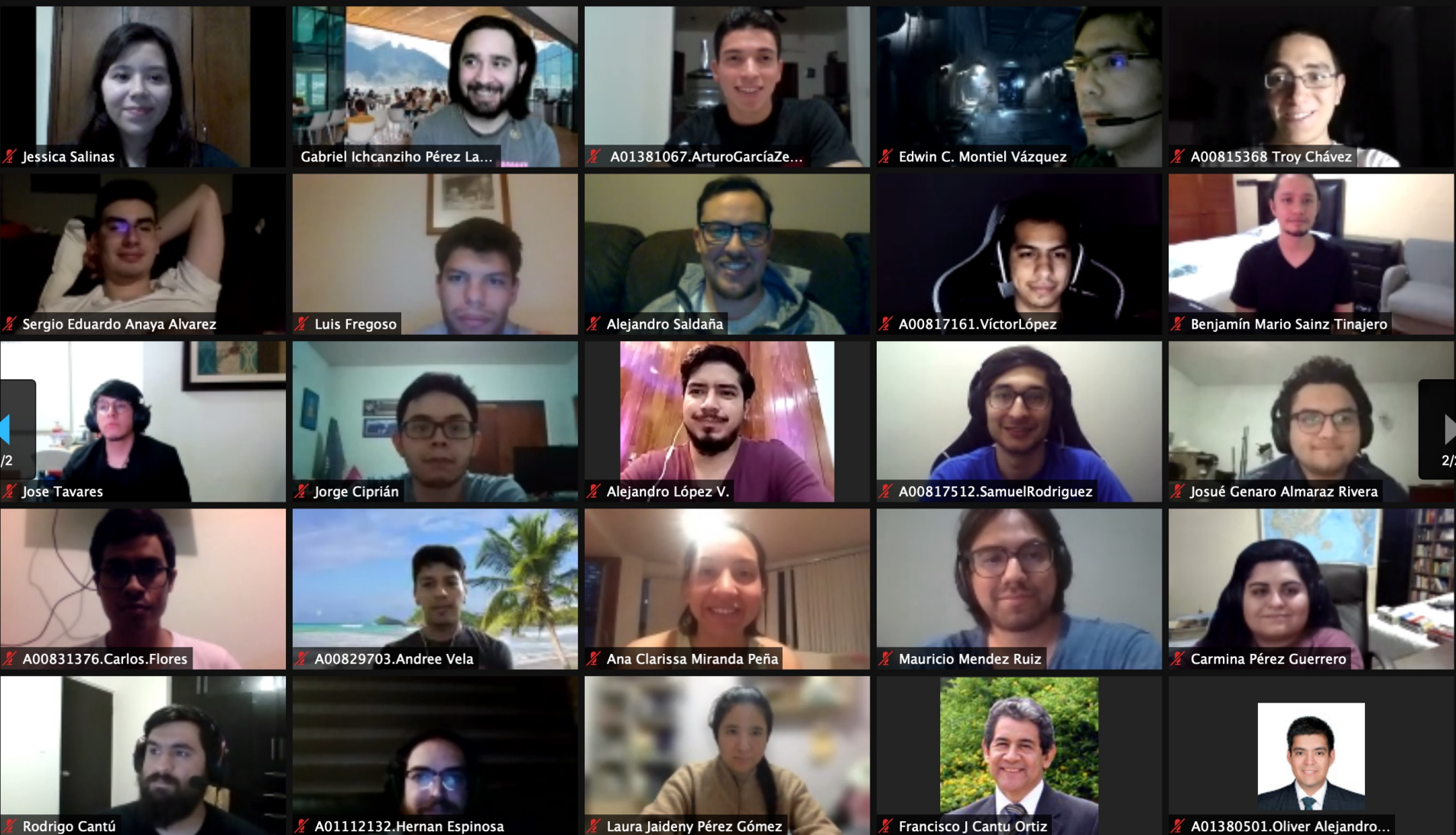
Active Learning

With Various Hands-on Exercises in Python and R

Enjoy the Class and have Fun!!!

