



Universidad Austral
Maestría en Ciencia de Datos

Laboratorio de Implementación III
edición Virtual

Libro de la Asignatura

Edición 2024

Autor: Gustavo Denicolay

última actualización: 2024-05-15 15:00

Tabla de Contenidos

1La Asignatura.....	3
1.1Sobre este documento.....	3
1.2Links fundamentales.....	4
1.3UAustral Cronograma de la Asignatura año 2024.....	5
1.4Modalidad y Criterios de Aprobación y Evaluación.....	6
1.4.1Modalidad y Criterios de Aprobación y Evaluación Competencia Kaggle.....	7
1.4.2Exposiciones en Clase.....	8
1.4.3Discusiones en Clase.....	8
1.4.4Informe final.....	8
1.4.5Evaluación participación en foro/chat Zulip.....	9
1.5Misión de la Asignatura.....	10
1.5.1Objetivos de la Asignatura.....	10
1.6 Unidades Temáticas.....	11
1.7Bibliografía General de la Asignatura.....	12
2Metodología de Enseñanza.....	14
2.1Generalidades.....	14
2.2Manifiesto Pedagógico.....	14
2.3Pensamiento Creativo - Tinkering.....	15
2.4Dinámica Constructivista en el Aula.....	16
2.5Authentic Learning.....	16
2.6Flipped Classroom Forte.....	16
2.7Conectivismo.....	17
2.8Active Learning.....	17
3Arranque en Frío.....	19
3.1Alta en Plataformas.....	19
3.1.1Zulip herramienta de chat y foros discusión.....	19
3.1.2Google Sheet Colaborativa de la Asignatura.....	20
3.1.3Plataforma Hypothes.is.....	20
3.1.4Plataforma ChatGPT.....	20
3.1.5Plataforma Kaggle.....	21
3.1.6Plataforma GitHub.....	22
4Herramientas, Conceptos, Operación y Buenas Prácticas.....	24
4.1Zulip.....	24
4.1.1Buenas prácticas para el uso de Zulip.....	25
4.1.2Participaciones extraordinariamente significativas en Zulip.....	26
4.1.3Participaciones significativas en Zulip.....	26
4.1.4Mensajes Privados a los profesores Permitidos.....	26
4.2Git y GitHub.....	28
4.2.1conceptos y operación.....	28
5¿Cómo seguir?.....	30
5.1Mantenerse actualizado.....	30
5.2Datasets Públicos.....	30
6Empiricismo en la Ciencia de Datos.....	31
6.1.1Bibliografía Introdutoria.....	31
7Las dos culturas en el modelado predictivo.....	32
8Estimando la Ganancia de un modelo predictivo.....	33
8.1Bibliografía Introdutoria.....	33
8.2Sesgo Varianza en la estadística clásica.....	33
8.2.1Bibliografía Inicial.....	33
8.3Sesgo Varianza en el moderno Machine Learning.....	34

1 La Asignatura

En Laboratorio de Implementación III usted se encuentra en el último bimestre de la Maestría en Ciencia de Datos. A lo largo de estos dos años ha desarrollado una buena cantidad de habilidades en la disciplina y es momento de enfrentarlo *en forma mínimamente guiada* a la toma de decisiones e implementación de un caso real simulando una auténtica consultoría.

En su versión 2024 esta asignatura será una **intensa** experiencia de exploración de soluciones sujeta a un activo debate grupal, en donde deberá poner a prueba gran parte de los conocimientos adquiridos en la maestría bajo situaciones adversas.

Enfrentará un problema real de forecasting, con todo el barro propio de una empresa inmersa en una macroeconomía inestable, baja de ventas y eventualmente incertidumbre económica producto de un proceso electoral.

La materia posee una competencia de ciencias de datos con el objetivo de construir el mejor modelo predictivo que resuelva un problema de forecasting de ventas al canal de distribución de productos de consumo masivo, utilizando la plataforma de competencias Kaggle.

La metodología es la de seminarios, deberá asistir a cada clase con los experimentos preparados y exponer en la misma los avances logrados la última semana y someter sus decisiones a un intenso intercambio de ideas con toda la cohorte.

El centro de esta asignatura es usted formando con sus pares una comunidad de aprendizaje colaborativo.

1.1 Sobre este documento

No imprima en papel este documento; está solo completo tan solo hasta el capítulo 4, qué es lo que necesita para iniciar la materia. Se le irá agregando contenido los próximos días a los primeros capítulos en función de la forma en que evolucione la dinámica del grupo.

1.2 Links fundamentales

UAustral Laboratorio de Implementación III , año 2024 virtual Links Fundamentales	
Zulip	https://austral2024.zulip.rebelare.com/
repositorio GitHub oficial	https://github.com/labo-imp/labo3-2024v
Invitación a la Competencia Kaggle	https://www.kaggle.com/t/b672b465347e4e2c83df0e23417a4c5c
Datasets Competencia primera entrega	https://storage.googleapis.com/open-courses/austral2024-fc72/sell-in.txt.gz https://storage.googleapis.com/open-courses/austral2024-fc72/tb_productos.txt https://storage.googleapis.com/open-courses/austral2024-fc72/tb_stocks.txt.gz
Google Sheet Colaborativa	https://docs.google.com/spreadsheets/d/1ngESOPFDKxeYS4oqmuxC6a7GO2QIVj7AW6UMZAtZN3c/edit?usp=sharing
Google Docs Colaborativo	TBD

1.3 UAustral Cronograma de la Asignatura año 2024

15-mayo miercoles	17:00	email	Envio por email de este documento a alumnos
15-mayo miercoles	18:00 a 22:00	Zoom	Clase 1 Presentación de la asignatura y su particular dinámica Planteo del Problema
22-mayo miercoles	18:00 a 22:00	Zoom	Clase 2 Primeras Presentaciones Estado del Arte + Análisis Exploratorio de Datos, Determinación de la Sub- Especialidad de cada grupo Estrategias generales de resolución Problemas encontrados en los Datos
29-mayo miercoles	18:00 a 22:00	Zoom	Clase 3 Primeras predicciones de cada Sub-Especialidad primeras grandes sorpresas
05-junio miercoles	18:00 a 22:00	Zoom	Clase 4
11-junio martes	18:00 a 22:00	Zoom	Clase 5
16-junio domingo	'23:59	Kaggle	<i>CheckPoint Kaggle</i>
18-junio martes	18:00 a 22:00	Zoom	<i>Clase 6</i>
25-junio martes	18:00 a 22:00	Zoom	Clase 7
02-julio martes	18:00 a 22:00	Zoom	Clase 8
14-julio domingo	'23:58	Google	Fecha límite cierre de Informe de Prediccion
		Zulip	Fecha límite de disponibilización GitHub con solución reproducible.
	23:59	Kaggle	Cierre automático de Competencia Kaggle
31-julio miercoles	11:59:00 PM	Zulip	Entrega de notas finales Se determina quienes pasan a recuperatorio.

1.4 Organización en Grupos

En la dinámica de la materia los alumnos se organizarán en grupos de tres personas con la excepción de que si la cantidad de alumnos no es múltiplo de tres, podrán haber hasta dos grupos de cuatro integrantes.

Dado que esta asignatura necesita de un debate amplio de los temas, con el propósito de romper grupos pre-existentes se ha dividido a la cohorte en dos mitades, la mitad naranja y la azul. Usted podrá formar grupo unicamente con personas de su mismo color, sin excepción.

