

## Esquema de trabajo grupal para el curso CM-274

- ✧ Equipos de 3 a 4 estudiantes.
- ✧ Entregable:
  - Archivo PDF (Informe).
  - Cuaderno de jupyter con el código y un resumen del proyecto.
  - Presentación en Beamer o Power Point para su exposición.

### Lista de temas:

- Estructuras de datos probabilísticos: Bloom filter, Count-Min Sketch, HyperLog Log .
- ✧ Inferencia variacional: Regresión lineal variacional, regresión logística variacional.
- ✧ Métodos de muestreo: Algoritmos de muestreo básico, MCMC, el algoritmo de Metropolis-Hasting.
- ✧ El algoritmo híbrido de Monte Carlo: Sistemas dinámicos.
- ✧ Cotas de Chernoff y Hoeffding: Análisis de esquemas de enrutamiento de paquetes aleatorios en las redes hipercubo y mariposa.
- ✧ Martingalas: el teorema de parada de martingalas, la desigualdad de Wald y la desigualdad de Azuma-Hoeffding.
- ✧ Modelos de grafos aleatorios: Ciclos Hamiltonianos.
- ✧ Aproximaciones Gaussianas: Binomial, Poisson y Normal.
- ✧ La paradoja de Parrondo.
- ✧ El algorítmico lema local de Lovász.
- ✧ El clasificador Bayesiano ingenuo.
- ✧ Algoritmos de corte mínimo en grafos.
- ✧ El tiempo de valor esperado en algoritmos de ordenación.
- ✧ Un algoritmo aleatorizado para calcular la mediana.
- ✧ Generación de valores aleatorios distribuidos normalmente.
- ✧ El teorema de Shanon.
- ✧ Aprendizaje PAC.
- ✧ Heavy hitter en secuencia de datos.
- ✧ Hash del cuco.

### Evaluación

La evaluación serán 5 entregables, lo cuales se detallan a continuación, que serán parte de las calificaciones de laboratorios y no son anulables.

Entregable	Fecha límite de entrega	Puntaje
1. Propuesta de proyecto	31 de agosto	Participación
2. (Introducción, estado del arte y diseño del experimento), en formato IEEE, tamaño A4, máximo 3 páginas de extensión.	21 de setiembre	Laboratorio 1
3. Código y/o scripts (jupyter notebooks) con la experimentación realizada	5 de octubre	Laboratorio 2
4. Informe final, en formato IEEE, tamaño	9 de noviembre	Laboratorio 3

A4, máximo 8 páginas de extensión		
5. Exposición de presentación de trabajo	30 de noviembre	Laboratorio 4

Se tomará más énfasis en lo siguiente para las evaluaciones:

- ☞ Contenido del Informe.
- ☞ La solución al problema a realizar y la mención de las variantes.
- ☞ Exposición y presentación de posters.

### Propuesta de proyecto

La propuesta de proyecto deberá incluir lo siguiente (máximo una hoja A4):

- ☞ Nombre del proyecto
- ☞ Miembros del equipo
- ☞ Proyecto a utilizar (dependiendo del orden de trabajos se hará las asignaciones repetidas).
- ☞ Objetivo del proyecto
- ☞ Artículos científicos relevantes
- ☞ Propuesta tentativa de modelos de clasificación a utilizar
- ☞ **Declaración de trabajo grupal (según formato de la Directiva y normas para la elaboración de trabajos grupales).**

### Informe del trabajo grupal:

El informe del proyecto deberá explicar claramente el objetivo del estudio, trabajos previos sobre el problema, código de solución (el más simple) del problema en cuestión, pruebas, conclusiones, etc. Para la mayoría de los problemas los paquetes de software de R o Python están disponibles como dominio público. No hay restricción para usarlos.

