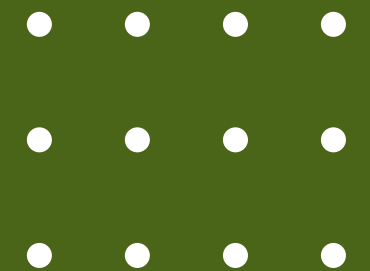
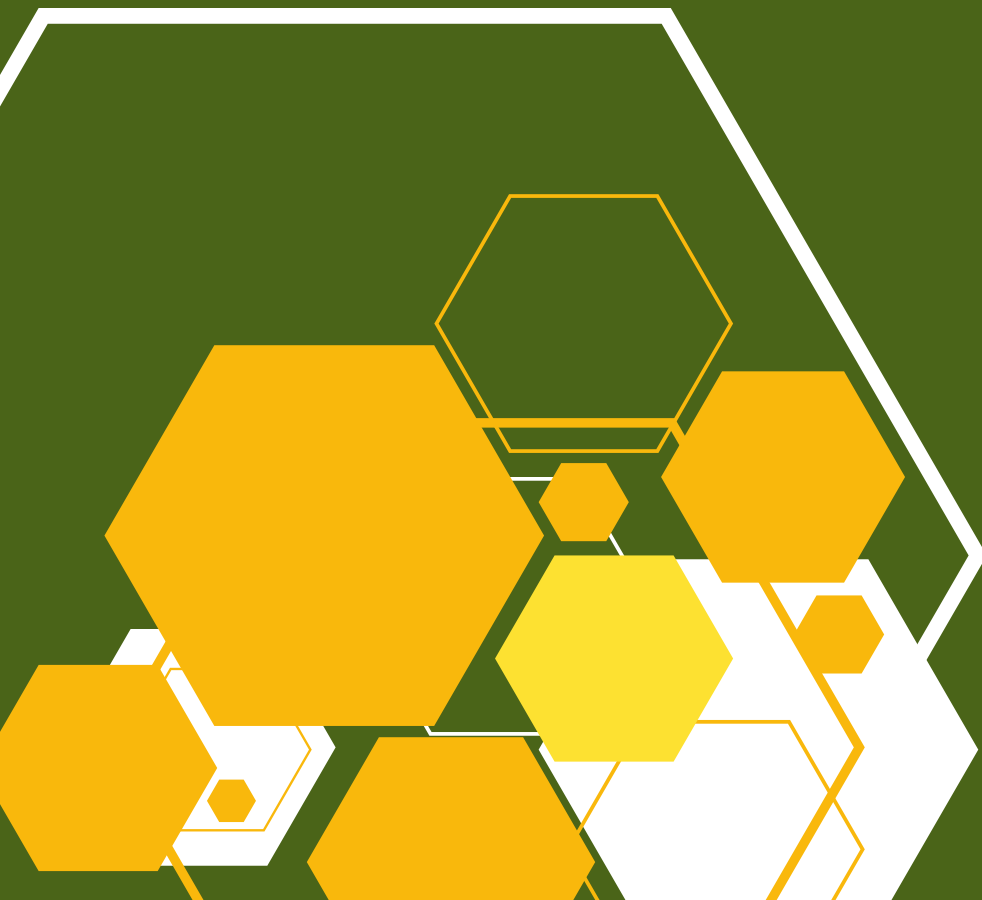


