### **Slide 1: Título de la Presentación**

Título: Predicción del Período Orbital de Asteroides Utilizando Machine Learning

Contenido:

* Tu nombre/mails
* Fecha
* Nombre del curso o programa
* Grupo Nº 5

### **Slide 3: Introducción al Proyecto**

Título: Introducción

Contenido:

* Breve descripción del proyecto
* Importancia de predecir el período orbital de los asteroides

Guión:

"En nuestro proyecto, nos enfocamos en predecir el período orbital de los asteroides en base a diversas características físicas y orbitales. Consideramos que esto nos ayudará a conocer la trayectoria de estos cuerpos celestes y a evaluar posibles riesgos de impacto con la Tierra." **PASAR SLIDE**

### **Slide 4: Datos Utilizados**

Título: Datos Utilizados

Contenido:

* Fuente de los datos: <http://neo.jpl.nasa.gov/> (no sé si convenga poner que es un banco de datos de la NASA)
* El dataset tiene 839714 observaciones y 31 atributos (poner al costado una tablita con nombre de algunos pocos features con su tipo de dato . Sí o sí que aparezcan per y a al menos)
* Descripción de las características (semieje mayor, excentricidad, inclinación, etc.)

Guión:

“Para trabajar en nuestro problema utilizamos un dataset de la Nasa. Este se compone de 839714 observaciones y 31 atributos que en su mayoría son de tipo flotante. Entre los que podemos destacar :

* **per (periodo orbital)** indica la cantidad de días que tarda el asteroide al completar una vuelta alrededor del Sol. Esta será nuestra variable de estudio o variable objetivo
* **a ( semieje mayor)** mide la distancia promedio desde el foco de la elipse hasta su centro en ua (1 UA equivale a 150\*10^6 KM)
* **e (excentricidad)** describe la órbita elíptica del asteroide y se mueve en el intervalo [0,1]. Siendo 0 una órbita circular perfecta y 1 una órbita altamente elíptica.
* **ad (distancia afelio)** es la distancia más lejana al Sol a lo largo de la órbita del asteroide medida en ua.

“

per\_y fue descartando por ser el periodo pero medido en años

Por hipótesis y posibles leaks solamente nos quedamos con 13 atributos. **PASAR SLIDE**

### **Slide 5: Exploración de Datos (EDA)**

Título: Exploración de Datos

Contenido:

* Tratar de meter los histogramas y boxplots en una diapo
* Hace falta una diapo más para poner la matriz de correlación creo
* Insights iniciales

Guión:

Realizamos una exploración de datos inicial para entender mejor la distribución y las relaciones entre las diferentes características.

En la tabla de resumen se puede ver que el mínimo del semieje mayor es negativo lo cual no tiene sentido. También la excentricidad tiene un máximo mayor a 1 lo cual no puede ser ya que tal atributo se mueve entre 0 y 1.

Histogramas: En estos estos gráficos podemos ver que la mayoría de los atributos no están normalizados.

Matriz de correlación: En el gráfico podemos ver que a (semieje mayor) y ad (distancia afelio) tienen una alta correlación con per (periodo en días). Esto nos da un indicio de que podrían ser útiles para estimar el período orbital. **PASAR SLIDE**

”

### **Slide 6: Preprocesamiento de Datos**

Título: Preprocesamiento de Datos

Contenido:

* Pasos de preprocesamiento (eliminación de outliers, escalado, etc.)
* Razonamiento detrás de cada paso
* Poner la tablita de los NA’s y Boxplots

Guión:

"El preprocesamiento de datos es importante para mejorar el poder predictivo del modelo. En esta etapa eliminamos outliers, borramos los NA y estandarizamos los atributos . Una vez concluida esta fase nos quedamos con 120.435 instancias

### **Slide 7: Modelado y Evaluación**

Título: Modelado y Entrenamiento

Contenido:

* Primer Modelo Regresión lineal con regularización LassoCV

Guión:

"Comenzamos con un modelo simple, una regresión lineal con regularización Lasso con validación cruzada para elegir el mejor lambda. Obtuvimos un error cuadrático medio de 21, que es muy bajo en comparación a la mediana del periodo orbital. Además vemos que haciendo el gráfico de coeficientes en función del lambda, el atributo del semieje-mayor es el más importante como habíamos visto en la matriz de correlación." **PASAR SLIDE**

### **Slide 8: Análisis de Residuos 1er modelo**

Título: Modelado y Evaluación

Contenido:

* Gráfico de residuo

Guión:

"Para comprobar si este modelo iba a ser el mejor, comprobamos con el gráfico de residuos y notamos que existe una tendencia polinómica, esto quiere decir que hay sesgo en los residuos." **PASAR SLIDE**

### **Slide 9: Modelo Polinómico**

Título: Análisis de Coeficientes de Lasso

Contenido:

* MSE vs Grado

Guión:

"Probamos con un modelo de regresión polinómica con lasso y vimos que en el grado 2 era el que tenía menor error cuadrático medio." **PASAR SLIDE**

### **Slide 10: Residuos Modelo Polinómico**

Título: Análisis de Coeficientes de Lasso

Contenido:

* Gráfico de coeficientes en función de lambda
* Interpretación de los coeficientes

Guión:

"Hicimos el análisis de los residuos de los diferentes grados y siguen teniendo una tendencia polinómica. **PASAR SLIDE** Esto nos quiere decir que aún se puede mejorar el modelo"

### **Slide 12: Modelo con grados polinómicos fraccionarios**

Título: Análisis de Coeficientes de Lasso

Contenido:

* Gráfico de coeficientes en función de lambda
* Interpretación de los coeficientes

Guión:

"Intentamos probar con grados fraccionarios para ver si el modelo mejoraba su desempeño pero el error cuadrático medio empeora, entonces decidimos que el mejor modelo para predecir el periodo orbital es el de grado 2."**PASAR SLIDE**

### **Slide 13: Mejor coeficiente**

Título: Análisis de Coeficientes de Lasso

Contenido:

* Gráfico de coeficientes en función de lambda
* Interpretación de los coeficientes

Guión:

"Creamos un gráfico de los coeficientes en función de los lamda y observamos que el semieje mayor es el que tiene más importancia a la hora de predecir el target.

¿Existirá una relación causal entre el semieje mayor y el periodo orbital?"**PASAR SLIDE**

### **Slide 14: Kepler**

Título: Conclusiones

Contenido:

* Resumen de los hallazgos
* Importancia de los resultados

Guión:

"Investigando si existía una relación causal, encontramos que hay una ley que relaciona ambos atributos. La tercera ley de kepler.**PASAR SLIDE**

Creamos 2 columnas, 1 con el target al cuadrado y otra con el semieje mayor al cubo y vimos una correlación del 100%.**PASAR SLIDE**

Al evaluar nuestro modelo polinómico en grado 1.5 sin la parte lineal el error cuadrático medio es ínfimo y los residuos carecen de sesgo"**PASAR SLIDE**

### **Slide 17: Conclusión**

Título: Conclusión

Contenido:

* Conclusión

Guión:

"Desarrollamos un modelo de regresión polinómica de grado 2 que predice bien nuestro target,

Comprobamos que el semieje mayor era la variable más importante a la hora de querer predecir el periodo orbital.

Investigando, encontramos la ley de la naturaleza de nuestro problema y lo comprobamos en base a nuestros datos."**PASAR SLIDE**