

**ENUNCIADO DEL PROYECTO DE CURSO
TECNOLOGÍAS EMERGENTES I
M.SC. ING. OSCAR CONTRERAS CARRASCO
GESTIÓN II-2023**

1. INTRODUCCIÓN

El proyecto de curso para la asignatura de Tecnologías Emergentes I consistirá en la realización de un trabajo de investigación en el cual se emplearán los modelos, herramientas y técnicas desarrolladas en clases con el fin de aplicarlas al análisis de un dataset para la resolución de una problemática determinada. Como resultado final, se tendrán dos componentes esenciales:

- Un documento que detalla todo el trabajo realizado, con los apartados que serán expuestos más adelante.
- Una presentación final que será realizada durante los últimos días de clases de la asignatura.

Para el estudiante, el propósito de esta actividad es la realizar un proyecto de investigación que permita poner en práctica los conceptos desarrollados a lo largo de la asignatura de Tecnologías Emergentes I, con énfasis en:

- Pre-procesamiento de datos
- Análisis de datos y feature engineering
- Selección y evaluación de modelos predictivos de Machine Learning

Para el docente, el propósito de esta actividad es el de evaluar las siguientes habilidades de los estudiantes:

- Capacidades de análisis requeridas para llevar adelante un proyecto de investigación, que incluyen: Identificación de una problemática a resolver, planteamiento de objetivos claros, análisis de datos, selección de la mejor alternativa de solución para el problema identificado, y presentación de los resultados obtenidos.
- Habilidades comunicativas y de trabajo en equipo: Comunicación efectiva con sus compañeros de grupo, grado de contribución, liderazgo y organización a fin de lograr los objetivos planteados.

2. DE LA CONFORMACIÓN DE EQUIPOS

El proyecto de curso será realizado en grupos de tres personas por afinidad. En casos muy extremos, se considerará la posibilidad de que puedan ser más estudiantes por grupo. Cada miembro debe contribuir de manera equitativa en la realización del proyecto, y por cada presentación de avances deberán proveer un reporte de su correspondiente grado de contribución.

Pueden existir varios grupos que estén trabajando en la misma problemática o similar. Sin embargo, cada grupo debe proceder de manera totalmente independiente del trabajo que estén realizando otros compañeros, ya que no está permitido realizar plagio del trabajo de otros grupos.

3. DE LAS TEMÁTICAS

El desarrollo del presente proyecto involucra:

- Identificación de una problemática de alcance local o global que pueda ser resuelta haciendo uso de modelos de Machine Learning predictivos.
- Para el abordaje de la problemática se podrán emplear modelos supervisados o no supervisados.
- Selección de un dataset que esté relacionado a la problemática que plantean resolver. Por ejemplo, si el proyecto es acerca de realizar predicciones para el diagnóstico médico de una enfermedad, se deberá conseguir un dataset que contenga los datos relevantes, y luego evaluar diversos modelos que permitan cumplir con ese objetivo.
- Pre-procesamiento de los datos que involucra: Limpieza del dataset, selección e ingeniería de features y análisis exploratorio.
- Evaluación de diferentes modelos haciendo uso de técnicas de cross-validation, selección de hiperparámetros, etc. para determinar la mejor solución para el problema planteado. Para este efecto, deberán hacer uso de diversas métricas relevantes a los ámbitos del aprendizaje supervisado y no supervisado.
- Interpretación de los resultados: Deben realizar un análisis y justificación de los resultados obtenidos en todas las fases del proyecto.

Se recomienda hacer uso de datasets de dominio público, para los cuales incluso es muy probable que encuentren modelos que ya fueron desarrollados y evaluados. Sin embargo, para este proyecto se valorará la contribución propia que cada grupo realice, y con preferencia enfocarse a la resolución de problemáticas relevantes a nuestro medio. De cualquier modo, si logran acceder a datasets provistos por una institución competente para el abordaje de la problemática planteada, esto será valorado positivamente en la calificación.

A continuación se tienen algunas opciones de proyectos que pueden desarrollar (Nota: Esta lista no es limitativa. Si desean realizar un proyecto que esté más allá de los aquí propuestos, pueden proceder de esa manera):

Proyecto	Problemática	Ejemplos de datasets
Diagnóstico de enfermedades	Existen una amplia variedad de enfermedades auto-inmunes y crónicas que representan un problema de salud pública a nivel mundial, entre ellas se	<ul style="list-style-type: none">• Diabetes: https://www.kaggle.com/datasets/mathchi/diabetes-data-set• Cáncer de mama:

