



ITESM- Campus Puebla

## **NAATIK | Entrega Solución del Reto**

### **Inteligencia artificial avanzada para la ciencia de datos II**

#### **Integrantes Equipo 1:**

Myroslava Sánchez Andrade A01730712  
José Antonio Bobadilla García A01734433  
Karen Rugerio Armenta A01733228  
Alejandro Castro Reus A01731065

Fecha: 30/11/2022

**Repositorio principal:** [https://github.com/myrosandrade89/TC3007C\\_AI/tree/main](https://github.com/myrosandrade89/TC3007C_AI/tree/main)

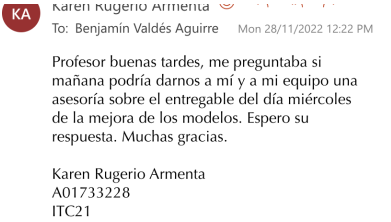

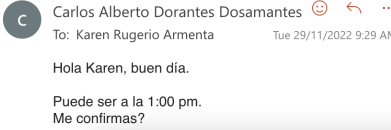
**Página WEB:**

- **Link al repositorio Backend:** <https://github.com/AlexReusC/server-naatik>
- **Link al repositorio Frontend:** <https://github.com/AntonioBobadilla/website-naatik>


Módulo Responsable de evaluar el rubro
Reto
Big Data
Cómputo en la nube
Técnicas y arquitecturas de deep learning.
Estadística avanzada para la ciencia de datos.

Rubros	Indicadores	Cambios Realizados	Link
Metodología	Hace un levantamiento de requerimientos completo y correcto	<p><b>Comentarios:</b> Sólamante se pidió un cambio en Memorandum of Understanding antes de ser firmado por los socios.</p> <div> <p>Hola equipo,</p> <p>Me pone nervioso firmar un acuerdo que tiene costo desglosado y en ningún lugar dice que se estima el costo como ejercicio, no como un compromiso de Naatik. Pueden corregirlo por favor?</p> <p>Saludos Pablol</p> </div> <p><b>Correcciones:</b> Se realizaron las correcciones correspondientes y se firmó el Memorandum.</p>	<p><b>Requerimientos:</b> <a href="https://github.com/myrosandrade89/TC3007C_AI/blob/main/documentation/RequerimientosOnePage.pdf">https://github.com/myrosandrade89/TC3007C_AI/blob/main/documentation/RequerimientosOnePage.pdf</a></p> <p><b>Memorandum of Understanding:</b> <a href="https://github.com/myrosandrade89/TC3007C_AI/blob/main/documentation/Memorandum%20of%20Understanding.pdf">https://github.com/myrosandrade89/TC3007C_AI/blob/main/documentation/Memorandum%20of%20Understanding.pdf</a></p>
	Aplica metodología de manera correcta	<p><b>Comentarios:</b> No se tuvieron correcciones de esta actividad.</p>	<p><b>Documentación:</b> <a href="https://github.com/myrosandrade89/TC3007C_AI/tree/main/documentation">https://github.com/myrosandrade89/TC3007C_AI/tree/main/documentation</a></p>

	Mantiene una comunicación clara y efectiva con el socio formador	<b>Comentarios:</b> No aplica	<b>Se tuvieron sesiones semanales vía zoom cada viernes con los socios formadores a través del siguiente link:</b> <a href="https://itesm.zoom.us/j/6115223557?pwd=cU12QVZZWE1VeEZHcjZvM0tUZGtIZz09">https://itesm.zoom.us/j/6115223557?pwd=cU12QVZZWE1VeEZHcjZvM0tUZGtIZz09</a>
Documentación	Genera documentación clara y completa del proceso	<b>Comentarios:</b> No se tuvieron correcciones de esta actividad.	<b>Documentación CRISP:</b> <a href="https://github.com/myr osandrade89/TC3007 C_AI/blob/main/docum entation/CRISP.pdf">https://github.com/myr osandrade89/TC3007 C_AI/blob/main/docum entation/CRISP.pdf</a> <b>Repositorio:</b> <a href="https://github.com/myr osandrade89/TC3007 C_AI">https://github.com/myr osandrade89/TC3007 C_AI</a>
	El repositorio del proyecto está bien estructurado y cumple con estándares.	<b>Comentarios:</b> No se tuvieron correcciones de esta actividad.	<b>Repositorio:</b> <a href="https://github.com/myr osandrade89/TC3007 C_AI">https://github.com/myr osandrade89/TC3007 C_AI</a>
Datos	Identifica si es necesario el uso de big data para la solución del reto y explica correctamente su razonamiento.	<b>Comentarios:</b> <a href="https://docs.google.com/spreadsheets/d/1xkPiQcRArJNv2ucWtkQ3Q_XrQWeG8tlo/edit#gid=1993151062&amp;range=A1:O23">https://docs.google.com/spreadsheets/d/1xkPiQcRArJNv2ucWtkQ3Q_XrQWeG8tlo/edit#gid=1993151062&amp;range=A1:O23</a>	<b>Documentación:</b> <a href="https://github.com/myr osandrade89/TC3007 C_AI/blob/main/docum entation/RetoDatos.pdf">https://github.com/myr osandrade89/TC3007 C_AI/blob/main/docum entation/RetoDatos.pdf</a>
	Diseña un modelo de almacenamiento y manejo de datos ad hoc al reto	<b>Correcciones:</b> 1. Se realizó un análisis de herramientas y tecnologías de acuerdo a los comentarios. 2. Se definió cómo sería el modelo de almacenamiento correcto, en caso de que el socio quisiera implementarlo en un futuro. 3. Se realizó k-folds validation para cada modelo. 4. Se realizó un análisis más profundo de los datos y si es Big-Data.	<b>Extracción y carga de datos:</b> <a href="https://github.com/myr osandrade89/TC3007 C_AI/blob/main/implementation/fragments/etl/etl.ipynb">https://github.com/myr osandrade89/TC3007 C_AI/blob/main/implementation/fragments/etl/etl.ipynb</a> <b>K-Folds validation:</b> <a href="https://github.com/myr osandrade89/TC3007 C_AI/tree/main/implementation/fragments/models">https://github.com/myr osandrade89/TC3007 C_AI/tree/main/implementation/fragments/models</a> (cada modelo tiene la implementación de K-Folds validation, a excepción del CNN )

