

**Resumen Elaborado Por:** *Jesús Alfredo Sanabria Mejía*

La clase del 17 de febrero se centró en dos temas principales:

## 0.1 Reglas de clase

El docente estableció las normas de comportamiento y participación para el curso, se indicó que la metodología estaría organizada alrededor de un datalab

Se indicaron un conjunto de herramientas que serán empleadas y de utilidad en el desarrollo del curso

1. Overleaf: Para facilitar la presentación de trabajos, en formato  $\text{\LaTeX}$
2. Github: Como repositorio de código y manejador de versiones para los avances y desarrollos que se presenten en la asignatura.
3. Uso de IA: Para explorar funcionalidades y contenidos como Gemini, OpenIA
4. DashBoard: Es un instrumento de acceso rápido en el que se registraran algunas tareas y se tienen links a materiales de clase.
5. Acceso a Contenidos CRAI: El Centro de Recursos para el Aprendizaje y la Investigación (CRAI) es un entorno común que reúne todos los servicios universitarios de apoyo al aprendizaje.
6. Grupo de Whataspp: Se recomienda participar activamente en el grupo de WhatsApp para estar al día con las noticias del curso y reportar por allí algún inconveniente si llegare a presentarse.

Entre las actividades Generales a Desarrollar se encuentran:

1. Resumen de Noticias: Se espera que los estudiantes registren semanalmente los acontecimientos mas relevantes que ocurran con relación al tema de automatización, inteligencia artificial, machine learning y demas propios de la asignatura.
2. Registro de Libros: Se deben registrar algunos libros sobre la asignatura para recomendar a los compañeros, de acuerdo al formato se hara una mini exposicion de cada uno.
3. Resúmenes de Clase: Se organizan grupos con el fin que posterior a cada una de las sesiones compendien en un documento las ideas principales de la clase y compartirlas con el grupo.
4. Exposiciones Grupales: Se organizan las fechas de exposicion y los temas a presentar.
5. Proyecto de Clase: se llevara a cabo un proyecto de flujo de datos como resultado de la asignatura.

## 0.2 Introducción al tema de automatización

Se realizó una introducción general al concepto de automatización en el contexto de flujos de datos para Machine Learning. La automatización se refiere al uso de herramientas y tecnologías para realizar tareas de forma automática. En el contexto de Machine Learning, la automatización es fundamental para el desarrollo de flujos de datos eficientes.

Durante la sesión de clase se comentó acerca de la importancia de la gestión de datos y la automatización en la construcción de sistemas. Destacando los desafíos que presentan la variedad, el volumen y la velocidad de los datos, y la necesidad de encontrar mejores enfoques para manejar estos desafíos en diferentes escenarios de uso.

Así también se indicó la importancia de comprender los conceptos fundamentales de la automatización y desarrollar habilidades en el uso de herramientas de automatización e integración de datos. Además, se menciona la implementación de soluciones de automatización y la integración de datos en proyectos de vida. Se hace referencia a la Robótica de Procesos Automatizados (RPA) y a cómo la inteligencia de datos permite analizar grandes cantidades de información de manera rápida y precisa, identificando patrones y tendencias que pueden pasar desapercibidos para los procesos tradicionales de tratamiento de información, se destaca que estas máquinas pueden ser programadas para tomar decisiones basadas en los datos analizados, lo que puede mejorar la eficiencia y la productividad.

Un flujo de datos pipeline es un conjunto de procesos secuenciales que trabajan juntos para mover, transformar y procesar datos de forma eficiente y automatizada. Es como una tubería que transporta los datos desde su origen hasta su destino final, realizando diferentes tareas en el camino. Componentes de un flujo de datos pipeline:

- **Entrada:** La fuente de datos de donde se originan los datos. Puede ser un archivo, una base de datos, una API o cualquier otro sistema que proporcione datos.
- **Extracción:** El proceso de recuperar datos de la fuente.
- **Transformación:** El proceso de modificar y limpiar los datos para que sean adecuados para su análisis. Esto puede incluir tareas como eliminar valores duplicados, corregir errores, convertir formatos, normalizar valores y agregar nuevas características.
- **Carga:** El proceso de almacenar los datos transformados en un destino final. Este destino puede ser un almacén de datos, un lago de datos, una base de datos o cualquier otro sistema que pueda almacenar datos.
- **Análisis:** El proceso de explorar y analizar los datos para obtener información y conocimiento.

Beneficios de un flujo de datos pipeline:

- **Eficiencia:** Automatiza el procesamiento de datos, lo que reduce el tiempo y el esfuerzo manual.
- **Escalabilidad:** Se puede adaptar a grandes volúmenes de datos.
- **Confiablez:** Reduce la posibilidad de errores humanos.

- Reproducibilidad: Permite ejecutar el mismo proceso con diferentes conjuntos de datos.
- Mejora la toma de decisiones: Facilita el acceso a información valiosa para la toma de decisiones estratégicas.

**Resumen Elaborado Por:** *Jesús Alfredo Sanabria Mejía*