ChatGPT 3.5 ▾



Tú

Create a random prime number with 11 digits



ChatGPT

Sure, here's a random prime number with 11 digits:

96005236357



Programa en lenguaje Julia `alumnos_particion.jl`

```
using CSV, DataFrames
using Random, StatsBase

dataset = DataFrame(CSV.File("alumnos.txt"))
mitad = trunc{Int, nrow(dataset)/2}

Random.seed!( 96005236357 )
naranja = sample(dataset[ :, :id], mitad, replace =false )

dataset[ [ x in naranja for x in dataset[:id]] , :]
```

El dataset ordenado alfabeticamente `alumnos.txt` y el script `alumnos_particion.jl` pueden encontrarse en el repositorio oficial de la cátedra, `./src/ParticionAlumnos`

La salida es la siguiente : mitad naranja

```
julia> dataset[ [ x in naranja for x in dataset[:id]] , :]  
28×2 DataFrame  
┌ Row │ id │ alumno │  
├───┼───┼────────┴───┤  
│ 1 │ 3 │ Bassan, Micaela │  
│ 2 │ 8 │ Canossini, Stefano │  
│ 3 │ 10 │ Carreño Patino, Ludwing Bernart │  
│ 4 │ 11 │ Cerutti, Leandro │  
│ 5 │ 12 │ Chavarria, Aureliano │  
│ 6 │ 15 │ Distefano, Marcela │  
│ 7 │ 18 │ Elia, Esteban │  
│ 8 │ 20 │ Falcones García, Johanna Estefanía │  
│ 9 │ 24 │ García, Matías │  
│ 10 │ 25 │ García Rio, Veronica │  
│ 11 │ 28 │ Gerschenfeld , Martín │  
│ 12 │ 29 │ Gimenez, Rafael │  
│ 13 │ 31 │ Ifran, Hernán │  
│ 14 │ 33 │ Ku Cuadra, Eduardo │  
│ 15 │ 35 │ Larregui, Gastón │  
│ 16 │ 42 │ Montes de Oca, Jose Andres │  
│ 17 │ 43 │ Monzon, Fernanda María │  
│ 18 │ 45 │ Núñez, Patricia Graciela │  
│ 19 │ 46 │ Orduz, Mónica │  
│ 20 │ 47 │ Piccin, Germán │  
│ 21 │ 48 │ Pijuan, Guillermo │  
│ 22 │ 49 │ Puyuelo, Corina │  
│ 23 │ 50 │ Reyes, Alejandra │  
│ 24 │ 51 │ Ríos, Soledad │  
│ 25 │ 52 │ Rodriguez, José Antonio │  
│ 26 │ 53 │ Rodríguez Saá, Milagros │  
│ 27 │ 55 │ Santamaría, César Germán │  
│ 28 │ 56 │ Valdés Castro, José Eduardo │
```

Tal como está especificado en la planilla colaborativa, los temas se dividirán en :

- Estadística Clásica : 2 grupos
- Auto Machine Learning : 3 grupos
- Gradient Boosting of Decision Trees : 6 grupos
- Deep Learning : 7 grupos

Los grupos se organizarán durante la primera clase o durante las primeras 24 horas posteriores, ya que deberán

1.5 Modalidad y Criterios de Aprobación y Evaluación

La nota mínima de aprobación de la materia es 5.50 , la nota final oficial se redondea a un número entero, la máxima nota final posible es 10.

Cada alumno en forma completamente individual puede parametrizar como desea ser evaluado en la materia.

Deberá cargar los porcentajes de contribución de la nota, en la Planilla Colaborativa de la Asignatura, antes de las 23:59 del día que se lleve a cabo la cuarta clase. En caso de no cargar nada, se le asignarán los porcentajes default.

Nota Ordinaria			
Contribución Nota			Actividades Obligatorias
Min	Default	Max	
15%	20%	25%	Discusiones en Clase, tarea Individual
1%	5%	10%	Exposiciones en Clase, tarea grupal
10%	15%	20%	Informe Final, tarea grupal
0%	10%	15%	Video Modelo Final, tarea grupal
30%	44%	74%	Competencia en la plataforma Kaggle, más scripts y experimentos en GitHub reproducibles. CheckPoint 20% nota kaggle Final 80% nota kaggle Tarea Grupal
0%	5%	15%	Participación <i>significativa</i> en Zulip. tarea Individual
1%	1%	1%	Juez en Torneo de Videos tarea individual

Todas las actividades anteriores con una contribución mínima mayor a 0% son obligatorias, la no participación/entrega en cualquiera de ellas implica el desaprobación de la materia y pasar directamente a recuperatorio.

El total de los porcentajes elegidos debe sumar 100.

Alumnos de un mismo grupo si lo desean pueden elegir porcentajes completamente distintos.

1.5.1 Modalidad y Criterios de Aprobación y Evaluación Competencia Kaggle

La competencia Kaggle de la asignatura se califica en función de la ganancia obtenida en el Private Leaderboard. Debe quedar claro que para la calificación no se tiene en cuenta el Public Leaderboard en absoluto.

Es una tarea grupal y la nota es grupal, pero cada alumno del grupo puede elegir en forma independiente a sus compañeros de grupo el porcentaje que Kaggle contribuirá a su nota individual.

Los alumnos deberán ya desde la primer semana formar grupo en Kaggle y quedarán limitados a los 20 submits diarios por grupo. No se podrán crear usuarios adicionales a la competencia. La plataforma Kaggle no admite divorcios.

Kaggle por default elige la predicción que más ganancia obtiene en el Public Leaderboard, pero esto puede y con altísima probabilidad deberá ser modificado por el alumno, el que elegirá el modelo que a pesar de no ser el de más ganancia en el Public Leaderboard a su entender es el que más ganancia obtendrá en el privado.

Habrán dos instancias Kaggle, utilizando la misma competencia:

1. Checkpoint a mitad del curso, aportará el 20% de la nota Kaggle
2. Final de la competencia, aportará el 80% de la nota Kaggle

Para la nota final de cada una de las instancias, las ganancias del Private Leaderboard se normalizarán y se transformarán al intervalo $[0,10]$ de nota.

Es parte de la filosofía de la materia la reproducibilidad de los experimentos. Su repositorio GitHub de la materia deberá permanecer completamente público y abierto, accesible para los profesores y sus compañeros, durante todo el transcurso de la asignatura hasta la entrega de las notas finales. Usted irá recibiendo feedback público en Zulip sobre sus scripts y parametrizaciones.

Como parte de la entrega usted deberá disponibilizar en su repositorio GitHub, en forma pública bajo una carpeta que creará exclusivamente a ese propósito, los scripts y archivos de parámetros que permitan generar la solución definitiva que está entregando en Kaggle a partir del dataset original. Los profesores tienen que ser capaces de correr esos scripts y generar exactamente el archivo que usted ha subido a Kaggle como entrega final.

En caso que los profesores posean inquietudes sobre su trabajo grupal o la participación individual de algún alumno, cada integrante del grupo pasará a una etapa de evaluación oral individual con live coding.

1.5.2 Exposiciones en Clase

Durante cada clase sincrónica, algunos grupos expondrán los avances en la construcción del modelo predictivo según la subespecialidad elegida, los problemas encontrados y las soluciones que proponen o ya han implementado.

1.5.3 Discusiones en Clase

Una vez que un equipo expone durante la clase se inician las discusiones (~ intercambio de ideas) que es abierta a toda la cohorte.-

1.5.4 Informe final

Para tomar buenas decisiones en la construcción de un excelente modelo predictivo se deben llevar a cabo una gran cantidad de experimentos para entender qué es lo que mejor funciona, en algunos casos solo para este dataset, en otros para la mayoría de los datasets.

Muchas veces esa necesaria cantidad de experimentos puede resultar abrumadora para quienes están enfrentando por primera vez un problema de dimensiones reales. Otras veces resolver una pregunta sobre la mejor decisión, debido a la naturaleza probabilística de los modelos predictivos requiere de repetir el experimento cambiando las semillas de los generadores de números pseudoaleatorios. En la asignatura vamos a repartir la carga de esta experimentación de forma colaborativa entre todos los alumnos, obteniendo conclusiones más robustas y reproducidas por pares.

Los instructores plantean problemas del modelado predictivo, de capital relevancia para el problema concreto propuesto en la asignatura, propondrán hipótesis experimentales y asignan a distintos grupos de alumnos a las tareas de diseñar y ejecutar, reproducir, presentar y revisar los experimentos, los que se presentarán en formato de video y se discutirán en clase y Zulip.

Los experimentos se vuelcan en un documento compartido accesible a todos los alumnos y constituyen la base para la toma de decisiones en la confección de un buen modelo predictivo final.

Es muy importante realizar una completa búsqueda bibliográfica para el experimento; lo que resulta un verdadero desafío para algunos experimentos.

Rúbrica de Informe Final	
Porcentaje	Concepto a Evaluar
10%	Hipótesis Experimental, será brindada por los profesores Los alumnos buscarán bibliografía que soporta, o no, la hipótesis experimental.
30%	Diseño experimental y materiales. El diseño del experimento muestra una acabada comprensión del problema a resolver, los experimentos están precisamente especificados y detallados, los scripts son adecuados.
20%	Procedimientos y recopilación de resultados La ejecución de los experimentos es correcta, los parámetros iniciales y resultados capturados son los adecuados.
20%	Análisis de los resultados. Se interpretan correctamente los resultados obtenidos
20%	Resultados, conclusiones y discusión con pares La conclusión determina claramente si la hipótesis experimental se cumple o no. Se presentan los resultados en forma clara y convincente, se debaten los mismos con el resto de los alumnos.