Modelo	Selecciona un modelo adecuado al problema presentado en el reto.	<p><b>Comentarios:</b>  <a href="https://docs.google.com/spreadsheets/d/1xkPiQcRARJNV2ucWtkQ3Q_XrQWeG8tlo/edit#gid=1060677826&amp;range=A1:G21">https://docs.google.com/spreadsheets/d/1xkPiQcRARJNV2ucWtkQ3Q_XrQWeG8tlo/edit#gid=1060677826&amp;range=A1:G21</a></p> <p>Los comentarios son de <b>falta de entrega del avance</b>. Es importante mencionar que los comentarios reales de las correcciones se hicieron por medio de un <b>zoom</b>:</p> <p> Karen Rugerio Armenta To: Benjamín Valdés Aguirre Mon 28/11/2022 12:22 PM</p> <p>Profesor buenas tardes, me preguntaba si mañana podría darnos a mí y a mi equipo una asesoría sobre el entregable del día miércoles de la mejora de los modelos. Espero su respuesta. Muchas gracias.</p> <p>Karen Rugerio Armenta A01733228 ITC21</p>	<p><b>Documentación:</b>  <a href="https://github.com/myrosandrade89/TC3007_C_AI/blob/main/documentation/ModeloRefinamiento.pdf">https://github.com/myrosandrade89/TC3007_C_AI/blob/main/documentation/ModeloRefinamiento.pdf</a></p> <p><b>(Es necesario dar click en ‘more pages’ para ver la documentación completa)</b></p> <p><b>Código:</b>  <a href="https://github.com/myrosandrade89/TC3007_C_AI/tree/develop/implementation/fragments/models">https://github.com/myrosandrade89/TC3007_C_AI/tree/develop/implementation/fragments/models</a></p>
	Emplea arquitecturas del estado de arte deep learning como parte de la solución del reto.	<p><b>Entre los comentarios de éste meeting se sugirió:</b></p> <ol style="list-style-type: none"> <li>1. Realizar matrices de confusión como métrica de comparación de modelos.</li> <li>2. Dar más contexto sobre el reto.</li> <li>3. Documentar claramente por qué se eligió el modelo que se eligió.</li> <li>4. Realizar un análisis de los resultados obtenidos</li> </ol>	
	Elabora un modelo benchmark (regresión lineal, logística o un modelo arima) para la evaluación de la solución y explica detalladamente la forma en la que se genera el modelo.	<p>Así mismo se pidió retroalimentación para el módulo de estadística avanzada:</p> <p> Profesor buenas tardes, me preguntaba si mañana podría darnos a mí y a mi equipo una asesoría sobre el entregable del día miércoles de la mejora de los modelos. Espero su respuesta. Muchas gracias.</p> <p>Karen Rugerio Armenta A01733228 ITC19</p> <p> Translate message to: English   Never translate from: Spanish</p> <p> Carlos Alberto Dorantes Dosamantes To: Karen Rugerio Armenta Tue 29/11/2022 9:29 AM</p> <p>Hola Karen, buen día.</p> <p>Puede ser a la 1:00 pm. Me confirmas?</p>	