ALGORITMOS II

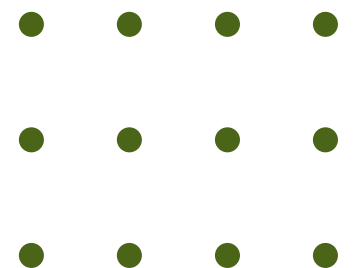
LIBRERÍA DE MODELOS PREDICTIVOS

Integrantes
Scafati, Ayelén
Llanes, Federico
Turri, Bautista

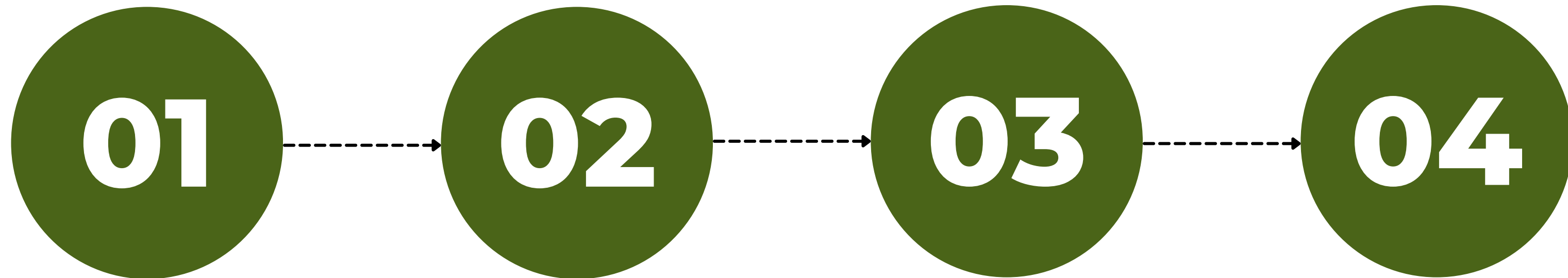


Objetivo

El objetivo de este proyecto es elaborar una librería que nos permita crear modelos predictivos mediante una versión simplificada del ensamble RandomForest. Buscamos entrenar un modelo con un conjunto de datos etiquetados, utilizando técnicas de aprendizaje supervisado, para predecir la variable objetivo. Inicialmente, nos enfocaremos en problemas de clasificación, con la posibilidad de extender la funcionalidad a problemas de regresión en el futuro.



Actividades y fases del proyecto



Investigación del tema

Planeación de diseño

Desarrollo en código

Revisión de los resultados

Metodología del proyecto

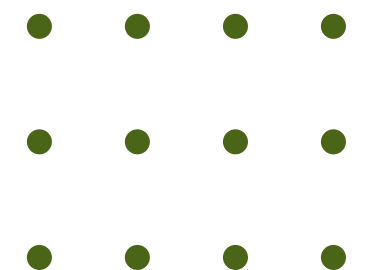
- Reuniones por Meet semanales.
- Grupo de WhatsApp para coordinar reuniones y comunicar avances.
- Repositorio GitHub para subir los avances.
- Calendar de Google para agendar las reuniones y fechas clave.