	Se relacionan los resultados con la bibliografía original. Se proponen nuevos experimentos en caso de cuestionamientos externos. Se responden adecuadamente las preguntas de los pares
--	--

Los formatos preferidos para los videos son: YouTube, Prezi, Loom y Twitch. No se acepta la entrega de links a Google Drive o Dropbox. El video debe ser accesible para todos por lo menos hasta seis meses después de haber sido entregadas las notas, para que tengan acceso a todos ellos los alumnos que deben realizar el recuperatorio de la materia. No es necesario que el video sea público, solamente podremos acceder los que tengamos el link y no podrá ser buscado en las redes.

Es parte fundamental de la tarea que los alumnos investiguen la forma de hacer una video presentación efectiva e intercambien ideas en Zulip. Ese descubrimiento será asistido por los profesores.

1.5.5 Evaluación participación en foro/chat Zulip

Zulip es el sistema nervioso de la asignatura ya que contribuye a generar una comunidad. Pasan más cosas en Zulip que en las clases sincrónicas.

Ver el capítulo Zulip dentro de Herramientas, Conceptos y Operación el detalle de lo que se espera de usted en Zulip

Rúbrica de participación en foro Zulip	
Porcentaje	Concepto a Evaluar
30%	Contenido. Genera preguntas o respuestas profundas, significativas, claras, fáciles de entender y con potencial para el proyecto de la asignatura. <u>Compartiendo</u> resultados logra atraer la atención y <u>colaboración</u> de sus compañeros a sus posts. Idealmente se transforma en un líder conceptual del grupo.
30%	Contribución al dinamismo de la comunidad. Plantea preguntas interesantes y relevantes, intenta motivar las discusiones grupales sobre tópicos relevantes a la asignatura, debate positivamente, participa de conversaciones iniciadas por otros.
20%	Colaboración en dudas operativas, colabora rápida y acertadamente con compañeros que requieren algún tipo de asistencia operativa en alguna de las herramientas del curso, lenguaje de programación, etc. Idealmente se transforma en un referente en un tema específico del grupo.
20%	Frecuencia. Al final del curso deberá tener como mínimo treinta participaciones. Una consulta técnica cuenta como participación !

Cada alumno participa con su usuario en forma individual en Zulip

1.6 Misión de la Asignatura

Lograr que los alumnos sean capaces de resolver un problema de dimensiones reales del mercado argentino con una *excelencia* que sorprenda a sus pares.

1.6.1 Objetivos de la Asignatura

1. Resolver un problema de dimensiones reales del mercado local utilizando las herramientas tecnológicas para manejar grandes volúmenes de datos y ser capaz de generar un forecasting competitivo en una situación desafiante.
2. Desarrollar una fuerte cultura de la experimentación siendo capaz de cuestionar todo lo que le han enseñado en la asignatura Series Temporales y otras.
3. Desarrollar la creatividad para el Feature Engineering
4. Conocer y utilizar efectivamente las técnicas “estado del arte” en cuanto a algoritmos y librerías de última generación para el forecasting.

1.7 Unidades Temáticas

1. Nociones elementales de la actividad de consumo masivo, tipos de actores, producción, cadena de distribución e información disponible
2. Planteo del problema de negocios y métrica a optimizar
3. Metodología de evaluación de la bondad de ajuste de una predicción
 1. El problema de la fuga de datos (*Data Leakage*)
 2. Método de *Walk Forward Validation*
4. Breve repaso de métodos de la estadística clásica
 1. *Autoregressive and Moving Average Models* (ARIMA)
 2. Auto Arima
 3. Seasonal ARIMA (SARIMA), SARIMAX
5. Modelos *vector autoregression* (VAR), *vector moving average* (VMA , VARMA)
6. Modelos generalized autoregressive conditional heteroskedasticity (GARCH)
7. Métodos derivados de las redes neuronales profundas aplicados a las series de tiempo
 1. *CNN Convolutional Neural Networks*
 2. *RNN Recurrent Neural Networks*
 3. *LSTM Long short-term Memory*
8. Métodos propietarios para series de tiempo
 1. *Prophet* de la compañía Facebook
 2. *AWS Forecast* de la compañía Amazon
9. Gradient Boosting of Decision Trees aplicado a las series de tiempo
 1. Feature Engineering necesario.
10. Clustering de Series de Tiempo
 1. Método basado en la Transformada Discreta de Fourier
 2. Método de *Dynamic Time Warping*

1.8 Bibliografía General de la Asignatura

La siguiente es la bibliografía general; en cada clase se brindará bibliografía específica sobre los temas vistos.

Sitios con muy buenos artículos

- <https://machinelearningmastery.com/> muy precisos los artículos de Jason Brownlee
- <https://www.kdnuggets.com/>

Sitios populares fáciles de leer que hay que tomar con pinzas sus artículos ya que por lo general carecen de profundidad, y lo que es muy grave hay muchos con enormes errores conceptuales y prácticos:

- <https://towardsdatascience.com/>
- <https://medium.com/tag/data-science>

Papers y otros Libros

Aghabozorgi, S., Shirkhorshidi, A. S., and Wah, T. Y. (2015). Time-series clustering—a decade review. *Information Systems*, 53:16–38.

Athanasopoulos, G., Poskitt, D. S., & Vahid, F. (2012). Two canonical VARMA forms: Scalar component models vis-à-vis the echelon form. *Econometric Reviews*, 31(1), 60–83

Agrawal, R., Faloutsos, C., and Swami, A. (1993). Efficient similarity search in sequence databases. In *International conference on foundations of data organization and algorithms*, pages 69–84. Springer.

Bandara, K., Bergmeir, C., and Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert systems with applications*, 140:112896.

Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, USA

Chen, Tianqi, XGBoost: A Scalable Tree Boosting System, *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.

Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27(1), 134–141.

Folger, J. (2015). Backtesting and Forward Testing: The importance of correlation. Investopedia, 2022 update

Ke, Guolin, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *NIPS'17: Advances in Neural Information Processing Systems Conference*, 3148-3156, 2017

Kuhn, Max, Feature Engineering and Selection: A Practical Approach for Predictive Models *Chapman&Hall/CRC Data Science Series*, 2019

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1), 388–427.

Hoarau, Michaël: Time Series Analysis on AWS, Packt Publishing 2022

Hyndman, Rob, Forecasting: Principles and Practice, 3rd Edition, Otexts, 2021

Rafferty, Greg : Forecasting Time Series Data with Facebook Prophet, 2nd Edition Packt Publishing; 2023

Tadayon, M. and Iwashita, Y. (2020). A clustering approach to time series forecasting using neural networks: A comparative study on distance-based vs. feature-based clustering methods. *arXiv preprint arXiv:2001.09547*

Taylor, S. J.; and Letham, B. 2017. Forecasting at Scale, *Am Stat.* 2018;72(1):37–45, 2017

[Tingyu Weng, Wenyang Liu, Jun Xiao](#) Supply chain sales forecasting based on lightGBM and LSTM combination model, [Industrial Management & Data Systems](#), Vol. 120 No. 2, pp. 265-279
Yu Y, Si X, Hu C, Zhang J (2019) A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput* 31(7):1235–1270

Para la búsqueda de papers se recomienda primero buscarlos en Google Scholar
<https://scholar.google.com/> y si no está disponible bajarlo del inframundo <https://sci-hub.se/>

Página para bajar libros <https://libgen.rs/>

Herramienta para visualizar conexiones entre papers <https://www.connectedpapers.com/>

2 Metodología de Enseñanza

2.1 Generalidades

Basados en el [authentic learning](#) se plantea un problema de negocios de la vida real tal cual un alumno podría encontrar en un trabajo en Argentina, con una cantidad de datos real, y una forma de evaluación igual a la que sería evaluado en una empresa : ganancia en pesos argentinos de la campaña de retención de clientes y exposición con storytelling de los resultados.

La asignatura se dicta con el espíritu de la [teoría del conectivismo](#), donde los alumnos deben crear colaborativamente contenido, reflexionar sobre contenido de sus pares, discutirlo y generar nuevo contenido.

La dinámica de las clases es la del [flipped classroom](#) (aula invertida) donde los alumnos ven los materiales *más simples* asincrónicamente **antes** de la clase sincrónica, y una vez en la clase sincrónica se utiliza una dinámica de [active learning](#) basado en pares con algunas exposiciones del profesor para los temas más complejos. Dado que los alumnos pueden consultar por Zulip en forma asincrónica antes de la clase, técnicamente el nombre correcto de la metodología es *active learning forte*.

