Universidad de los Andes
Minería de Datos
Profesor: Rafael Amaya Gómez
Fecha: 15/05/2025

PROYECTO FINAL

Este curso brinda una introducción a la minería de datos como parte fundamental de la analítica de datos. Se puede visualizar como una caja de herramientas para la resolución de problemas propios que involucren diferentes tipos de manejo de datos, contemplando el análisis de criterios para determinar su validez en su aplicación y brindar herramientas para soportar el proceso de toma de decisiones. Siguiendo la lógica de los contextos actuales, es usual trabajar en equipo y complementarnos en pro de un objetivo claro y definido en una compañía o proyecto. Es por esto, que este trabajo contempla la creación de un equipo de trabajo que proponga, desarrolle y resuelva un problema real, aprovechando todas las herramientas que se han aprendido hasta el momento. Además, reconociendo que los tomadores de decisiones no son expertos como ustedes en este tipo de herramientas, por lo que requieren una forma visual y fácil de obtener los resultados esperados para soporta la toma de decisiones.

METODOLOGÍA

Cada grupo tendrá que proponer un problema a partir de datos reales en los que se resuelva una pregunta de investigación o de negocio y que contemple las herramientas vistas en el curso. Dentro de estas preguntas se pueden contemplar ejemplos como:

- ¿Cómo podemos identificar grupos de clientes con características similares para personalizar nuestras estrategias de marketing?
- ¿Qué patrones de comportamiento se observan en diferentes segmentos de clientes y cómo podemos adaptar nuestras ofertas para satisfacer sus necesidades específicas?
- ¿Cuáles son los factores clave que influyen en las ventas de nuestros productos o servicios?
- ¿Podemos utilizar datos históricos para predecir las ventas futuras y ajustar nuestra estrategia de inventario y marketing en consecuencia?
- ¿Cómo podemos identificar transacciones sospechosas o fraudulentas en tiempo real?
- ¿Qué patrones anómalos en los datos financieros podrían indicar actividades fraudulentas?
- ¿Cómo podemos reducir los tiempos de producción, minimizar los errores y optimizar los recursos utilizando técnicas de minería de datos?

Para ello, se propone que cada grupo haga las veces de un equipo consultor que va a proponer una solución a esta problemática, implementando una aplicación o interfaz interactiva que muestre los resultados obtenidos a una junta directiva. Por lo que se le solicita como posibles soluciones una aplicación con base en R usando Shiny o en Python usando Dash (no se debe hacer una solución en ambos casos, no habrá bonificación).

Para el caso de Shiny puede consultar los siguientes recursos en línea:

- Shiny Basics
  https://shiny.posit.co/r/getstarted/shiny-basics/lesson1/index.html
- Minicurso Shiny (@Enriqueloper)
  https://www.youtube.com/watch?v=vjbdGteuasU&list=PLdV8ntSOIL5TgPHo4Gp2esAaW2M4sDsCg&index=1

Para el caso de Dash puede con los siguientes recursos en línea

- Tutorial Dash Plotly
  https://dash.plotly.com/tutorial

- Minicurso Plotly Dash (Data Science Tutorials)
  https://www.youtube.com/watch?v=Ma8tS4p27JI&list=PLH6mU1kedUy8fCzkTTJlwsf2EnV_UvOV-&index=1

Cada herramienta diseñada en Shiny (R) o Dash (Python) debería incluir a lo sumo:

- Texto fijo con una descripción breve del problema a resolver.
- Texto fijo con la descripción de los datos usados: tipo de datos, fuente y tratamiento usado.
- Visualizaciones de los resultados que permitan facilitar la comunicación de estos.