	distinguen: El cáncer, la diabetes y la esclerosis múltiple entre otros. Existen una variedad de datasets públicos que pueden emplear para este análisis.	https://archive.ics.uci.edu/dataset/14/breast+cancer <ul style="list-style-type: none"> Enfermedades cardíacas: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset
Calidad del agua y temas afines	El acceso al agua potable es una problemática de alcance global y representa un aspecto esencial para la preservación de la salud de los habitantes. Existen diversos datasets sobre el tema del agua y problemáticas relacionadas. Sin embargo sería interesante si logran encontrar información provista por fuentes locales.	<ul style="list-style-type: none"> https://www.kaggle.com/datasets/adityakadiwal/water-potability https://data.world/datasets/water
Desempeño académico de estudiantes.	Existen una variedad de factores que influyen en el rendimiento académico de estudiantes a diversos niveles, desde la escolaridad hasta la universidad. Existen datasets para modelos predictivos del desempeño académico de un estudiante. Asimismo, existen investigaciones realizadas sobre los factores que influyen en la selección de una carrera versus otra.	<ul style="list-style-type: none"> https://archive.ics.uci.edu/dataset/320/student+performance https://www.kaggle.com/datasets/spscientist/students-performance-in-exams
Sistemas de recomendación	Los sistemas de recomendación se emplean para proponer productos o sugerencias que estén alineadas a las preferencias de un usuario en particular, o bien de usuarios con características similares. Es así que sitios como Amazon, eBay entre otros implementan este tipo de sistemas para ofrecer a sus usuarios diversas sugerencias.	<ul style="list-style-type: none"> https://github.com/caserec/Datasets-for-Recommender-Systems
Análisis de sentimientos (Sentiment analysis)	Este tipo de aplicaciones tienen como objetivo determinar si un comentario emitido por un usuario es positivo o negativo.	<ul style="list-style-type: none"> https://www.repustate.com/blog/sentiment-analysis-datasets/

	Variaciones de los sistemas de análisis de sentimientos, también incorporan rankings del grado de satisfacción expresado en un comentario textual.	
--	--	--

Por supuesto que pueden elegir otros proyectos más allá de los descritos en este listado o que estén relacionados, sin embargo deben tomar en cuenta las consideraciones que se presentan a continuación:

- **Proyectos con modelos para visión computacional:** En este ámbito se tienen los modelos para clasificación de imágenes, reconocimiento facial, detección de objetos, etc. Si bien este ámbito está muy avanzado en la actualidad, para trabajar en este tipo de proyectos, ustedes requieren conocer sobre redes neuronales convolucionales y modelos más avanzados que normalmente son desarrollados en Tecnologías Emergentes II. Si eligen un proyecto de estas características, asegúrense de solicitar la guía del docente.
- **Chatbots, modelos de lenguaje, modelos generativos, etc:** La tecnología ha tenido avances muy importantes en últimos años. Sin embargo, una gran parte de estos modelos están basados en tecnologías específicas y redes neuronales transformer que también son desarrolladas en Tecnologías Emergentes II. Si eligen un proyecto de estas características, deben consultar con carácter obligatorio al docente para analizar la factibilidad.

Como alcances del proyecto, no es necesario que desarrollen una aplicación móvil o web funcional implementado el mejor modelo. Sin embargo, si las características del proyecto lo ameritan, pueden además crear una aplicación dirigida al usuario final. Sobre estos temas, deben consultar necesariamente al docente.

4. DE LA EVALUACIÓN DEL PROYECTO FINAL

Se realizará en tres fases:

Momento evaluativo	Aspectos mínimos a evaluar	Porcentaje
Primero	<ul style="list-style-type: none"> • Título del proyecto. • Análisis de la problemática. • Objetivos: General y específicos. • Justificación técnica, económica y social. 	5%
Segundo	<ul style="list-style-type: none"> • Marco teórico (relevante para el proyecto) • Pre-procesamiento del dataset al 100% 	10%

Final	<ul style="list-style-type: none"> • Evaluación de modelos concluida. • Interpretación de resultados concluida. • Conclusiones, recomendaciones. • Presentación oral del proyecto. 	70%
-------	--	-----

Los avances del proyecto serán reportados en un documento escrito con los siguientes apartados:

Introducción

Capítulo 1: Generalidades

- 1.1. Análisis de la problemática
- 1.2. Objetivos:
 - a) Objetivo general
 - b) Objetivos específicos
- 1.3. Justificación
 - a) Técnica
 - b) Económica
 - c) Social

Capítulo 2. Marco teórico: Debe ser relevante al proyecto que desarrollan y no simplemente una copia de fuentes externas. Todos los autores consultados deben tener citas en formato APA. No está permitido hacer plagio de las fuentes consultadas. Proyectos que incurran en esta irregularidad tendrán cero puntos sin derecho a reclamo posterior. Se les recomienda consultar con el docente si sus citas bibliográficas son adecuadas para evitar cualquier observación posterior.

Capítulo 3. Desarrollo del proyecto:

- 3.1. Descripción del dataset
- 3.2. Pre-procesamiento del dataset: Haciendo énfasis en selección de features, feature engineering, limpieza del dataset, etc.
- 3.3. Entrenamiento de modelos
- 3.4. Evaluación de modelos
- 3.5. Interpretación de resultados

Conclusiones

Recomendaciones

Bibliografía (En formato APA y por cada cita bibliográfica en el documento)

Al finalizar el semestre, cada grupo realizará una presentación oral en clases de todo el trabajo realizado. Tendrán un tiempo de 15 minutos para realizar su intervención.

La evaluación del proyecto será rigurosa, por lo que solamente los proyectos que ostenten una calidad como para ser publicados en jornales científicos tendrán calificaciones de 90% o superior.

5. PUBLICACIÓN DE PROYECTOS

Los estudiantes que estén interesados en que sus proyectos sean publicados en algún jornal científico, por ejemplo el Journal Boliviano de Ciencias (Univalle), deben hacerlo conocer al docente a la brevedad, ya que se deben coordinar los aspectos de pertinencia, así como la autoría y el formato del artículo científico a ser publicado.