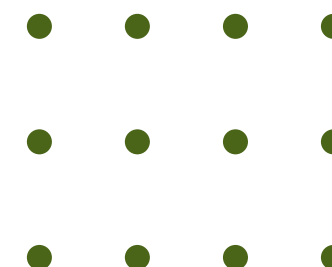
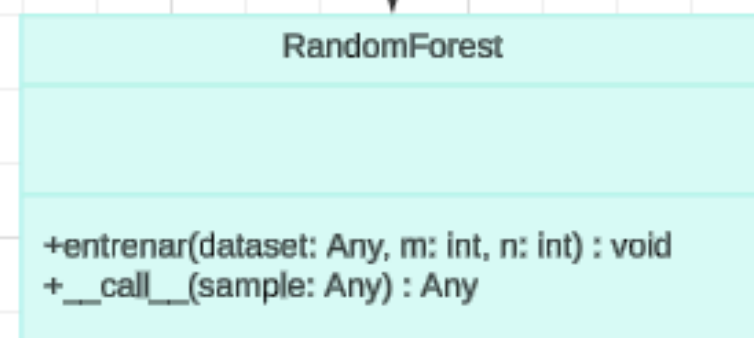
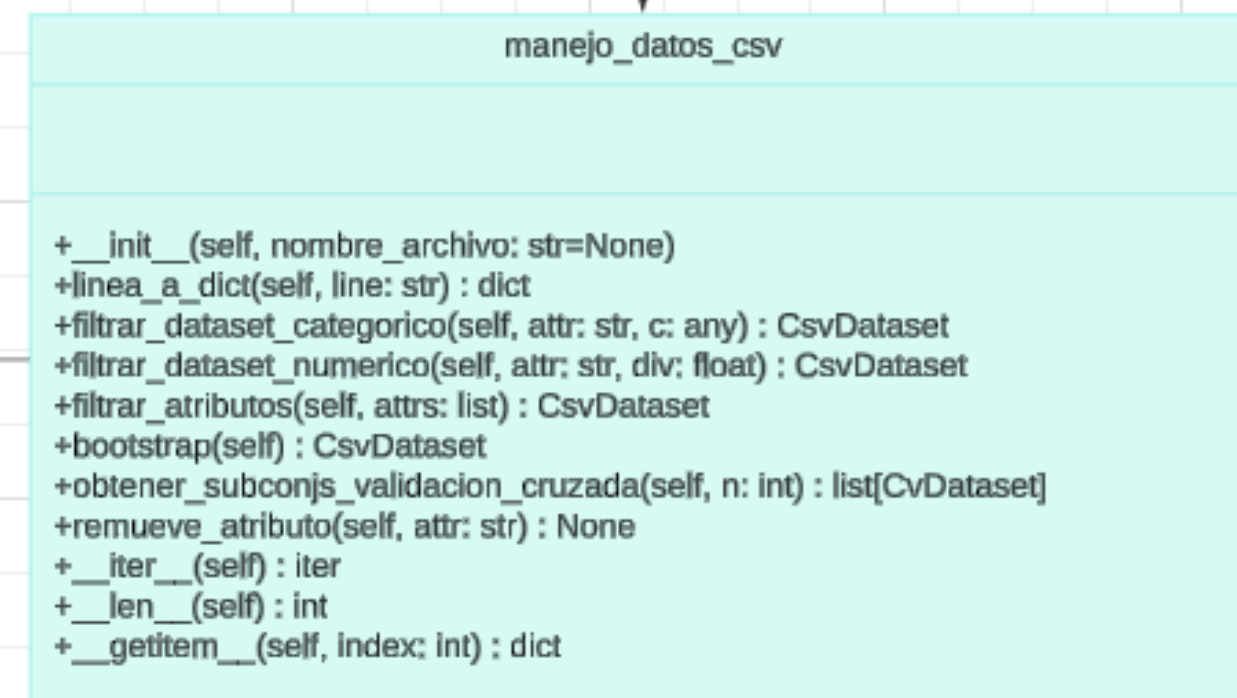
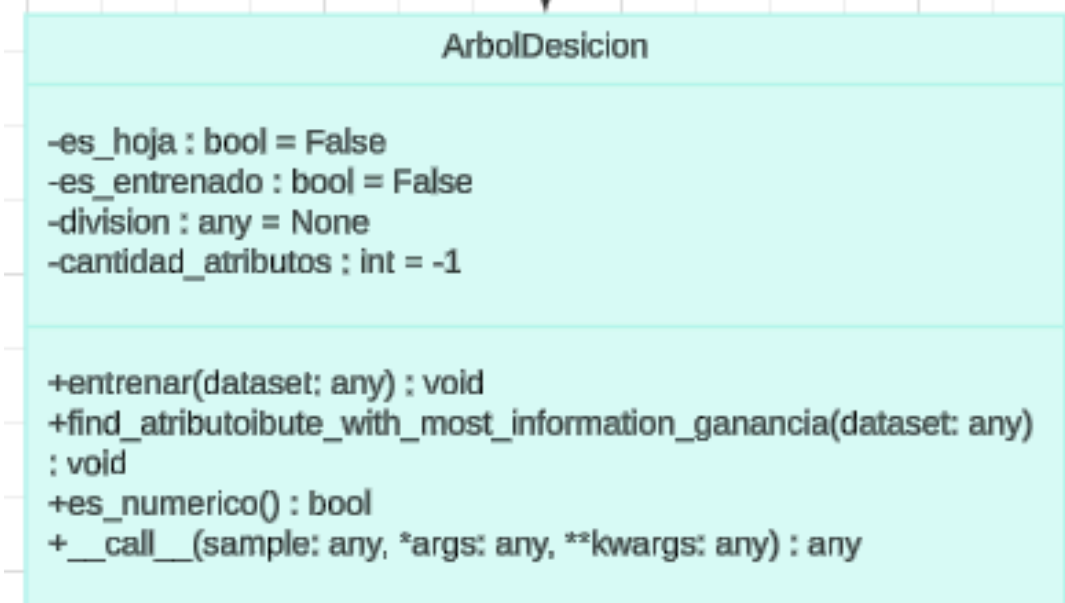
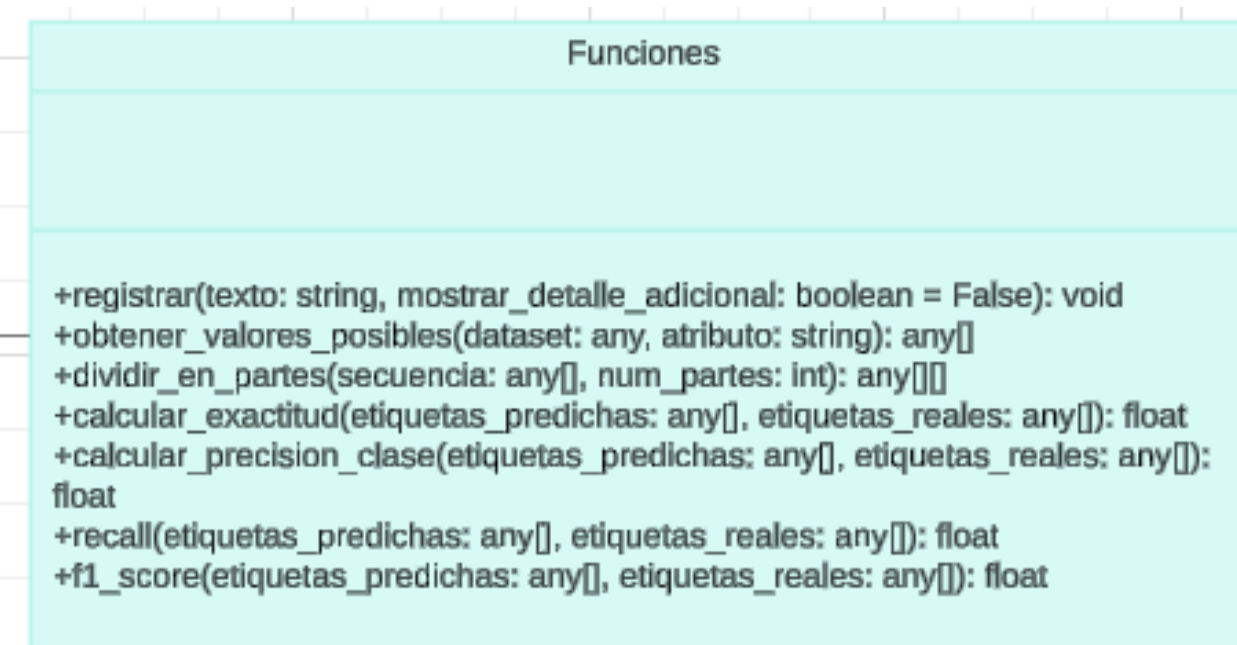
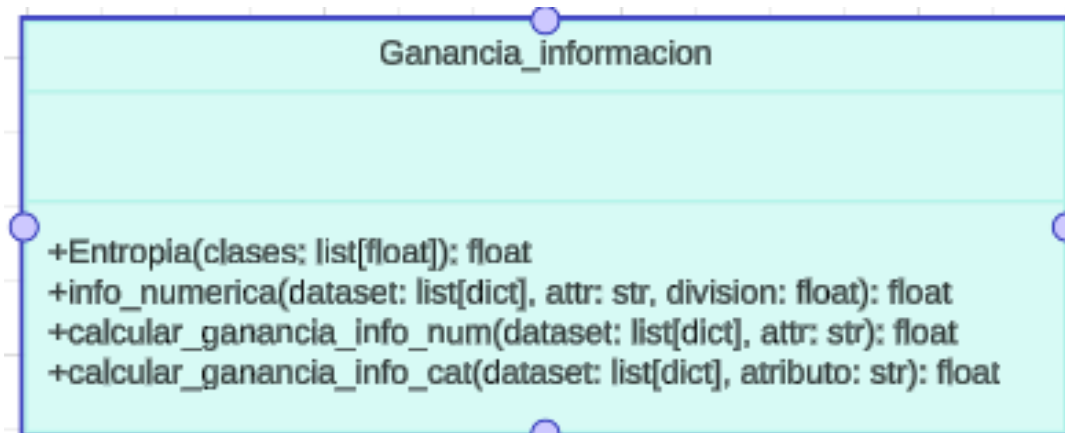