2.2 Manifiesto Pedagógico

Si una sola frase debiera resumir la pedagogía de esta asignatura sería la expresión de Plutarco:

Enseñar, más que llenar un recipiente es encender un fuego.

Las filosofías pedagógicas se pueden dividir a grandes rasgos en estas teorías antagónicas:

- el conductismo- instructivismo de Edward Thorndike y B. F. Skinner que es la predominante en el mundo y Argentina
- el constructivismo social de John Dewey y Seymour Papert que será utilizada en esta asignatura.

El estilo en esta asignatura es el del *aprendizaje basado en pares*, en donde a pesar de existir un sendero diseñado por el profesor se alienta continuamente a los alumnos a recorrer su propio camino, observar y aprender de sus pares, diseñar creativamente sus propios experimentos reflexionando profundamente sobre los resultados que se van obteniendo, cuestionando al establishment. Como referencia se puede tener la filosofía y comunidad que existe alrededor del lenguaje de programación Scratch y los conceptos de experiencias educativas de Mitchel Resnick.

La asignatura es una experiencia de gran intensidad intelectual y emocional que busca constantemente alentar a que cada alumno supere lo que se espera de él; se intenta lograr un efecto de resonancia grupal que amplifique el aprendizaje.

El estilo de dictado está basado en la frase "la emoción es el timón de la razón" donde cada tema es presentado de forma que genere gran emocionalidad en los alumnos, por momentos usted sentirá profundo amor y fascinación por las ideas que se estarán viendo, en otros literalmente sentirá odio.

Finalmente, aunque no menos importante, la asignatura busca la reflexión profunda y el cuestionamiento crítico de lo enseñado en otras asignaturas e incluso en ella misma. Citando a Richard Feynman

"The problem is not people being uneducated. The problem is that people are educated just enough to believe what they have been taught, and not educated enough to question anything from what they have been taught".

Dos videos y un artículo que reflejan aspectos de la forma de enseñanza

<https://www.youtube.com/watch?v=uRxD-pe3PN0>

Mitchel Resnick, Lifelong Kindergarten Group,

Media Lab, MIT

<https://web.media.mit.edu/~mres/papers/constructionism-2014.pdf>

Mitchel Resnick, Constructionism

<https://www.youtube.com/watch?v=9vp824ZbksI>

George Siemens, Connectivism

2.3 Pensamiento Creativo - Tinkering

Para ilustrar lo que se espera del alumno en esta asignatura, van unos párrafos del libro Resnick, Mitchel , *Lifelong Kindergarten: Cultivating Creativity through Projects, Passion, Peers, and Play*, The MIT Press, 2018

One of the students, named Nicky, started by building a car out of LEGO bricks. After racing the car down a ramp several times, Nicky added a motor to the car and connected it to the computer. When he programmed the motor to turn on, the car moved forward a bit—but then the motor fell off the body of the car and began vibrating across the table on its own.

Rather than trying to repair the car, Nicky became intrigued with the vibration of the motor. He played and experimented with the vibrating motor, and began to wonder whether he might be able to use the vibrations to power a vehicle. Nicky mounted the motor on a platform atop four “legs” (LEGO axles).

After some experimentation. He realized that he needed some way to amplify the motor vibrations.

To do that, he drew upon some personal experiences. Nicky enjoyed riding a skateboard, and he remembered that swinging his arms gave him an extra push on the skateboard. He figured that a swinging arm might accentuate the vibrations of the motor as well, so he connected two LEGO axles with a hinged joint to create an arm and attached it to the motor. As the motor turned, the arm whipped around—and amplified the motor vibrations, just as Nicky had hoped. In fact, the system vibrated so strongly that it frequently tipped over.

A classmate suggested that Nicky create a more stable base by placing a LEGO tire horizontally at the bottom of each leg. Nicky made the revision, and his “vibrating walker” worked perfectly. Nicky was even able to steer the walker. When he programmed the motor to turn in one direction, the walker vibrated forward and to the right. When he programmed the motor to turn in the other direction, the walker vibrated forward and to the left.

I was impressed with Nicky’s vibrating walker—but even more impressed by the strategies he used in creating it. As Nicky worked on his project, he was constantly tinkering. Throughout the process, he was playfully experimenting, trying out new ideas, reassessing his goals, making refinements, and imagining new possibilities.

Like all good tinkerers, Nicky was:

- **Taking advantage of the unexpected.** When the motor fell off of his car, Nicky didn’t see it as a sign of failure; he saw it as an opportunity for new explorations.
- **Drawing on personal experience.** When Nicky needed to amplify the vibrations of the motor, he relied on his experiences as a skateboarder and knowledge of his own body.
- **Using familiar materials in unfamiliar ways.** Most people don’t imagine LEGO axles as arms or legs, nor do they imagine LEGO wheels as feet—but Nicky was able to look at objects in the world around him and see them in new ways.

Usted no jugará con bloques de un LEGO , sino que configurará y reconectará scripts y bloques de código brindados por la cátedra, y si tiene la fortuna de ser atrapado por el entusiasmo y motivación de Nicky, desarmará scripts y creará algo totalmente innovador.

2.4 *Dinámica Constructivista en el Aula*

En caso de preguntarse cómo concretamente difieren el conductismo del constructivismo, aquí va un ejemplo de dinámica sobre como medir para niños de 6 años

- https://storage.googleapis.com/open-courses/videos-d0a6/whale_measuring.mp4 (4 minutos)
- paper <http://www.gphillymath.org/ConstructivistLearn/Constructivist.pdf>

Sin embargo esta dinámica en el aula puede ser muy difícil para algunos alumnos adulto, ya que suele estar acostumbrado a la "Educación Colonial" https://storage.googleapis.com/open-courses/videos-d0a6/colonial_education.mp4 (1 minuto) y esperar "recetas" del profesor.

2.5 *Authentic Learning*

El enfoque pedagógico de esta asignatura está fuertemente relacionado con el *Authentic Learning* e indirectamente al Constructivismo Social en cuanto a:

- El aprendizaje comienza con el planteo de un problema del mundo real, de dimensiones reales, que la mayoría de alumnos muy probablemente deban enfrentar su actividad profesional.
- Algunos aspectos del problema están intencionalmente definidos de forma difusa para que los alumnos reflexionen profundamente sobre las mejores soluciones posibles.
- La evaluación de la asignatura refleja la evaluación del mundo real profesional, basada en la rentabilidad, comunicación de resultados y colaboración con pares.
- La teoría está al servicio de la práctica, y aparece justo a tiempo; jamás a la inversa.
- El profesor colabora con *Instructional Scaffolding* ayudando a escalar los distintos niveles de complejidad conceptual.

2.6 *Flipped Classroom Forte*

La idea es que el alumno reciba la ayuda directa del profesor en los momentos de aprendizaje que más lo necesita que generalmente es cuando el alumno se enfrenta a resolver un problema en soledad (no en el aula); para generar esa disponibilidad de ese tiempo lo fácil, la transmisión de información, se pasa para antes de la clase en modalidad asincrónica, cuando los profesores no están disponibles en tiempo real.

Además se evalúa lo asincrónico con quizzes y con las subidas a Kaggle, permitiendo a los profesores responder por el foro/chat Zulip preguntas simples, y dejar lo que no ha sido entendido para reforzarlo en la siguiente clase presencial. Los quizzes muestran que se está entendiendo y que no, por más que el alumno crea que está entendiendo todo lo que lee.

Una presentación sobre Flipped Learning Forte clara, concisa y carente del usual laberinto literario de las ciencias sociales, es <https://www.slideshare.net/alfredo.prietomartin/intercambio-hispano-chileno-de-experiencias-de-flipped-learning-en-enseanza-blended-y-online> junto con este artículo

<https://revistaventanaabierta.es/flipped-classroom-sus-limitaciones-y-aparicion-de-una-nueva-variante-flipped-learning-forte/>

2.7 Conectivismo

Las ideas de teoría del aprendizaje del conectivismo tienen gran semejanza con el deep learning, <https://www.youtube.com/watch?v=yx5VHpaW8sQ> (3 min)

- El aprendizaje ocurre a través de la red neuronal del cerebro, redes de humanos y redes de entidades no humanas.
- La red de conocimiento se desarrolla a partir de las interacciones con otras entidades de la red y el mundo en general. El objetivo de la enseñanza es estimular dichas interacciones.
- La red crece:
 - nodos
 - incorporando nuevos nodos
 - eliminando nodos ya no útiles
 - conexiones entre nodos
 - creando conexiones entre nodos que antes no estaban conectados
 - reforzando la conexión entre dos nodos, al ser transitado frecuentemente
 - disminuyendo el peso de la conexión, al ser poco transitado ese link
- El conocimiento reside en los pesos de las conexiones
- Aprender es el proceso de conectar nodos específicos o fuentes de información
- **La capacidad de aprender cosas nuevas es más importante de lo que se ya se conoce.**
- Tomar decisiones es un proceso de aprendizaje, de transitar reiteradamente y en formas novedosas la red

2.8 Active Learning

Active Learning significa que se llevará al mínimo lo que en la jerga se denomina "clase magistral" que es básicamente un profesor hablando y los alumnos escuchando; los alumnos se deberán involucrar activamente durante la clase con actividades diseñadas y guiadas por el profesor, en donde deban reflexionar profundamente sobre el tema que se está tratando.

