

## PROYECTO FINAL

### 1. Introducción

El proyecto final del curso Inteligencia Artificial 1 es un trabajo grupal (**mínimo 2 y máximo 3 estudiantes por grupo**) que busca que los estudiantes desarrollen una solución a un problema real usando modelos de analítica y conjuntos de datos de diferente formato. Cada grupo debe entender el problema, investigar su contexto, los antecedentes, establecer la metodología de trabajo y proponer las métricas de desempeño para evaluar el progreso. Se espera que en el desarrollo se entrenen y evalúen diferentes modelos de analítica. Para cada modelo deberá ajustar adecuadamente los hiperparámetros y evaluar los resultados que obtenga con base en métricas predefinidas. Cada grupo podrá usar la metodología CRISP-DM que se verá en el curso y adaptarla a las necesidades de su proyecto.

### 2. Caso de Estudio Propuesto: Clasificar géneros de canciones

**Recomendaciones muy acertadas** - ¿Cómo es que una lista de reproducción conoce tan bien a su usuario?

En los últimos años, los servicios de transmisión (streaming) con grandes catálogos se han convertido en el medio principal a través del cual la mayoría de las personas escuchan su música favorita. Pero al mismo tiempo, la gran cantidad de música que se ofrece puede significar que los usuarios pueden sentirse un poco abrumados cuando intentan buscar música más nueva que se adapte a sus gustos.

Por esta razón, los servicios de streaming han buscado formas de categorizar la música para permitir recomendaciones personalizadas. Un método involucra el análisis directo de la información de audio sin procesar en una canción dada, calificando los datos sin procesar en una variedad de métricas. En este proyecto, examinarán los datos recopilados por un grupo de investigación conocido como The Echo Nest. El objetivo es revisar este conjunto de datos y clasificar las canciones como 'Hip-Hop' o 'Rock', **todo sin escuchar ni una sola**. Al hacerlo, deberá limpiar los datos, realizar algunas visualizaciones exploratorias de los mismos, incrementar la cantidad de registros del dataset original obteniendo nuevos datos de fuentes confiables, limpiar y preparar los datos para usarlos con modelos de ML. También deberá usar la reducción de características con el objetivo de alimentar los datos a través de algoritmos de aprendizaje automático.

### 3. Evaluación y entregables

Se evaluará la calidad del trabajo mediante respuestas a las siguientes preguntas: ¿la metodología es clara y robusta? ¿son razonables las aproximaciones que se hicieron en el proyecto? ¿los datos se exploraron y procesaron de forma adecuada? ¿las soluciones propuestas son ingeniosas e interesantes? ¿explican correctamente los impactos de la solución en el ámbito global, ambiental,

social y económico? ¿complementaron los datos iniciales? ¿los estudiantes desarrollaron y transmitieron conocimientos no triviales sobre el problema y/o los algoritmos y modelos? ¿el trabajo demuestra el desarrollo de las competencias que se definieron para este curso? Recuerde que es importante que presente una discusión sólida sobre el diseño, desarrollo y resultados de su proyecto. Debe mostrar cómo se compara su trabajo con resultados de otros autores.

Previo a la primera entrega, un representante de cada grupo debe enviar un correo con el nombre del proyecto en cuestión, los integrantes de grupo y el enlace al repositorio de entrega en **OneDrive** a: [kelyn.botina@u.icesi.edu.co](mailto:kelyn.botina@u.icesi.edu.co). El repositorio debe contener tres carpetas, una para cada entrega (i.e.: “Entrega1”, “Entrega2”, “Entrega3”). En cada carpeta sólo deben estar los siguientes archivos y subcarpetas:

- ☐ Copia en .mp4 del video de la presentación (no debe exceder 10 minutos).
- ☐ Copia en PDF de la presentación en inglés.
- ☐ Copia en PDF del reporte con los avances en Inglés (máximo 3 páginas para el reporte 1, máximo 5 para el reporte 2 y máximo 7 páginas para el reporte 3)\*.
- ☐ Copia en PDF de un informe en inglés sobre el análisis del impacto global, ambiental, social y económico de las soluciones basadas en IA, con una sección dedicada a la solución que implementan en el proyecto. (Máximo 4 páginas).
- ☐ Carpeta de “Datos” con todos los datos que utilizó en el proyecto.
- ☐ Carpeta de “Código” con todo el código fuente que desarrolló, debidamente comentado.

