

Trabajo 1 (40%) - Introducción a la analítica

IMPORTANTE: Tenga en cuenta lo siguiente:

- Los siguientes ejercicios deben realizarse utilizando **Rmarkdown** del **R-Studio** y **subir en MOODLE el pdf** con la solución.
- El trabajo debe tener máximo 40 páginas, por lo cual es importante que seleccionen los resultados y gráficos más importantes, además de que controlen el tamaño de tablas y figuras.
- **LA FECHA MÁXIMA DE ENTREGA ES EL VIERNES 3 DE DICIEMBRE DE 2021 A LAS 12:00 DE LA MEDIA NOCHE. NO MODIFICARÉ LA FECHA DE ENTREGA, POR LO CUAL TENGAN EN CUENTA QUE EL CRONÓMETRO QUE APARECE EN MOODLE ES INMODIFICABLE Y LUEGO DE LA FECHA ESTIPULADA NO LOS DEJARÁ SUBIR EL TRABAJO Y LA NOTA SERÁ CERO.**

Planteamiento del trabajo:

1. Descargue de Yahoo Finance la base de datos de los precios de cierre diarios de la acción que se le asignó a su grupo en el periodo que va del 1 de enero de 2016 hasta el 31 de diciembre de 2020.
- **Grupo 1:** United Airlines Holdings, Inc. (UAL).
 - **Grupo 2:** Pfizer Inc. (PFE).
 - **Grupo 3:** JPMorgan Chase & Co. (JPM).
 - **Grupo 4:** Banco Santander, S.A. (SAN).
 - **Grupo 5:** Oracle Corporation (ORCL).
 - **Grupo 6:** Royal Caribbean Group (RCL).
 - **Grupo 7:** Chevron Corporation (CVX).
 - **Grupo 8:** CEMEX, S.A.B. de C.V. (CX).

- a. Contextualice a qué se dedica y dónde opera principalmente la empresa que se le asignó a su grupo. Luego, construya una base de datos con la misma estructura de los datos **Smarket** que se encuentran en el paquete **ISLR** (ver diapositiva 29 de la Clase 1). Realice un análisis descriptivo con los resultados y los gráficos que usted considere pertinentes, explicando lo que observa en cada uno.
 - b. Utilizando validación cruzada, encuentre el K (el número de vecinos), del modelo KNN, que mejores resultados arroje en términos del error de prueba estimado para predecir si el precio de la acción sube (o se mantiene igual) o baja en función de los 5 “lags” pasados y el volumen.
 - c. Con los datos de entrenamiento, ajuste un modelo logístico, un KNN con K encontrado en el item (b), un LDA y un QDA para predecir si el precio de la acción sube (o se mantiene igual) o baja en función de los 5 “lags” pasados y el volumen. Para cada modelo obtenga la matriz de confusión y el estimador del error de prueba. ¿Cuál modelo arroja mejores resultados y por qué?
 - d. Saque conclusiones de los resultados obtenidos en el item (c).
2. Realice las simulaciones descritas en las diapositivas Clase 1 y replique los gráficos de las diapositivas 22 y 24 sin tener en cuenta el método KNN-CV (pero sí KNN-1).
 3. Pruebe que si en el modelo de regresión lineal múltiple se tiene que $p > n$ (el número de covariables es mayor que el tamaño muestral) entonces los estimadores de los coeficientes, $\hat{\beta}$, no son únicos. ¿Cuáles son las alternativas para “resolver” este problema?
 4. Realice la **Actividad** propuesta en la última diapositiva de la **Clase 3**.
 5. Realice la **Actividad** propuesta en la última diapositiva de la **Clase 4**.
 6. Realice la **Actividad** propuesta en la última diapositiva de la **Clase 5**.