Active Learning es *la metodología* en la que toda la comunidad educativa está de acuerdo que brinda resultados superadores. Puede verse este artículo "Active learning increases student performance in science, engineering, and mathematics" <https://www.pnas.org/doi/full/10.1073/pnas.1319030111>

Es tan distinto a la mecánica tradicional que muchos alumnos se sentirán incómodos al comienzo porque jamás han sido expuestos a la misma. No se confunda, active learning no tiene que ver con ser extrovertido, sino con ser reflexivo !

Videos cortos que explican los fundamentos del active learning:

https://www.youtube.com/watch?v=zoa2pKYp_fk

(5 min) Janet Rankin, MIT

<https://www.youtube.com/watch?v=Z9orbxoRofI>

(14 min) Eric Mazur, Harvard University

Un artículo sobre aulas de active learning <https://er.educause.edu/articles/2017/12/creating-active-learning-classrooms-is-not-enough-lessons-from-two-case-studies>

Esta es una lista de ejemplo *no vinculante* del tipo de actividades que se llevarán a cabo (ir a Large Group) https://www.queensu.ca/teachingandlearning/modules/active/12_exampls_of_active_learning_activities.html siendo la actividad más tradicional el Think Pair Share que es la que habla Eric Mazur

Al siguiente extenso video es una clase en el MIT donde en un posgrado de educación la profesora demuestra a los alumnos diversas dinámicas de active learning <https://www.youtube.com/watch?v=hGBNi4P9OfA> , no hace falta que lo vea, ya lo experimentará.

3 Arranque en Frío

En la materia trabajaremos con una gran variedad de herramientas, plataformas y conjuntos de datos, las que deben ser instaladas, configuradas, interconectadas y finalmente usted deberá aprender a utilizarlas. No se sienta desbordado por el shock inicial, aprenderá a utilizarlas. En palabras de un ex alumno "todas las semanas tuve en simultáneo la frustración de no entender alguna implementación y la satisfacción de dominar lo que una semana atrás parecía imposible."

Intente avanzar por su cuenta con los siguientes pasos de instalación, aunque no entienda en detalle lo que está haciendo, ya se le explicará el uso de las herramientas en clase.

En caso de tener dificultades con algún punto, o consúltelo en Zulip o espere a la primera clase en donde se verán conceptualmente estos pasos.

Palabras de un ex alumno : "Unos pocos empezamos la maestría con PC de Empresa con restricciones, bloqueos y rigideces. La fuimos "piloteando" pero cada vez se fue haciendo más cuesta arriba... Creo que en esta materia finalmente todos terminamos con PC libre. En mi caso fue un salto cuántico de comodidad."

3.1 Alta en Plataformas

3.1.1 Zulip herramienta de chat y foros discusión

En la materia no se utilizará el email, toda la comunicación y discusión se llevará a cabo en Zulip. Los profesores serán por lejos quienes más participen en las conversaciones de Zulip.

Para darse de alta en el Zulip de la materia usted debe ingresar al link de Zulip de la sección 1.1 Links Fundamentales y Regístrese (abajo a la derecha del recuadro)

Usted puede acceder a Zulip de tres formas:

- En su PC desde algún browser
- En su PC utilizando la app, instálela desde <https://zulip.com/apps/>
- Desde su smartphone, busque la app en su *App Store* (alerta : no es buena la app)

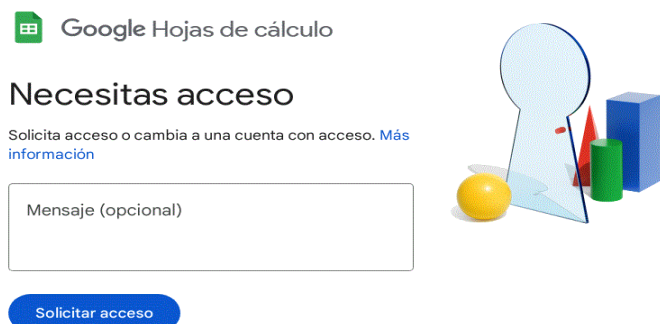
En el capítulo Herramientas, Conceptos y Operación se le enseñará cómo utilizar efectivamente Zulip a lo largo de la asignatura, ya que parte de la nota final depende de ello.

Una vez dentro de Zulip, vaya a la rueda dentada ubicada arriba a la derecha, Manage streams, elija el stream cartelera, y a la derecha marque Email notifications

3.1.2 Google Sheet Colaborativa de la Asignatura

Para proceder con este paso es necesario que usted disponga de una cuenta de gmail la que además le será indispensable para trabajar con la plataforma Google Cloud.

Siga el link de Google Sheet Colaborativa de la sección 1.2 Links Fundamentales de este documento. Le aparecerá una pantalla de este estilo



presione el botón de Solicitar acceso , a las pocas horas recibirá un email en su gmail informándole que ya tiene acceso a la planilla colaborativa, en donde deberá ir cargando resultados de las corridas que realice a lo largo de la asignatura y podrá comparar sus resultados con los de sus compañeros.

3.1.3 Plataforma Hypothes.is

Hypothes.is es una herramienta de anotación colaborativa de páginas web y documentos .pdf, que también utilizaremos para anotar scripts y Jupyter Notebooks . Le permitirá ver anotaciones que el profesor ha realizado en papers a utilizar en la asignatura, y usted mismo podrá agregar anotaciones.

La siguiente es una razonable introducción a Hypothes.is <https://www.youtube.com/watch?v=87h0nYi-i9o> presta atención a cómo se hacen las anotaciones y a cómo se participa en una discusión sobre una anotación.

Ir a la página <https://web.hypothes.is/start/> y seguir los pasos para crear un usuario, se recomienda enfáticamente crear un usuario con el formato que sea <su nombre> punto <su apellido> . Adicionalmente instale la extensión para el navegador Google Chrome, que le será indispensable para la anotación de documentos .pdf

No haremos públicas las anotaciones en la asignatura, sino que las acotaremos al grupo privado [uaustral](#), para sumarse a estos grupos estando logueado a Hypothes . is en su browser y luego seguir este link <https://hypothes.is/groups/E6iXgnWN/uaustral>

3.1.4 Plataforma ChatGPT

Utilizaremos el novedoso ChatGPT para que escriba scripts simples por nosotros.

Ingresa a <https://chat.openai.com/auth/login> y cree un usuario gratuito. Dada la alta demanda de ChatGPT es posible que deba probar en distintos horarios hasta lograrlo.

3.1.5 Plataforma Kaggle

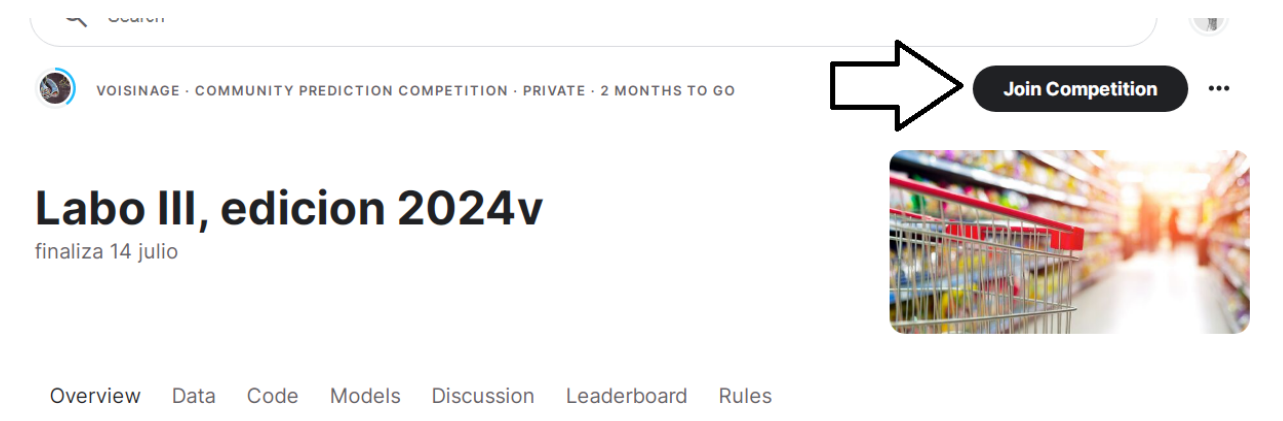
Utilizaremos la plataforma Kaggle para alojar la Competencia Kaggle de la Materia la que será cerrada, limitada a los alumnos de la materia que deberán registrarse en Kaggle con el email de Universidad Austral.

