



**UAI**  
**UNIVERSIDAD ADOLFO IBÁÑEZ**

**UNIVERSIDAD ADOLFO IBÁÑEZ – FACULTAD DE INGENIERÍA Y CIENCIAS**  
**Magíster en Ingeniería Industrial e Investigación de Operaciones**  
**Año Académico 2020**

## **PRUEBA II**

**Curso** : **Análisis Predictivo**  
**Profesor** : **Abel Valdebenito Sanhueza**  
**E-mail** : **abel.valdebenito@edu.uai.cl**  
**Fecha entrega** : **04-12-2020**

**Pronosticar el tiempo de vida de unidades de almacenamiento.**

### **Contexto**

Cada día, Backblaze toma una foto de cada disco duro operativo que incluye información básica del disco duro (por ejemplo, capacidad, falla) y S.M.A.R.T. estadísticas reportadas por cada unidad. Este conjunto de datos contiene datos de los dos primeros trimestres de 2016.

### **Contenido**

Este conjunto de datos contiene información básica del disco duro y 90 columnas o valores sin procesar y normalizados de 45 S.M.A.R.T. estadísticas diferentes. Cada fila representa una observación diaria de un disco duro.

date: Fecha en formato yyyy-mm-dd

serial\_number: Número de serie del disco asignado por el fabricante.

model: Número de modelo del disco asignado por el fabricante.

capacity\_bytes: Capacidad del disco en bytes.

failure: Contiene un "0" si el disco está OK. Contiene un "1" si el último día la unidad estuvo operativa antes de fallar.

90 variables que comienzan con smart': valores sin procesar y normalizados para 45 estadísticas SMART diferentes según lo informado para una unidad dada

### **Hints**

Algunos elementos a tener en cuenta al procesar los datos:

- Las estadísticas SMART pueden variar en significado según el fabricante y el modelo. Puede ser más informativo comparar unidades que sean similares en modelo y fabricante.
- Algunas columnas SMART pueden tener valores fuera de límites
- Cuando falla una unidad, la columna 'falla' se establece en 1 el día de la falla y, a partir del día siguiente, se eliminará la unidad del conjunto de datos. Cada día, también se agregan nuevas unidades. Esto significa que la cantidad total de unidades por día puede variar.
- SMART 9 es la cantidad de horas que una unidad ha estado en servicio. Para calcular la edad de una unidad en días, divida este número entre 24.

### **Parte I (20%). Limpieza de datos e ingeniería de variables**

1. Transforme los datos a tiempos de vida de unidades.
2. Calcule features de las variables SMART correspondientes a promedio, desviaciones estándar y coeficiente de asimetría, de la última semana, últimas dos semanas, último mes

### **Parte II (40%) Análisis descriptivo de los tiempos de sobrevivencia.**

3. Realice un análisis descriptivo bivariado entre tiempo de vida de las unidades a través de las curvas de Kaplan-Meier con cada una de las features.
4. Realice Test de comparación de curvas
5. Inspeccione gráficamente el supuesto de riesgos proporcionales.

### **Parte III (40%) Modelos de Sobrevivencia.**

6. Ajuste un modelo exponencial y un modelo weibull, ¿cuál de estos modelos explica mejor los tiempos de vida?, fundamente su respuesta. Interprete los resultados del modelo.
7. Ajuste un modelo lognormal y gamma ¿en comparación al modelo que seleccionó en 6, es uno de estos modelos mejor?.
8. Del modelo paramétrico seleccionado, compare las curvas de sobrevivencia en relación a la curva de Kaplan-Meier
9. Ajuste un modelo de regresión de Cox. Compare los resultados en comparación a los modelos desarrollados previamente. Interprete los resultados en términos de los coeficientes de regresión.