The student would need to consider the following instructions (a - d) during the development of this project.

a) Justificación lógica basada en el razonamiento de la elección específica de enfoques de aprendizaje automático.

b) Múltiples modelos de aprendizaje automático (al menos dos) utilizando hiperparámetros y una comparación entre

los enfoques de modelado elegidos.

c) Visualice su comparación de los resultados del modelado de ML. Puede utilizar un enfoque estadístico para argumentar que

una característica es más importante que otras.

d) Se deben utilizar métodos de validación cruzada para justificar la autenticidad de sus resultados de ML.

Presentará los hallazgos y defenderá los resultados en el informe (MS Doc) destacando su trabajo. Su El informe debe capturar los siguientes aspectos que sean relevantes para las investigaciones de su proyecto.

1. Una introducción clara, motivación, una descripción del dominio del problema y una explicación de cómo Los objetivos del proyecto se justifican mediante algoritmos de Predicción / Clasificación.

(20 puntos)

2. Caracterización de los datos, preprocesamiento, explicación y descripción de las técnicas utilizadas para la variación en la precisión en tres divisiones de entrenamiento (20%, 25% y 30%) usando validación cruzada técnicas.

(30 puntos)

3. ¿Cuál es el propósito principal del ajuste de hiperparámetros en el aprendizaje automático? ¿Podrías dar más detalles?

Técnicas específicas de ajuste de hiperparámetros (por ejemplo, GridSearchCV) aplicadas a modelos de aprendizaje automático.

para encontrar parámetros óptimos?

(25 puntos)

4. Interpretar y explicar los resultados obtenidos, discutir sobreajuste/infraajuste/generalización, proporcionar una Justifique los modelos elegidos y utilice visualizaciones para respaldar sus hallazgos. Comentarios en Python código, las conclusiones del proyecto deben especificarse al final del informe. El estilo Harvard debe ser utilizado para citas y referencias.

(25 puntos)

Selecting only numeric columns

numeric data = data.select\_dtypes(include=['float64', 'int64'])

correlation\_matrix = data.corr()

print(correlation\_matrix)

Todas las presentaciones de evaluación deben cumplir con los requisitos mínimos que se enumeran a continuación. No hacerlo puede tener

implicaciones para las calificaciones otorgadas.

● Todos los archivos deben cargarse por separado en Moodle.

● Detallar claramente el número de palabras utilizadas en el informe.

● Número de palabras en el informe (1250 palabras +/-10%) sin incluir diagramas, códigos, referencias ni citas.

y títulos.

● Utilice control de versiones como Github o cualquier otra herramienta para mostrar el progreso en CA1. deberías tener en al menos 5 confirmaciones en Github antes del envío.

● La rúbrica se proporciona para el desglose detallado de las calificaciones al final de esta CA.

● Utilice las referencias de Harvard cuando cite material de terceros.

● Ser trabajo propio del alumno.

● Incluir la portada de la evaluación CCT.

● Presentarse antes de la fecha límite especificada o estar sujeto a sanciones por presentación tardía.

● Debe especificarse claramente el número de palabras utilizadas después de cada sección del informe.