



## Proyecto Primer Bimestre

### Objetivo:

Desarrollar un modelo de regresión para determinar cuáles variables son capaces de explicar un fenómeno determinado, por lo tanto, se desea identificar qué variables aportan de manera sustancial a la creación de un modelo estadístico basado en datos.

### Instrucciones del Informe:

1. El proyecto se realizará en grupos de hasta 5 personas. Se entregará un solo proyecto por grupo con todos los ficheros que se generan en latex incluyendo, por supuesto, el .pdf final. Además, se deberá adjuntar el fichero de R, con el modelo que se ha construido.
2. La extensión del informe final no debe ser menor a 5 páginas ni mayor a 10.
3. El informe puede ser redactado en español o inglés y tiene que poseer, al menos, la siguiente estructura:
  - Resumen (200 a 250 palabras)
  - Palabras Clave (3 a 10 palabras)
  - Introducción (explicación del contexto y revisión del estado de arte)
  - Metodología (qué modelo van a usar, por qué y cuáles variables van a utilizar)
  - Experimentos (se pueden incluir gráficos del dataset y cualquier otro tipo de ayuda que permita explicar los experimentos realizados)
  - Conclusiones
  - Referencias (incluir sólo referencias de artículos científicos o libros. Recomendación de fuentes: ScienceDirect, IEEE Xplore, MDPI, Springer, Taylor & Francis; Cengage Learning, Wiley)
4. Se habilitará un enlace en AVAC para la entrega del Proyecto 1.

### Instrucciones del Dataset:

Ya que el proyecto consiste en la generación de un modelo de Regresión Lineal Múltiple, es menester que se seleccione un dataset de manera libre, pero con al menos las siguientes características:

- Tamaño de aproximadamente 1000 instancias, con mínimo 4 y máximo 8 variables.
- La lista de datasets que **NO** podrán ser utilizados son:
  - <https://www.kaggle.com/nasa/asteroid-impacts>
  - <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>
  - <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
  - <https://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set>
  - <https://www.kaggle.com/pavanraj159/predicting-a-pulsar-star>
  - <https://archive.ics.uci.edu/ml/datasets/air+quality>
  - <https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>
  - <https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice>
  - <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>
  - <https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>
  - <https://archive.ics.uci.edu/ml/datasets/QSAR+fish+toxicity>

<https://archive.ics.uci.edu/ml/datasets/Wireless+Indoor+Localization>  
<https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choic>  
<http://archive.ics.uci.edu/ml/datasets/Computer+Hardware>  
<https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise>  
<https://archive.ics.uci.edu/ml/datasets/Energy+efficiency>  
<https://archive.ics.uci.edu/ml/datasets/glass+identification>  
<https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame>  
<https://archive.ics.uci.edu/ml/datasets/Ecoli>  
<https://www.kaggle.com/mirichoi0218/insurance>  
<https://www.sciencebase.gov/catalog/item/580e72cee4bof497e794b81e>  
<https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring>  
<https://archive.ics.uci.edu/ml/datasets/Bias+correction+of+numerical+prediction+model+temperature+forecast>  
<https://archive.ics.uci.edu/ml/datasets/Forest+Fires>  
<https://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant>  
<https://www.kaggle.com/uciml/pima-indians-diabetes-database>  
<https://archive.ics.uci.edu/ml/datasets/MAGIC+Gamma+Telescope>  
<https://www.kaggle.com/rajyellow46/wine-quality>  
<https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients#>  
<https://www.kaggle.com/selfishgene/youtube-faces-with-facial-keypoints>  
<https://www.kaggle.com/tarunpaparaju/apple-aapl-historical-stock-data>  
<https://www.kaggle.com/rpaguirre/tesla-stock-price>  
[https://www.kaggle.com/adityasakare/bitcoin-dataset?select=Bitcoin\\_data.csv](https://www.kaggle.com/adityasakare/bitcoin-dataset?select=Bitcoin_data.csv)  
<https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>  
<https://www.kaggle.com/yassershrief/breath-alcohol-ames-dataset>  
<https://archive.ics.uci.edu/ml/datasets/South+German+Credit>  
<https://archive.ics.uci.edu/ml/datasets/Facebook+metrics>  
<https://www.kaggle.com/gavincanacam/home-loan-predictions>  
<https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings#>  
<https://www.kaggle.com/deepshah16/meme-cryptocurrency-historical-data?select=Bitcoin.csv>  
<https://www.kaggle.com/adityadesai13/used-car-dataset-ford-and-mercedes?select=hyundi.csv>  
<https://www.kaggle.com/datasnaek/youtube-new?select=CAvideos.csv>  
<https://www.kaggle.com/lislejoem/us-minimum-wage-by-state-from-1968-to-2017>  
<https://www.kaggle.com/keplersmachines/kepler-labelled-time-series-data>  
<https://www.kaggle.com/jaepin/shopeeph-koreantop-clothing>  
<https://archive.ics.uci.edu/ml/datasets/Productivity+Prediction+of+Garment+Employees>  
<https://www.kaggle.com/varpit94/american-airlines-group-stock-data>  
<https://www.kaggle.com/uciml/pima-indians-diabetes-database>  
<https://www.kaggle.com/uciml/student-alcohol-consumption>  
<https://archive.ics.uci.edu/ml/datasets/Abalone>  
<https://www.kaggle.com/datasets/surajjha101/stores-area-and-sales-data>  
<https://www.kaggle.com/datasets/alanjo/graphics-card-full-specs>  
<https://www.kaggle.com/datasets/surajjha101/stores-area-and-sales-data>  
<https://www.kaggle.com/datasets/yasserh/song-popularity-dataset>