Kaggle es, entre otras cosas, una plataforma que permite participar de competencias mundiales de ciencias de datos. Es muy útil para el dictado de cursos ya que genera un notable involucramiento de los alumnos, los que todo el tiempo van teniendo feedback de qué tan buenos son los modelos predictivos que van probando en comparación a los compañeros.

Si no posee un usuario de Kaggle deberá registrarse a la plataforma Kaggle utilizando este link <https://www.kaggle.com/account/login?phase=startRegisterTab> se sugiere que para ello utilice su verdadero nombre (no un pseudónimo).

Anote el nombre de usuario ya que lo requerimos más adelante.

Una vez que ya posea un usuario de Kaggle, deberá registrarse en la competencia, para ello debe seguir el link de invitación a la competencia del capítulo *1.2 Links Fundamentales* y luego presionar el botón negro que dice Join Competition.



Preste atención a que hay dos links de Kaggle, uno es la invitación a la competencia, que será necesario solo la primera vez, y el otro link es con el que ingresará asiduamente.

Lea con atención la sección Overview del menú principal, que contiene las subsecciones Description, Evaluation, Evaluation Details, Q&A y Past Opinions.

Una vez registrado en Kaggle con su número telefónico verificado, vaya a <https://www.kaggle.com/settings/account> y presione

Settings

Control over your Kaggle account and all communications

Account Notifications

Your email address

gustavo.denicolay@gmail.com

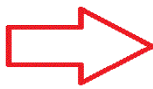
Change email

Phone verification

Verified

API

Using Kaggle's beta API, you can interact with Competitions and Datasets and more via the command line. [Read the docs](#)



Create New Token

Expire Token

Quotas

Guarde en un lugar seguro el archivo que se bajará, contiene su usuario Kaggle y una key. Lo necesitará en la instalación de Google Cloud, para poder subir en forma automática submits.

3.1.6 Plataforma GitHub

Para poder compartir nuestros scripts y experimentos habrá un repositorio oficial de la asignatura y uno suyo, ambos residirán en la nube. Para ello usted dará de alta una cuenta en los servicios de la plataforma GitHub

Registrarse en https://github.com/signup?ref_cta=Sign+up

Una ayuda simple es <https://www.classicpress.net/github-desktop-a-really-really-simple-tutorial/>

Una vez ya registrado y dentro de su sesión de GitHub, es momento de crear su Personal Access Token que será algo así como su usuario y clave para acceder desde su PC local y desde Google Cloud.

Dentro de su sesión de GitHub vaya a la página <https://github.com/settings/tokens/new> ,

- donde dice Note escriba "mi primer token",
- en Expiration marque Custom y elija como fecha el 31-dic-2024 (realmente es muy recomendable que ponga una fecha de expiración)
- luego en Select Scopes marque TODOS los recuadros que le aparecen

- al final de la página presione el botón **Generate Token**

Le aparecerá una pantalla como esta:


Personal access tokens

Generate new token

Revoke all

Tokens you have generated that can be used to access the [GitHub API](#).

Make sure to copy your personal access token now. You won't be able to see it again!

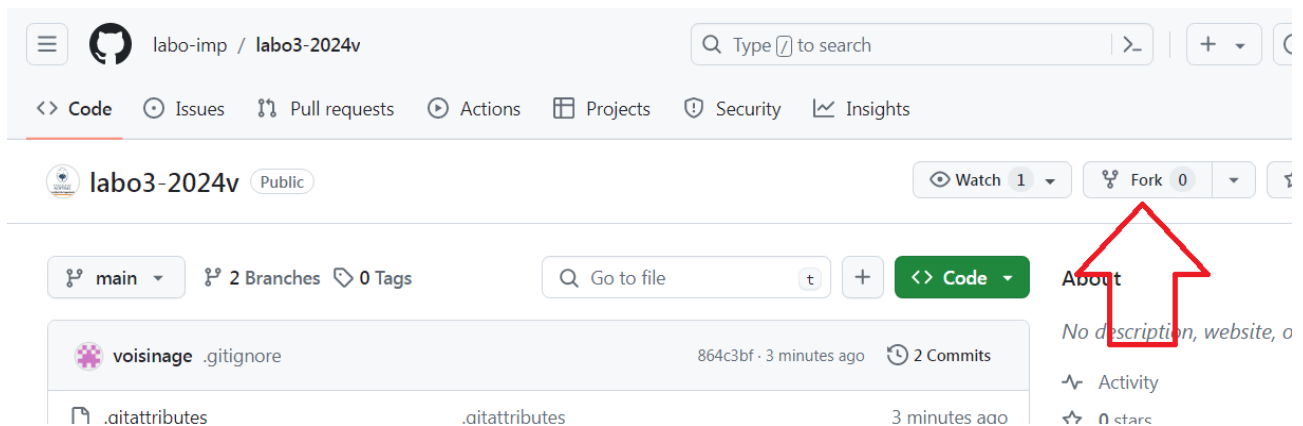
✓ ghp_bG7T3t7vMRrUHlbrMYIVHzJnOJFDs11pCGxz 

Delete

allí copie y guarde en un lugar seguro el token, que en el ejemplo de la imagen es el string `ghp_bG7T3t7vMRrUHlbrMYIVHzJnOJFDs11pCGxz`

usted volverá a utilizar ese token cuando instale Google Cloud, guárdelo en un lugar seguro por favor

Una vez hecho el login a su GitHub usted no tendrá ningún repositorio creado inicialmente, ir al repositorio GitHub de la materia cuyo link está en el capítulo *Links Fundamentales* y hacer un fork del repositorio presionando el botón que está arriba a la derecha de la página de GitHub .



Le aparecerá un formulario para crear el fork

De esta forma usted pasará a tener en su cuenta de GitHub una copia del repositorio oficial de la asignatura. Ambos están en la nube, pero usted no puede trabajar directamente con esos archivos en la nube sino que debe generar una copia en su computadora de su repositorio GitHub local.

4 Herramientas, Conceptos, Operación y Buenas Prácticas

4.1 Zulip

Zulip es el sistema nervioso de la asignatura. Pasan más cosas en Zulip que en las clases presenciales.

Utilizaremos Zulip para

- Solicitar y brindar soporte técnico.
- Mantener discusiones en foros, asincrónicos fuera de las clases.
- Intercambiar mensajes, links, artículos, archivos entre alumnos entre sí, alumnos y profesores
- quizzes en la clase sincrónica y también en la tarea para el hogar asincrónica
- Cartelera de información de la asignatura
- resultados de Experimentos Colaborativos
- Entregas de la materia

Zulip es una aplicación, open source, de chat y foros de discusión con videoconferencia integrada, que permite organizar las conversaciones en *streams* y *topics* posibilitando mantener intercambio de ideas en paralelo, lo que es una gran ventaja si no se está online todo el tiempo. La existencia de hilos de conversación (*streams*) y tópicos permite una eficiente conversación no lineal y asincrónica.

Zulip es open source y está corriendo en un servidor propio, Slack no es open source. Zulip no está tan pulido y es menos intuitivo que Slack. Sin embargo Slack tiene la contra que revisar canales con mucha actividad se vuelve muy tedioso ya que no se dividen en tópicos.

Zulip permite enviar estos tipos de mensajes :

- privados a otro usuario o grupo de usuarios o a un stream privado
- públicos, a streams públicos, que actúan como salas de chat/foros de discusión

Todas las preguntas a profesores, sobre cualquier tema de la asignatura, deben ser siempre públicas. Es mandatoria esta transparencia. Escribir mensajes privados es caer en el oscurantismo de los emails, es dejar de compartir con el resto y evita que se pueda leer lo que otros están pensando.