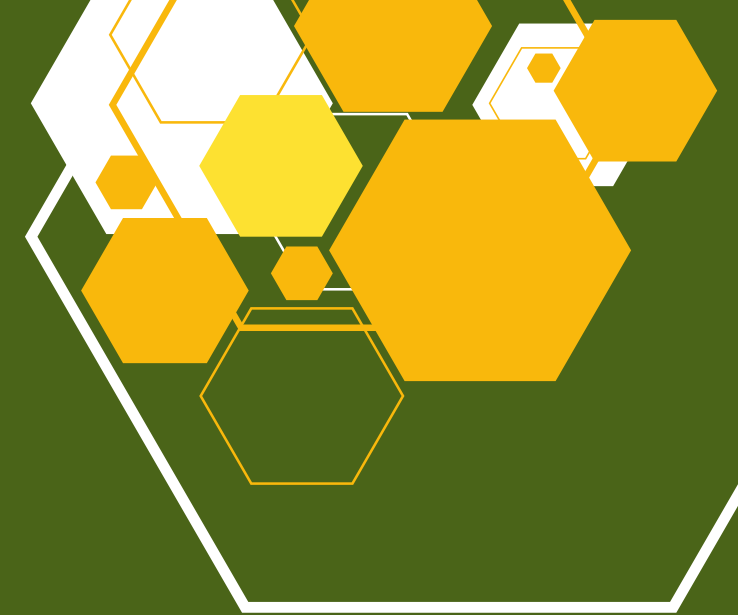
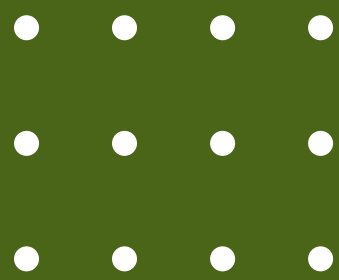
Decisiones de diseño