El informe debe incluir la siguiente información:

- ☞ Introducción
  - Presentación del problema general sobre el que versará el trabajo y cómo se integra dentro del uso del lenguaje R y del curso.
  - Objetivo del estudio
  - Organización del informe (secciones).
- ☞ Estado del arte
  - Breve mención del aporte que otros artículos científicos han realizado para este problema.
  - Mención de al menos 3 artículos científicos que mencionan el problema y las variantes realizadas.
- ☞ Diseño del experimento
  - Descripción de los objetos, funciones y técnicas a utilizar.
- ☞ Experimentos y resultados
  - Línea base: Reproducción de resultados reportados en un artículo científico anterior.
  - Evaluación del rendimiento de los modelos ensayados.
  - Comparación de línea base y resultados propios.
- ☞ Discusión
  - Interpretación de los resultados obtenidos.
  - ¿Cómo podría ser mejorado sus resultados?

- ☞ Conclusiones y trabajos futuros.
- ☞ Bibliografía o Referencias.

## Código y/o scripts (Jupyter Notebooks )

- ☞ El código será trabajado en equipo utilizando GitHub, de manera que se pueda verificar los aportes hechos por cada uno de los integrantes del curso.
- ☞ Se recomienda trabajar con jupyter notebooks o Jupyter lab en R o Python. Si el equipo desea trabajar en otro entorno, o en un lenguaje de programación diferente, consultarlo previamente con el profesor.
- ☞ Se deberá asignar nombres representativos a los archivos, de manera que se pueda identificar su orden relativo y el propósito de cada uno. No hay restricciones para tomar como base código tomado de otras fuentes, siempre y cuando se cite debidamente la fuente y se realice las adaptaciones que requiera el propio trabajo.
- ☞ El código deberá estar mínimamente comentado, siempre en español. Se ignorará cualquier comentario en otro idioma.
- ☞ Asimismo, se ignorará cualquier código simplemente copiado cuya fuente no haya sido citada, y se asignará el puntaje mínimo al grupo.
- ☞ Subir los archivos de código fuente y el notebook de jupyter del informe a un repositorio común del grupo, de manera que se puede revisar los aportes de todos los estudiantes.

## Exposición

El trabajo final, tendrá exposiciones. Se recomienda seguir las guías indicadas abajo:

## Recursos

- ☞ Git y Github | Curso Práctico de Git y Github Desde Cero: <https://www.youtube.com/watch?v=HiXLkL42tMU>
- ☞ Editor colaborativo LaTeX en línea: <https://www.overleaf.com/>.
- ☞ **Formato IEEE (MS Word y LaTeX):**  
<https://www.ieee.org/conferences/publishing/templates.html>
- ☞ Buscador de literatura académica: <https://scholar.google.com.pe/>
- ☞ Jupyter Notebook Tutorial: The Definitive Guide:  
<https://www.datacamp.com/community/tutorials/tutorial-jupyter-notebook>
- ☞ Notebooks With R: <https://www.datacamp.com/community/blog/jupyter-notebook-r>
- ☞ Guías para elaborar un poster:
  - ☞ <https://guides.nyu.edu/posters>
  - ☞ <https://nau.edu/undergraduate-research/poster-presentation-tips/>
  - ☞ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1876493/>

## Herramientas opcionales

## **Rstudio**

Rstudio: <https://www.rstudio.com/> , es un IDE para R. Es software libre con licencia GPLv3 y se puede ejecutar sobre distintas plataformas.

## **Git-Github**

Git: <https://git-scm.com/> es un sistema de control de versiones de gran potencia y versatilidad en el manejo de un gran número de archivos de código fuente a través del desarrollo no lineal, es decir vía la gestión rápida de ramas y mezclado de diferentes versiones.

Github: <https://github.com/>, es una plataforma de desarrollo colaborativo de software utilizado para alojar proyectos (muchos proyectos importantes como paquetes de R, Django, el Kernel de Linux, se encuentran alojados ahí) utilizando Git y el framework Ruby on Rails.

## **Anaconda y Jupyter**

Anaconda: <https://www.continuum.io/downloads> es una distribución completa de python que incluye el paquete Jupyter.

El proyecto Jupyter: <http://jupyter.org/>, es una aplicación web que permite crear y compartir documentos que contienen código de diversos lenguajes de programación, ecuaciones, visualizaciones y texto en diversos formatos.

## **Libros de referencia**

- ☞ Tilman M. Davies, The Book of R, No Starch Press, INC 2016
- ☞ Using R for Introductory Statistics, John Verzani, June 26, 2014 by Chapman and Hall/CRC.