Lo más probable es que usted acceda a Zulip desde su computadora y además tenga instalada la app en su smartphone (Android y iOS están soportados)

Zulip como aplicación permite esto <https://zulip.com/features/> , una relativa "falta" de Zulip es que no permite enviar mensajes de voz en forma nativa.

Un video que muestra como utilizar Zulip es <https://www.youtube.com/watch?v=xWa56KdgYZM>

Stream	Utilidad
3-La Multinacional responde	Preguntas dirigidas a La Multinacional, responderá las que por privacidad pueda, utilizando su particular lenguaje alejado del tecnicismo de la ciencia de datos.
3-cartelera	alumnos no pueden enviar mensajes, profesores anuncian <ul style="list-style-type: none"> • recordatorios fechas importantes • commits del repositorio oficial • actualizaciones Libro de la Asignatura • bibliografía nueva • comunicaciones generales
3-Clase n	Cada clase tendrá su correspondiente stream, y estos dos tópicos estarán siempre presentes: <ul style="list-style-type: none"> • antes : de que tratará la clase y que debo traer hecho/leído • despues: resumen de lo que fue visto en clase, de las conclusiones
3-Feature Engineering	Intercambio de ideas sobre la creación de atributos nuevos
3-Estrategias	Discusión de estrategias sobre cómo mejorar los modelos predictivos. Temas principalmente conceptuales, pero también aplicados.
v-Exp Colaborativos	Intercambio Experimentos Colaborativos
Python	Consultas de Python
off-topíc	estados de ánimo generales, memes generales, todo lo que no valdría la pena ir a buscar a los threads oficiales
3-entregafinal	será utilizado para entregar los videos

Los streams en fondo amarillo serán habilitados a su debido momento.

Si usted desea que un profesor preste atención a su mensaje, simplemente incluya @profesores en el texto de su mensaje.

Alumnos de promociones anteriores han manifestado la importancia que tiene elegir correctamente el thread y tópico de cada mensaje : poder encontrarlos en el futuro, tener lo de un mismo tema todo en un solo lugar.

Según la dinámica del curso es posible que se agreguen threads o que jamás se activen algunos de la tabla.

4.1.1 Buenas prácticas para el uso de Zulip

Zulip es una herramienta asincrónica, sus compañeros y profesores pueden escribir en cualquier momento del día o la noche, incluso durante el fin de semana.

En caso que usted no pueda acceder en forma diaria a Zulip, al ingresar se sentirá abrumado por la enorme cantidad de mensajes que han enviado sus compañeros y se planteará : “¿Qué es lo importante y que es lo que puedo obviar?

El stream cartelera es fundamental, ya que es ahí donde los profesores pasan unos pocos anuncios vitales.

Todo mensaje privado que reciba de compañeros, también será vital que los responda cuanto antes. Toda respuesta a un previo mensaje suyo también será importante que la vea cuanto antes. Si configuró bien Zulip, en estos dos casos usted recibirá un email para avisarle que tiene algo importante.

Los streams clase n y sus tópicos antes y después también son vitales y debe leer lo antes posible todo mensaje que aparezca, ya que lo orientarán.

Es una maravillosa práctica que usted cree un stream personal en Zulip con su nombre completo, que solamente usted podrá ver y escribir en él, totalmente invisible a todos, incluso invisible a los profesores. En ese stream personal usted podrá escribir sus ideas, y también guardarse links (punteros) a otros mensajes que encontró interesantes.

Para todo el resto, le recomendamos esta estrategia que le ha resultado muy útil a alumnos de años previos:

1. Empezar a leer superficialmente todos los mensajes no leídos
2. Si el mensaje vale la pena, marcarlo con la estrella “Star this message Ctrl + s”
3. Luego, cuando disponga un momento de tranquilidad, ir a la carpeta de ★ Starred Messages y leer cada mensaje en detalle, quizás hasta preguntando algo a su autor. Una vez leído a conciencia el mensaje, es muy importante guardar un link del mismo en su stream privado (el que tiene su nombre) y lo desmarca.
4. Con este método, ★ Starred Messages contiene los mensajes que temporalmente están pendientes de revisión, y su stream privado contiene punteros a los mensajes que le interesaron.

4.1.2 Participaciones extraordinariamente significativas en Zulip

Son participaciones extraordinariamente significativas :

- Responder una pregunta técnica a un compañero.
- Realizar una pregunta conceptual (no de soporte técnico).
- Recomendar bibliografía junto con un comentario por haber sido leída previamente.
- Compartir resultados de análisis exploratorio de datos.
- Compartir un problema que se tuvo y la forma en que se soluciona.
- Compartir tips sobre como operar eficientemente alguna aplicación.

4.1.3 Participaciones significativas en Zulip

Son participaciones significativas en Zulip:

- Una pregunta técnica que aún no fue realizada por un compañero

La cátedra prefiere que usted realice preguntas técnicas por Zulip en lugar del grupo privado de whatsapp para así poder entender los temas que no están quedando claros y reforzarlos ya sea por Zulip mismo o en la clase.

Es preferible curar a ese paciente aún con vida, antes que una vez finalizada la materia realizar una autopsia para entender qué salió mal y salvar vidas futuras.

4.1.4 Mensajes Privados a los profesores Permitidos

Todas las preguntas a profesores, sobre cualquier tema de la asignatura, deben ser siempre públicas.

La heterogeneidad de alumnos en las maestrías de ciencias de datos genera que por un lado nos encontramos con alumnos que ya saben programar, se encuentran trabajando en ciencia de datos y realizan preguntas avanzadas en clase, mientras que otros recién están aprendiendo lo básico. Por lo general, varios alumnos se sienten tentados a preguntar en forma privada al profesor, manifiestan tener miedo que su pregunta sea elemental, o incluso mal formulada, que pueda ser mal vista por el resto. Se sienten incómodos con dicha exposición. Le doy una maravillosa noticia, el 65% de los alumnos está en su misma situación, con lo que al hacer en forma pública la pregunta que considera elemental, dará una alentadora señal al resto y se producirá la propiedad emergente que muchos más se animen a preguntar, y tal cual lo expresa Eric Mazur será muy posible que obtenga una clara respuesta de un compañero que acaba de aprender el tema hace muy poco.

Estos son ejemplos de mensajes privados dirigidos a los profesores, que por no hablar de la asignatura, y ser sobre temas privados, si están permitidos.

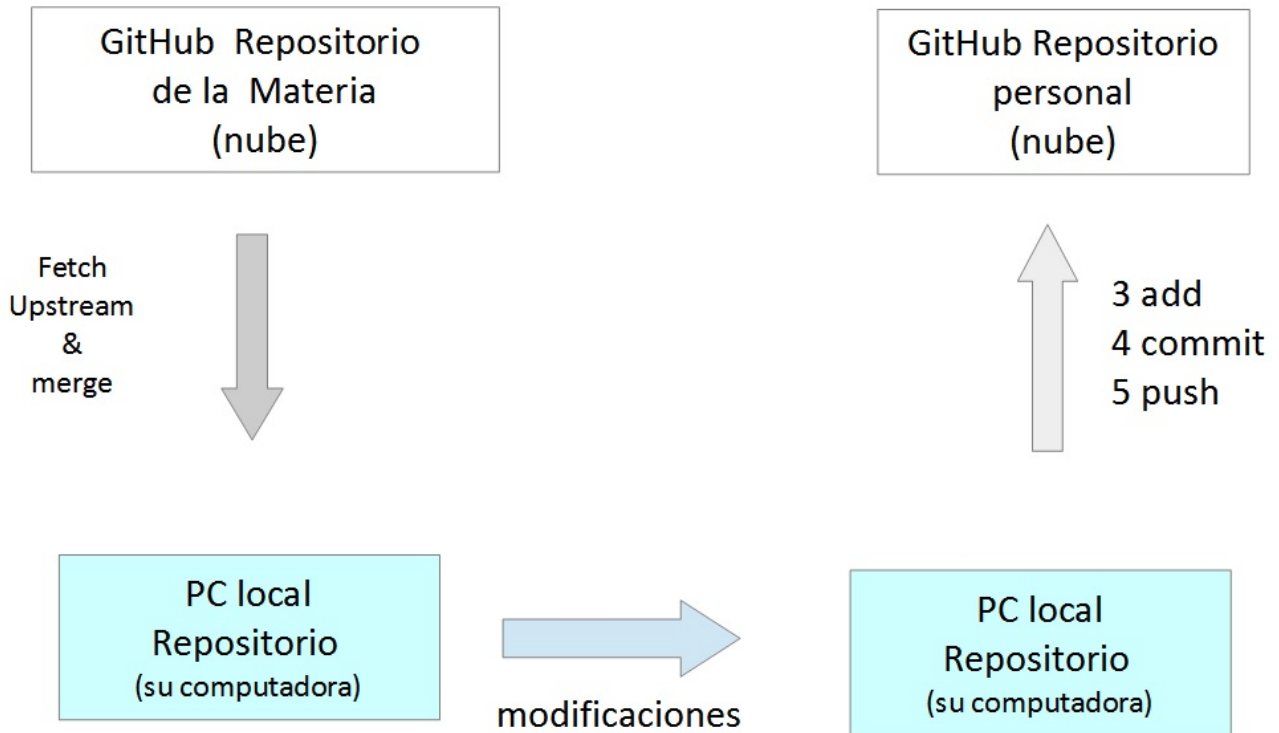
"Me surgió una posibilidad laboral en Madrid y la acepté, por lo que me voy a estar yendo en las próximas semanas y estoy haciendo gestiones a destajo. Como te podrás imaginar voy a tener que dejar el posgrado, aunque pienso empezar otro allá. Y en este punto, si sabés y tenés ganas de pasarme info, me vendría muy bien recomendaciones de buenos programas de maestría en ciencia de datos."