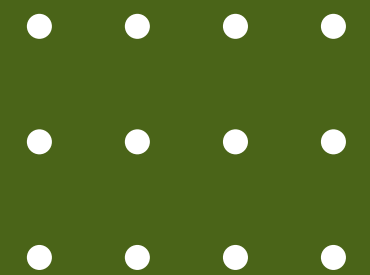
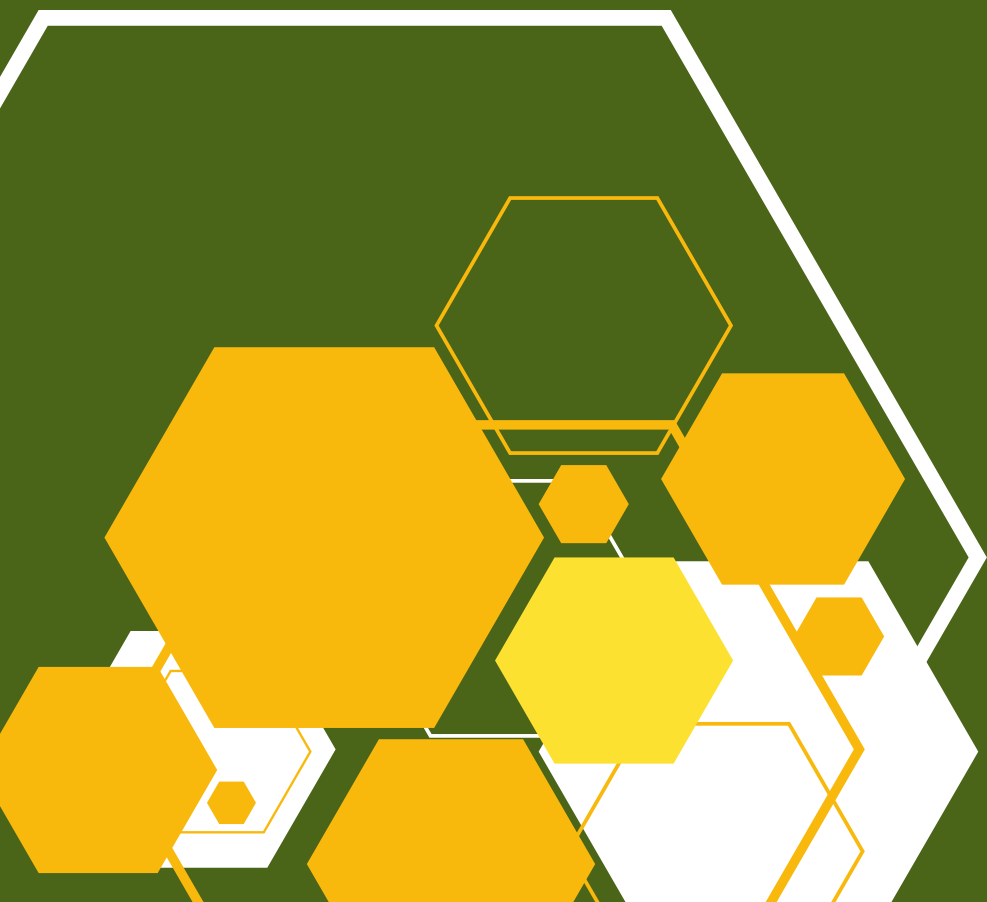
- Se utilizó la librería de pandas para trabajar y manipular los datasets.
- Se utilizó la librería random para la selección aleatoria de columnas para el Random Forest.
- Se implementó una estructura de tipo árbol N-ario.
- El código se implementó utilizando el paradigma de programación orientada a objetos.
- El código fue modularizado en distintas clases.
- Se ha implementado el algoritmo ID3, con extensiones características del C4.5 que permiten el manejo de atributos continuos.