Evaluación	Plantea pruebas del estado del arte respaldadas en la literatura, que sean válidas y relevantes en el contexto del reto.	<p><b>Las correcciones sugeridas en este meeting fueron:</b></p> <ol style="list-style-type: none"> <li>1. Añadir más contexto sobre la generalización necesaria para el reto.</li> <li>2. Se sugirió añadir un threshold</li> <li>3. Analizar los resultados de las matrices de confusión.</li> </ol> <p><b>Correcciones:</b> Se realizaron las correcciones correspondientes, a excepción del threshold, ya que se necesita un modelo generalizado</p>	
	Selecciona métricas (o indicadores de desempeño) adecuadas tanto para el modelo como para el reto.	<p><b>Comentarios:</b></p> <p>Hiola Karen, La tabla de comparación me parece bien, tengo dos observaciones.</p> <ol style="list-style-type: none"> <li>1. ¿Cuánto gamos con respecto a una predicción basada en el elemento más común? como les he comentado en clase ese es un dato sumamente importante, comparar con el "baseline" (ver <a href="https://machinelearningmastery.com/implement-baseline-machine-learning-algorithms-scratch-python/">https://machinelearningmastery.com/implement-baseline-machine-learning-algorithms-scratch-python/</a>). Por ejemplo si el más común representa 90 por ciento del dataset y tenemos una precisión del 89% con nuestro algoritmo, quiere decir que el algoritmo es deficiente. Tienen que poner la columna del método "baseline".</li> <li>2. Las conclusiones son un tanto cortas, hay más que comentar al respecto.</li> </ol> <p><b>Correcciones:</b></p> <ol style="list-style-type: none"> <li>1. Se realizó la comparación con el Baseline.</li> <li>2. Se hizo un análisis más profundo de los resultados.</li> </ol>	<p><b>Documentación:</b> <a href="https://github.com/myr-osandrade89/TC3007_C_AI/blob/main/documentation/Evaluaci%C3%B3n.pdf">https://github.com/myr-osandrade89/TC3007_C_AI/blob/main/documentation/Evaluaci%C3%B3n.pdf</a></p> <p><b>Código:</b> <a href="https://github.com/myr-osandrade89/TC3007_C_AI/blob/main/implementation/fragments/evaluation/models_evaluation.ipynb">https://github.com/myr-osandrade89/TC3007_C_AI/blob/main/implementation/fragments/evaluation/models_evaluation.ipynb</a></p> <p><b>Gráfica polar:</b> <a href="https://github.com/myr-osandrade89/TC3007_C_AI/blob/main/implementation/fragments/evaluation/polar_plot.ipynb">https://github.com/myr-osandrade89/TC3007_C_AI/blob/main/implementation/fragments/evaluation/polar_plot.ipynb</a></p>
	Interpreta los resultados del modelo de manera correcta		
Refinamiento	Refina el modelo después de su primera iteración para mejorar su desempeño.	<p><b>Comentarios:</b> <a href="https://docs.google.com/spreadsheets/d/1xkPiQcRArJNv2ucWtkQ3Q_XrQWeG8tlo/edit#gid=1060677826&amp;range=A1:G21">https://docs.google.com/spreadsheets/d/1xkPiQcRArJNv2ucWtkQ3Q_XrQWeG8tlo/edit#gid=1060677826&amp;range=A1:G21</a></p> <p>Los comentarios son de <b>falta de entrega del avance</b>. Es importante mencionar que los comentarios reales de las correcciones se hicieron por medio de un <b>zoom</b>:</p>	<p><b>Documentación:</b> <a href="https://github.com/myr-osandrade89/TC3007_C_AI/blob/main/documentation/ModeloRefinamiento.pdf">https://github.com/myr-osandrade89/TC3007_C_AI/blob/main/documentation/ModeloRefinamiento.pdf</a> (Es necesario dar click en 'more pages' para ver la documentación completa)</p> <p><b>Código:</b></p>

	<p>Prueba con otras arquitecturas (además del benchmark) para ver si obtiene mejores rendimientos</p>	<div><div><div><div><div>KA</div><div>Karen Rugerio Armenta</div></div><div><div>To: Benjamin Valdés Aguirre</div><div>Mon 28/11/2022 12:22 PM</div></div></div><div><p>Profesor buenas tardes, me preguntaba si mañana podría darnos a mí y a mi equipo una asesoría sobre el entregable del día miércoles de la mejora de los modelos. Espero su respuesta. Muchas gracias.</p><p>Karen Rugerio Armenta A01733228 ITC21</p></div></div><div><div><div><div></div><div>Benjamin Valdés Aguirre</div></div><div><div>To: Karen Rugerio Armenta</div><div>Mon 28/11/2022 2:46 PM</div></div></div><div><p>Si Karen, puede ser 800 am Saludos!</p></div></div></div> <div><p><b>Entre los comentarios de éste meeting se sugirió:</b></p><ol style="list-style-type: none"><li>1. Realizar matrices de confusión como métrica de comparación de modelos.</li><li>2. Dar más contexto sobre el reto.</li><li>3. Documentar claramente por qué se eligió el modelo que se eligió.</li><li>4. Realizar un análisis de los resultados obtenidos</li></ol><p><b>Correcciones:</b> Se realizaron todas las correcciones sugeridas.</p></div> <div><p><a href="https://github.com/myrosandrade89/TC3007_C_AI/tree/develop/implementation/fragments/models">https://github.com/myrosandrade89/TC3007_C_AI/tree/develop/implementation/fragments/models</a></p></div>
--	---	---