

Escuela de Ciencia de la Computación
Sustitutorio
CC531 A Análisis en Macrodatos 2023-I

Desarrollar un Proyecto Final en **Spark**.

- Programación en Apache Spark Scala de los algoritmos de Multilayer perceptron classifier y Linear Support Vector Machine
- Entrenar los modelos de Multilayer perceptron classifier y Linear Support Vector Machine, con sus respectivas base de datos.
- Enviar desde un **nodo** a apache Spark mediante el Spark Streaming datos para test, para que procesen los modelos ya entrenados de Multilayer perceptron classifier y Linear Support Vector Machine.
- Se procesará en el Apache Spark, los modelos ya entrenados de Multilayer perceptron classifier y Linear Support Vector Machine, y devolverá los resultados a los clientes que los solicitaron, donde esto imprimieran el resultado en su **nodo**.

Realizar la configuración de Spark en donde realizar los siguiente:

- Buscar una Base de Datos de un caso específico para Multilayer perceptron classifier y Linear Support Vector Machine
- Describir el caso de Estudio para Analizar.
- Uso de librerías Spark Scala

Comprimir todos los archivos:

- Utilizar la Plantilla en un archivo PDF:
 - Para describir detalles de la base de datos, el rubro de la organización de la base de datos, describir la teoría que implica los datos, que características tiene la base de datos, campos detalles, reglas de los datos, que información usted cree, que se podría obtener base de datos.
- Adjuntar los códigos fuentes con sus formatos respectivos (no subir otros archivos).

No usar bases de datos relacionadas a:

1. Abandono en el sector bancario.
2. Riesgo Financiero.
3. Enfermedades cardiovasculares (o enfermedades cardiacas).
4. Base de datos de información de LinkedIn.
5. Incidentes delictivos.

6. Personas afectadas por Covid.
7. Estilos de vestir.
8. Casos de incendio.
9. Pacientes.
10. Hospitales.
11. Vinos.
12. Cáncer del pulmón.
13. Venta de Casas.

Parte 1