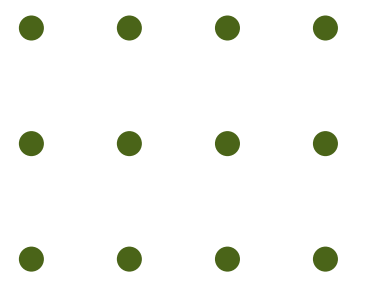
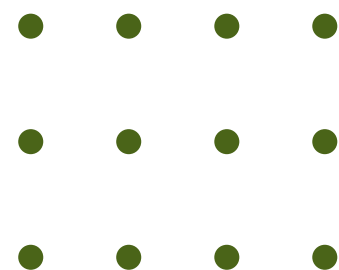


Demostración del código



Dificultades encontradas

- Falta de conocimiento general en herramientas de aprendizaje automático.
- Adaptación del código para valores categóricos y numéricos.
- El Random Forest tenía problemas para dar buenas precisiones con datasets pequeños.
- Intento de adaptación del ID3 al C4.5.

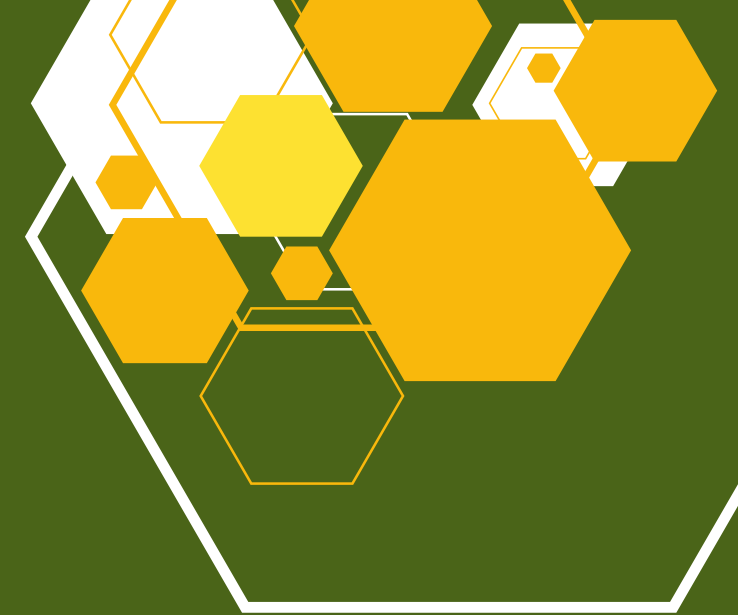
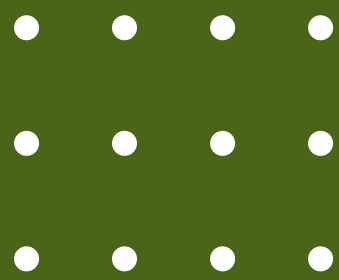


A futuro



- Mejorar la impresión de los árboles.
- Implementar podado.
- Manejar de forma más eficiente las variables continuas.
- Extender nuestro algoritmo mediante la incorporación de Gain Ratio y otras características del algoritmo C4.5.





GRACIAS POR SU ATENCIÓN

¿Preguntas?