"Más adelante... si es posible y no es molestia, quisiera consultarte si podrías orientarme en encontrar una aplicación (PROYECTO DE TESIS) en la Industria Petroquímica / Generación de Energía....que es donde me desempeño actualmente.

Por supuesto que encontrar la aplicación es mi tarea, pero quizás algún exalumno hizo o está haciendo algo dentro del campo...o hay algún paper que se me esté escapando... . Cualquier dato es bienvenido. "

4.2 Git y GitHub

4.2.1 conceptos y operación



o desde la línea de comando de la PC local

```
git checkout main
git pull

git fetch upstream
git checkout main
git merge -Xtheirs upstream/main -m "Z manda"

git checkout main
git add .
git commit -m "otro commit"
git push
```


5 ¿Cómo seguir?

5.1 *Mantenerse actualizado*

¿En que sitios es posible mantenerse actualizado luego de finalizada la asignatura ?

- <https://machinelearningmastery.com/>
- <https://www.kdnuggets.com/>
- <https://www.analyticsvidhya.com/>
- <https://neptune.ai/blog>
- <https://www.fast.ai/#category=technical>
- <https://openai.com/blog>
- <https://www.deepmind.com/blog>
- <https://blog.ml.cmu.edu/>
- <https://news.mit.edu/topic/machine-learning>

5.2 *Datasets Públicos*

¿Dónde encontrar datasets públicas para una Tesis de Maestría o practicar ?

- <https://cloud.google.com/datasets>
- <https://github.com/awesomedata/awesome-public-datasets>
- <https://www.kaggle.com/datasets>
- <https://www.datos.gob.ar/dataset>
- <https://data.buenosaires.gob.ar/dataset/>
- <https://data.gov/>
- <https://careerfoundry.com/en/blog/data-analytics/where-to-find-free-datasets/>
- <https://medium.com/analytics-vidhya/top-100-open-source-datasets-for-data-science-cd5a8d67cc3d>
- <https://www.v7labs.com/open-datasets>
- <https://azure.microsoft.com/en-us/products/open-datasets/>
- <https://registry.opendata.aws/>

6 Empiricismo en la Ciencia de Datos

En el Machine Learning muchos algoritmos son encontrados por prueba y error, al inicio se los prueba en decenas de datasets, se divulgan, los usa la comunidad, ganan competencias Kaggle, pasan a ser usados por practicantes en innumerables situaciones, pero carecen de demostraciones matemáticas sobre su robustez- Recién luego de varios años, o más de una década se demuestra la razón por la que funcionan, los efectos que se observan y las razones de su poder (por ejemplo que converjan o que el overfitting esté controlado).

Esta situación colisiona violentamente con quienes provienen de la estadística clásica.

6.1.1 Bibliografía Introductoria

- <https://www.kdnuggets.com/2015/07/deep-learning-triumph-empiricism-over-theoretical-mathematical-guarantees.html>
- <https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2016.0153>
- <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124>
- https://ris.utwente.nl/ws/files/177935074/Boon_2020_BioTechnoPractic_How_scientists_are_brought_back_into_science_preprint_May_9_2019.pdf
- <https://hdsr.mitpress.mit.edu/pub/139rpgyc/release/1>
- <https://www.youtube.com/watch?v=oONHlua2gBY>
- <https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0145>

Arquetípico ejemplo de empiricismo, acontecido en la medicina

- <https://www.redalyc.org/journal/920/92044744013/html/> leer la sección **Anexo Semmelweis y la fiebre puerperal** (está después de la bibliografía)

7 Las dos culturas en el modelado predictivo

- El paper original de Leo Breiman Statistical Modeling, The Two Cultures
<http://www2.math.uu.se/~thulin/mm/breiman.pdf>
- <https://towardsdatascience.com/thoughts-on-the-two-cultures-of-statistical-modeling-72d75a9e06c2>

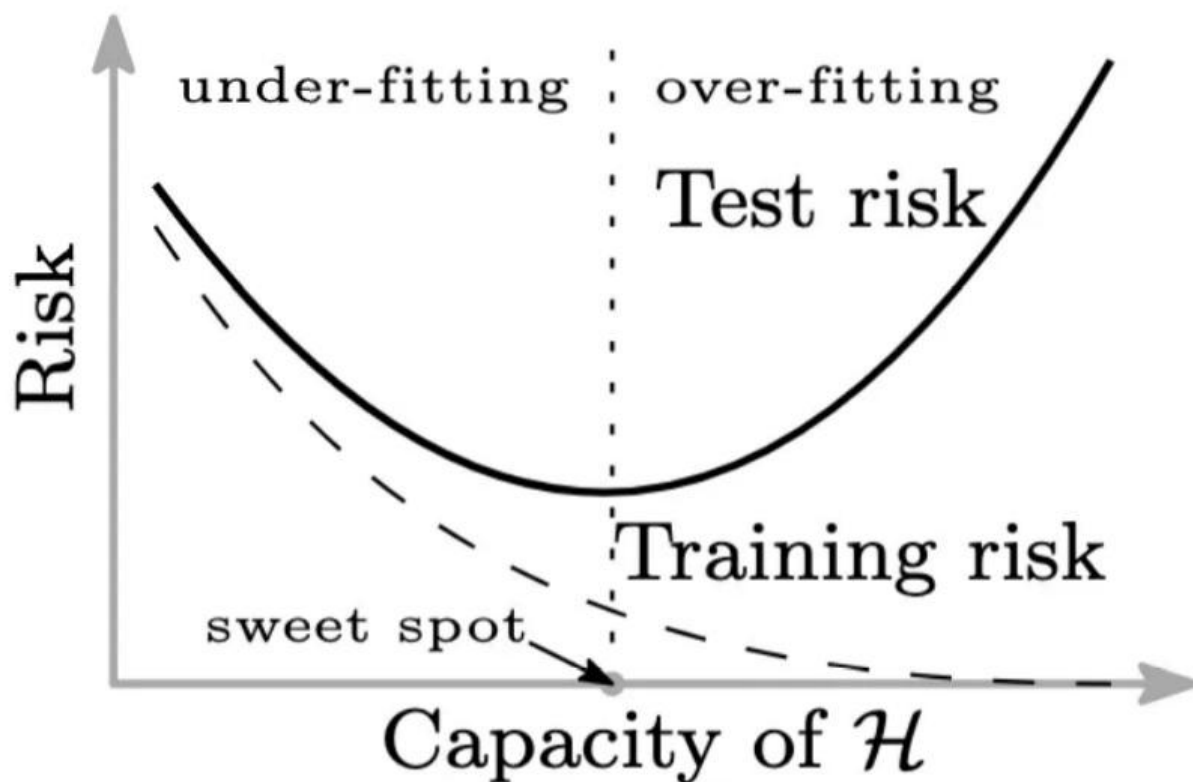
8 Estimando la Ganancia de un modelo predictivo

8.1 Bibliografía Introductoria

Bibliografía

- <https://machinelearningmastery.com/statistical-significance-tests-for-comparing-machine-learning-algorithms/>
- <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.183.534&rep=rep1&type=pdf>

8.2 Sesgo Varianza en la estadística clásica



8.2.1 Bibliografía Inicial

- <https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>
- <http://scott.fortmann-roe.com/docs/BiasVariance.html> (15 minutos)

8.3 Sesgo Varianza en el moderno Machine Learning