**Primer conjunto de entregables: semana 7 (septiembre 11 y septiembre 15 de 2022).** El enfoque de esta entrega es (son) la(s) pregunta(s) de interés, el tipo de problema al que pertenece, la metodología, las métricas que utilizarán para medir el progreso, los datos recolectados, el análisis exploratorio de los datos y los siguientes pasos que se tomarán en el proyecto.

En esta entrega deben también proponer estrategias para conseguir más datos que les permita incrementar su conjunto de datos, pueden revisar los siguientes enlaces:

<https://www.kaggle.com/datasets/geomack/spotifyclassification>  
<https://opendatascience.com/a-machine-learning-deep-dive-into-my-spotify-data/>  
<https://developer.spotify.com/documentation/web-api>  
<http://millionsongdataset.com/tasteprofile/>

En esta entrega también se debe comenzar el análisis de los impactos en el ámbito global, ambiental, social y económico de las soluciones basadas en IA, en un texto que no supere las dos páginas.

**Segundo conjunto de entregables: semana 12 (octubre 16 y octubre 20 de 2022).** El enfoque de esta entrega son la estrategia implementada para la obtención de nuevos datos, la preparación de los datos, el entrenamiento de los modelos (incluido el ajuste de hiperparámetros), los resultados obtenidos (métricas, gráficas, etc.), y el plan de despliegue. Además, se deberá realizar un análisis

inicial de los impactos en el ámbito global, ambiental, social y económico de la solución que implementaron.

**Tercer conjunto de entregables: semana 17 (noviembre 20 y noviembre 24 de 2022).**

El enfoque de esta entrega son la reducción de características, evaluación de resultados, el despliegue de la solución, la entrega al “cliente” y el reporte final. Se espera que el nivel de profundidad de su análisis y la calidad de sus resultados haya mejorado desde la segunda entrega. Además, deberá presentar el análisis final de los impactos en el ámbito global, ambiental, social y económico de la solución que implementaron.

**\*Notas:**

1. Es muy importante que el código fuente esté bien documentado. Si utiliza datos y código fuente de terceros debe referenciarlos de forma clara y explícita, de lo contrario se considerará como fraude.
2. Los informes deben presentarse en Inglés y contener detalles explicativos claros y concisos. Procure incluir diagramas de flujo, diagramas de bloques u otras figuras que ilustren su metodología, arquitectura de software y resultados tanto en el reporte como en la presentación. Procure utilizar gráficos con calidad vectorial.

**Estructura básica del reporte final (máximo 7 páginas):**

- i. Title.
- ii. Abstract.
- iii. Introduction: context, description of the problem, why it is interesting?
- iv. Theory: What should the reader know to understand your development? Note: carefully select what should go in this section, avoid including generalities.
- vi. Methodology: How did you approach your project? Note: I'm not interested in seeing an exact copy of the CRISP-DM diagram.
- vii. Results: How did the models perform on different datasets? Metrics of interest for the specific problem.
- viii. Results analysis: What do you observe in your results? Are your models generalizing well? Is there overfitting? What's working well? What's failing? How do your results compare to others reported in the literature?
- ix. Conclusions and Future Work: What did you do? What did you learn? What could be improved?
- x. Bibliographic References: Only include articles, books, or digital material that you have read and used.

**Observación:** Revise con detenimiento artículos publicados en conferencias de interés como NIPS, ICML o ICLR, etc. antes de redactar sus reportes finales.