HWA3
Q1 – Overfitting
Q2 – QS Rankings
Q3 – House Pricing √

FPA2

Q1 – Fitting a Model to Data

Q2 – Keystroke Dynamics

Q3 – Model Selection (Frydman)

FPA3

Q1a – KNN (Ch6)

Q1b – Unbalanced DS (Ch7)

Q1c – ROC/AUC (Ch8)

Q2 – FP (Model Validation)

Conference Presentations

Gregorio

Gabriel

Genaro

Víctor

Welcome to the 17th Int. Conference on Data Science (ICDATA'21)

Non-Attendance Policy/Presentation Modes due to Coronavirus/COVID-19 – please see info here.

We invite you to attend ICDATA'21 and to submit papers. ICDATA'21 offers a 4 day single-track conference, keynote speeches by world renowned scientists, special sessions and free tutorials on all aspects of data science. In the past, the data science conference has attracted a lot of presentations in formal sessions plus keynotes, and excellent tutorials, with a lot of attendees plus many more visitors from other conferences of the congress.
For further information on the 2021 World Congress in Computer Science, Computer Engineering, and Applied Computing (CSCE'21) and its conferences, please use the following link: CSCE'21

ICDATA'21 is part of CSCE'2021. ICDATA'21 is an international conference for academics and practitioners held simultaneously with other joint conferences as part of CSCE'21, the 2021 World Congress in Computer Science, Computer Engineering, and Applied Computing. CSCE'21 is one of the largest annual gatherings of researchers and practitioners in computer science, computer engineering and applied computing. Each of the joint conferences in CSCE is a premier conference for presentation of advances in their respective fields. The congress includes **20 major tracks** composed of 122 technical, research, and panel sessions as well as a number of keynote lectures and tutorials. All will be held simultaneously, same location and dates (July 26-29, 2021) to facilitate cross-disciplinary knowledge transfer. This year's Congress will be a hybrid event that combines a "live" in-person event with a "virtual" online component.

It is anticipated that CSCE'21 will attract between 1,000 and 2,000 participants from about 75 countries/territories. Participants range from established researchers on Professorial level, starting researchers (post-doc, PhD, Masters and Bachelor level) and contract researchers unto practitioners applying all aspects of data science in private or public companies or research institutions. 2019, 54% of attendees were from academia, 25% from industry; 20% from government and funding agencies; and 1% unknown. About half of the attendees were from outside USA; from 69 nations.

ICDATA'21 attracts quality papers & presentations. The data science conference has been highly ranked in the past. Please to not confuse ICDATA (formerly DMIN) with other „multi-conferences" with approximately 100% acceptance rates. The conference maintains its own program committee and manages its own thorough and objective review process of at least 2 independent reviewers (leading academics & practitioners). Over the past ten years, ICDATA has an average acceptance rate of approximately 50%. While this rate is not as low as for some top-tier conferences, it is respectable and still allows researchers from less developed countries to actively participate in the conference.

Publication & Indexing. Due to the high quality of published papers in the proceedings of CSCE in recent years, this year (2020) ALL accepted papers will be published by **Springer Nature**.

Publisher: Springer Nature – Book Series: Transactions on Computational Science & Computational Intelligence, https://www.springer.com/series/11769