Cada grupo estará compuesto a lo sumo de cuatro integrantes, que pueden adoptar alguno de los siguientes roles sugeridos*:

- Diseñador herramienta. Esta persona lidera la creación de la aplicación en Shiny o Dash y acopla las soluciones obtenidas por los demás miembros del equipo
- Escritor: Esta persona lidera la escritura de un documento corto, pero autocontenido que explique a la junta directiva la problemática, metodología propuesta y resultados obtenidos en un resumen ejecutivo.
- Líder solución: Persona que lidera la solución del problema a través de un código en R o Python. Se encarga de coordinar a los demás integrantes para llegar a resultados que respondan la pregunta de negocio o investigación.
- Apoyo de solución: Persona que apoya al líder en la solución del problema y se encarga de hacer un R MarkDown o archivo ipynb en Python explicando paso a paso cada segmento del código propuesto.

*Cabe resaltar que estos roles sugeridos no implican que no haya una contribución en los diferentes ámbitos del proyecto y que estos roles no necesariamente son fijos, sino que por el contrario se sugieren desde una perspectiva más flexible.

CONTENIDO DE LA ENTREGA

La entrega consiste en cuatro elementos: 1) Herramienta en Shiny o Dash, 2) archivo R MarkDown o ipynb, 3) Resumen Ejecutivo y 4) vídeo en YouTube explicando su solución. Los cuatro elementos se deben entregar en un archivo .zip con el número del grupo como se muestra en el siguiente ejemplo: ***Grupo1_Proyecto.zip***. Su número de grupo estará disponible en Bloque Neón.

A. Sobre la herramienta se espera que:
   1. Muestre los resultados obtenidos (preferiblemente a través de visualizaciones)
   2. Si es posible, permita una interacción con el usuario a través de una predicción o clasificación de un nuevo registro
   3. Responda a la pregunta de investigación o de negocio.

B. Sobre el archivo MarkDown o ipynb:
   1. Describa y caracterice cada variable de la base de datos
   2. Explique cada función o herramienta de tratamiento y visualización de datos
   3. Explique el funcionamiento de cada método implementado para la solución del problema
   4. Incluya todo el código usado para llegar a la solución
   5. Especifique las versiones de R o Python implementada y todas las librerías cargadas.

C. El resumen ejecutivo se espera que incluya (máx. 5 páginas):
   1. Breve introducción donde se menciona el tipo de problemas que van a resolver.
   2. Descripción de los métodos implementados en la solución del problema
      a. Explicación de por qué se seleccionó ese método

    b. Consideraciones tenidas en cuenta, por ejemplo, Cross-validación, Bootstrapping, Bagging, entre otras, para la selección de parámetros.
3. Resultados obtenidos con el modelo planteado.
4. Presentación básica del funcionamiento de la herramienta
    a. ¿Cuáles funcionalidades usaron?
    b. ¿Cómo se usa su herramienta?

D. Sobre el vídeo en YouTube se espera que:
1. Dure máximo 5 minutos.
2. El grupo decide la forma de exponer, pero todos deben tener las cámaras prendidas en la sustentación.
3. Presente claramente la problemática, aproximación implementada y resultados obtenidos.
4. Expliquen cómo funciona su herramienta y como responde a la pregunta de negocio o investigación.

CRONOGRAMA

La entrega del proyecto se debe realizar a más tardar el viernes 31 de mayo a las 11:59pm. Deben hacer una entrega por grupo y enviar el link del vídeo en Youtube.

EVALUACIÓN Y CALIFICACIÓN

El proyecto se evaluará usando los siguientes criterios:

1. Herramienta (20%)
    a. Diseño visual (10%)
    b. Funcionalidad (10%)
2. Archivo MarkDown o ipynb (30%)
    a. Caracterización base de datos (10%)
    b. Explicación métodos de tratamiento y visualización de datos (10%)
    c. Explicación del método usado y sus consideraciones (10%)
3. Resumen ejecutivo (40%)
    a. Presentación de la problemática (10%)
    b. Estrategia solución (10%)
    c. Explicación de resultados obtenidos (10%)
    d. Presentación de la herramienta (10%)
4. Vídeo (10%)
    a. Uso del tiempo (5%)
    b. Organización del vídeo (5%)