- El dataset puede ser extraído de cualquier repositorio público, ejemplo: UCI Machine Learning Repository, Kaggle, KDnuggets, Mendeley Data, Gapminder, Data.gov, MLdata.org, etc.
- Una vez que se ha seleccionado el dataset se debe registrarlo, para su aprobación (hasta el 27/11/2022), en:

[https://docs.google.com/document/d/1hHQAdp2xjklVWoTo8H1KV4OJs8eWUA\\_7q3fkVxSMsjY/edit?usp=sharing](https://docs.google.com/document/d/1hHQAdp2xjklVWoTo8H1KV4OJs8eWUA_7q3fkVxSMsjY/edit?usp=sharing)

- Sobre los datos se espera:
  - Limpieza de datos (data curation).
  - Representación gráfica de los datos.
  - Representación de correlaciones entre los datos.
  - Ejecución de la tarea de regresión.
  - Pruebas de significancia y bondad de ajuste del modelo.
  - Comprobación de las suposiciones del modelo.
  - Identificación de outliers y valores influyentes.

### Defensa Oral del Proyecto:

Se escogerá una persona, al azar, del grupo de estudiantes para que realice la defensa oral del proyecto. El estudiante podrá apoyarse de cualquier material que previamente haya preparado (código, diapositivas, prezi, etc.)

### Rúbrica:

La rúbrica pretende ser una guía general de evaluación de los proyectos de manera que el alumno sepa qué se le va a exigir en el trabajo y cómo se le va a evaluar. La rúbrica con la que evaluaremos este proyecto es la siguiente:

- **Informe y Contenidos:**

Categoría	Excelente	Bien	Mejorable	Pésimo
Redacción (10 puntos)	No hay errores de gramática, ortografía o puntuación. 10 puntos	Casi no hay errores de gramática, ortografía o puntuación. 8 puntos	Unos pocos errores de gramática, ortografía o puntuación. 4 puntos	Muchos errores de gramática, ortografía o puntuación. 2 puntos
Organización (10 puntos)	La información está muy bien organizada con párrafos bien redactados y con subtítulos. 10 puntos	La información está organizada con párrafos bien redactados. 8 puntos	La información está organizada, pero los párrafos no están bien redactados. 4 puntos	La información proporcionada no parece estar organizada. 2 puntos
Calidad de la Información (60 puntos)	La información está claramente relacionada con el tema principal y proporciona varias ideas secundarias y/o ejemplos. 60 puntos	La información da respuesta a las preguntas principales y 1-2 ideas secundarias y/o ejemplos. 50 puntos	La información da respuesta a las preguntas principales, pero no da detalles y/o ejemplos. 30 puntos	La información tiene poco o nada que ver con las preguntas planteadas. 10 puntos
Diagramas e Ilustraciones (10 puntos)	Los diagramas e ilustraciones son ordenados, precisos y añaden al entendimiento del tema. 10 puntos	Los diagramas e ilustraciones son precisos y añaden al entendimiento del tema. 8 puntos	Los diagramas e ilustraciones son ordenados y precisos y algunas veces añaden al entendimiento del tema. 4 puntos	Los diagramas e ilustraciones no son precisos o no añaden al entendimiento del tema. 2 puntos
Fuentes (10 puntos)	Todas las fuentes de información y las gráficas están documentadas y en el formato deseado. 10 puntos	Todas las fuentes de información y las gráficas están documentadas, pero unas pocas no están en el formato apropiado. 8 puntos	Todas las fuentes de información y gráficas están documentadas, pero muchas no están en el formato apropiado. 4 puntos	Algunas fuentes de información y gráficas no están documentadas. 2 puntos

- **Presentación Oral:**

DIMENSIÓN	EXCELENTE (8-10 PUNTOS)	BUENO (5-7 PUNTOS)	MALO (2-4 PUNTOS)	DEFICIENTE (0-1 PUNTO)
Social (3,0/10)	El estudiante socializa con mucha facilidad al interior de su grupo las definiciones para el planteamiento de un proyecto y se socializan los procedimientos	El estudiante interactúa de manera adecuada pero tiende a distraerse. Resuelven varios estudiantes y luego comparan respuestas, les falta un poco de tiempo	El estudiante casi no interviene en la discusión, quienes saben calcular límites los hacen y los demás observan pasivamente. Se forman subgrupos aislados de trabajo	Los estudiantes no interactúan, cada uno actúa independientemente, quienes saben no comparten el conocimiento
Cognitiva (4,0/10)	El estudiante aporta con ideas muy positivas en el contexto académico exigido. Todos llegan al conocimiento suficiente. Saben plantear un proyecto	El estudiante aporta unas cuantas ideas positivas pero a veces sale del contexto planteado. Se discuten excesivamente (en tiempo) la utilidad del planteamiento de un proyecto.	El estudiante casi no aporta conocimientos, sin embargo algunos obtienen respuestas no todas correctas	Los conocimientos no son comunes en el grupo, hay confusión. No está clara como plantear un proyecto.
Análisis (8,0/5)	El estudiante plantea correctamente un proyecto	Los procedimientos Son correctos, pero contienen errores de concepto	Los procedimientos Son correctos, pero contienen errores de concepto	Existen graves errores conceptuales sobre planteamiento de un proyecto.