**Prueba técnica para el cargo de Analista de Cartera del Grupo MP**

Política de Confidencialidad y Tratamiento de Datos Personales

Al participar en esta entrevista, usted se compromete a mantener la confidencialidad de toda la información de la empresa y del contenido de la prueba técnica, utilizando dicha información únicamente para este proceso y sin divulgarla a terceros. De igual forma, usted autoriza a la empresa a consultar y utilizar sus datos personales con fines relacionados con la selección. La empresa garantizará el tratamiento adecuado de sus datos conforme a la Ley 1581 de 2012 y el Decreto 1377 de 2013 de Colombia.

Su participación implica la aceptación de estas condiciones.

Esta prueba evalúa tus conocimientos y habilidades en estadística, indicadores, Power BI, R o RStudio, Python, Machine Learning y SQL. Cada pregunta y situación están diseñadas para medir tu competencia en áreas críticas para el rol. Por favor, responde de forma breve y precisa. ¡Te deseamos mucho éxito!

**Preguntas:**

1. En el contexto estadístico, ¿qué es un dato?
2. Describa la diferencia entre variable estadística e indicador.
3. ¿Cuáles son algunos tipos de análisis estadísticos fundamentales en machine learning y cómo pueden influir en la eficacia de los modelos?
4. ¿Cuál es la diferencia entre adivinar y predecir en el contexto del análisis de datos?
5. ¿Qué realiza el siguiente código en Python y por qué es importante aplicarlo a un conjunto de datos?

Texto

Descripción generada automáticamente

1. En el contexto de datos estructurados temporalmente (por ejemplo, registros diarios, semanales, mensuales), ¿cómo debería particionarse el conjunto de datos para entrenamiento y pru­eba al aplicar técnicas de machine learning? Explica brevemente tu enfoque y la razón detrás de esa estrategia.
2. Describe la metodología general de modelado que ofrece el paquete gamlss en R y cómo puede ser útil en el análisis de datos estadísticos.
3. ¿Qué es un procedimiento almacenado en SQL?
4. ¿Qué es Power Query en Power BI y cómo se utiliza para la preparación de datos?
5. Imagina que la empresa necesita proyectar el recaudo futuro basado en datos históricos de ventas. Tienes disponibles varios modelos de machine learning para predecir esta cantidad: regresión lineal, regresión polinomial, y un modelo de redes neuronales. Dado que la empresa toma decisiones basadas en su flujo de caja, y que dichas decisiones dependen directamente de la proyección que realices, ¿cómo decidirías qué modelo utilizar y por qué?
6. Ejercicio Práctico: Proyección del Recaudo.

**Objetivo:** El objetivo de este ejercicio es proyectar el recaudo basado en datos históricos.

**Indicaciones:** Inicia realizando un análisis descriptivo y exploratorio de la variable recaudo. Normaliza los datos de entrenamiento y de prueba, dejando el recaudo en su escala original. Ajusta al menos dos algoritmos predictivos de Machine Learning al conjunto de datos de entrenamiento *(datos\_train)* y predice con el conjunto de datos de prueba *(datos\_test).* Elige el mejor modelo ajustado utilizando el RMSE. Calcula la asertividad de los modelos. *(Puedes hacerlo en R o en Python).*

**Descripción de las variables del conjunto de datos, datos\_train:** en este conjunto de datos vas a encontrar un conjunto de 11 variables (columnas); la variable de respuesta: Recaudo y otras 10 variables de entrada, explicativas o regresoras. A continuación, su descripción:

* **Recaudo:** Dinero que la empresa ha recolectado.
* **Semana:** Número de la semana en la que se ha recolectado o se recolectará la venta realizada.
* **Venta:** Total de ventas realizadas por los grupos que facturan en esa semana.
* **Cons:** Proporción de clientes que harán pedidos consecutivos.
* **Grupos:** Número de grupos de clientes de diferentes zonas del país que facturan en una semana.
* **Festivos:** Número de días festivos en la semana.
* **Ing:** Número de ingresos realizados o planificados.
* **Egr:** Número de egresos realizados o planificados.
* **Ped:** Número de pedidos realizados o planificados.
* **NS:** Nivel de servicio de la operación.
* **OP:** Promedio de órdenes de los grupos que facturan en una semana.

¡Éxitos!