https://icdatascience.org/

Data Analytics

Articles Final Projects, Feb-June 2021

Id	Authors	Project Name	Status & Learning Outcomes
1	Diana Laura Aguilar	A Contrast Pattern-Based Analysis of theRestaurant Industry in the US in the Year 2020	
2	Josué Genaro Almaraz Rivera	COVID-19 Vaccination Progress Prediction inMexico	
3	Cristian Yair Alvarado Elizalde Mario Alberto Mendoza Cuevas	Analyzing Factors That Generate CompanyBankruptcy	
4	Luis Alfonso Rojo S ́anchez Sergio Eduardo Anaya Alvarez	Current Abortion Situation Analysis in Mexico	
5	Andree Vela Juan Carlos Ángeles	INEGI 2020 Population Survey Analysis ofDevelopment Disparity of Disabled Individuals inYucatan	
6	Rodrigo Alejandro Cant ́u Hern ́andez	Fintech vs. Traditional Banks:Mexican User Experience Market Analysis	
7	Alma Socorro Torres Torres Troy David Ch ́avez Gut ́errez	COVID-19 Impact on Labour Productivity byIndustry and Region in the United Kingdom	
8	J. F. Ciprian-Sanchez G. I. Perez-Landa	Sentiment analysison information related to news coverage on well-known newsnetworks such as BBC, CNN, NBC, and Fox News	
9	Alejandra De Luna P ́amanes Mauricio Mendez Ruiz	A sport climbing study: A data science approach	
10	Hernán Espinose Rodríguez		
11	Jessica Salinas Espinosa Carlos Flores Munguia	Sentiment Analysis for the Identification ofMisinformation on COVID-19 Vaccines	
12	Luis Fregoso Aparicio	Data Analysis for a automobile accident databasefrom Mexico to prevent accidents and reduceinsurance payment	
13	Andres Alberto Lopez Esquivel Arturo Garcıa Zendejas	Electric motor data analysis for temperatureestimation	
14	Ramon Hinojosa Alejandro	Twitter and Reddit post analysis on the subject of Cryptocurrencies	

Data Analytics

Articles Final Projects, Feb-June 2021

Id	Authors	Project Name	Status & Learning Outcomes
15	Victor Ivan Lopez Rodriguez	COVID-19 New Cases Prediction after Lockdownusing Mobility and Recovery Data in Monterrey	
16	Alejandro Saldana Rodrıguez Alejandro Lopez Vazquez	Technical Indicators in Indexed Stocks:ACRISP-DM Approach	
17	Clarissa Miranda-Pẽna	The Winning Formula for Football Competitions	
18	Edwin C. Montiel-V́azquez	Natural Language Processing for COVID-19vaccine perception study	
19	Laura Jaideny Ṕerez Ǵomez	Discussion topics on COVID-19 vaccine tweets	
20	Carmina Ṕerez Guerrero	Metric Analysis for the Semantic Segmentation ofJet Fire Temperature Zones	
21	Gregorio Arturo Reyes Gonźalez	A Bayesian Method Approach to Digital ViolenceAgainst Women	
22	Astrid Rivera Partida Diego Villatoro Geronimo	Biodiversity of the Worldwide Food Production	
24	Alfonso Angel Medina-Santana	Global Environment Facility impact on renewableenergy-based rural electrification: A data analyticsperspective	
25	Samuel Guadalupe Rodrıguez Rodrıguez	Food retail sales prediction using weather forecast	
26	Benjamín Mario Sainz Tinajero	Performance Analysis and Comparison of Multi-objective Evolutionary Clustering Akgorithms	
27	Juan Andres Talamas Carvajal	Social media Effects on the market: Reddit andTwitter Data analysis on Stocks	
28	Oliver Alejandro Veĺazquez Flores	League of Legends and Teamfight Tactics (TBD)	

Francisco J. Cantú-Ortiz, PhD

Professor of Computer Science and Artificial Intelligence
Tecnológico de Monterrey
Enago-Academy Advisor for Strategic Alliances

E-mail: fcantu@tec.mx, fjcantor@gmail.com

Cel: +52 81 1050 8294, SNI-2 CVU: 9804

Personal Page: http://semtech.mty.itesm.mx/fcantu/

Facebook: fcantu; Twitter: @fjcantor; Skype: fjcantor

Orcid: 0000-0002-2015-0562

Scopus ID:6701563520

Researcher ID: B-8457-2009

https://www.researchgate.net/profile/Francisco_Cantu-Ortiz

https://scholar.google.com.mx/citations?hl=es&user=45-uuK4AAAAJ

https://itesm.academia.edu/FranciscoJavierCantuOrtiz

Ave. Eugenio Garza Sada No. 2501, Monterrey N.L., C.P. 64849, México