

Experimentando herramientas sin código

Clasificación de razas de gatos mediante técnicas de aprendizaje automático en KNIME



Proyecto de Ciencia de Datos con KNIME



Autor: Alec Jonathan Montaña Romero

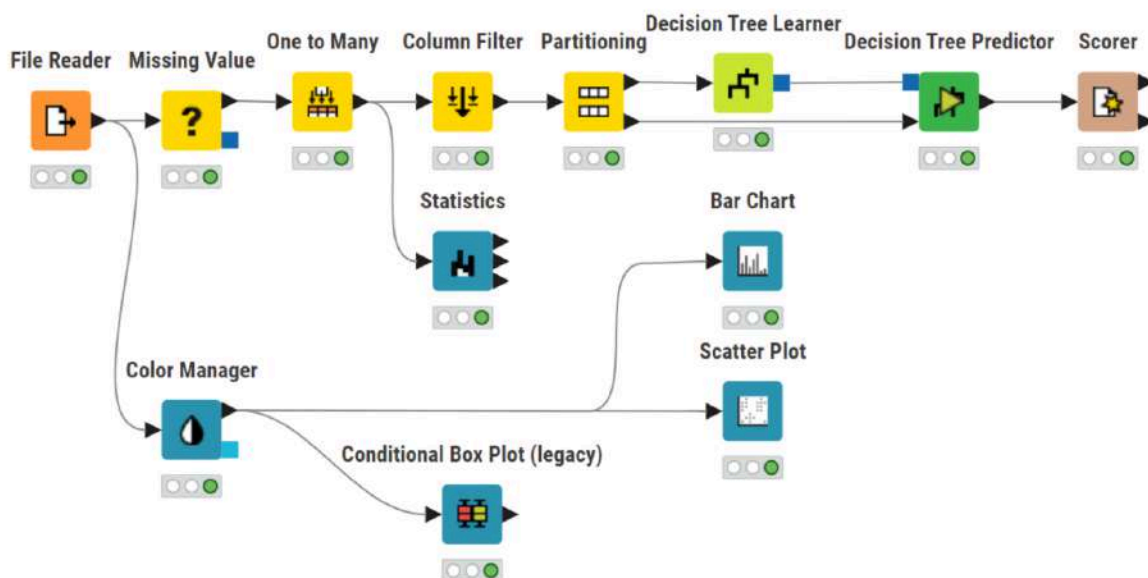
Fecha de entrega: 26/06/2025

Introducción y Objetivo

En el presente proyecto se hace uso de la herramienta no-code KNIME para abordar un problema de clasificación multiclase: identificar la raza de un gato a partir de sus características físicas y comportamentales. Se parte de un conjunto de datos preprocesado que incluye información como peso, longitud del cuerpo, color de ojos, entre otros.

Objetivo general: Desarrollar un modelo predictivo en KNIME capaz de clasificar la raza de un gato entre tres opciones (Angora, Maine Coon y Ragdoll), utilizando una serie de atributos relevantes presentes en el dataset.

Flujo del proceso



Descripción del Dataset

Fuente: Kaggle (<https://www.kaggle.com/datasets/joannanplkrk/its-raining-cats>)

Nombre del archivo: cat_breeds_clean.csv

Características principales del dataset:

- **Total de registros:** 1071
- **Columnas relevantes:**
 - Breed (objetivo)
 - Weight
 - Body_length
 - Eye_colour
 - Fur_colour_dominant
 - Country



#	Breed	Age_in_years	Age_in_months	Gender	Neutered_or_spayed	Body_length	Weight	Fur_colour_dominant	Fur_patterns	Eye_colour	Altered_neutered	Preferred_food	Shedding_season	Shedding_season	Country	Latitude	Longitude
1	Royal	0.05	3	female	FALSE	19	3	white	solid	blue	FALSE	wet	86	16	France	45.794	3.57
2	Royal	0.05	4	male	FALSE	19	2.5	white	solid	blue	FALSE	wet	42	16	France	45.812	3.878
3	Royal	0.05	6	male	FALSE	30	3.5	black	solid	green	FALSE	wet	41	11	France	44.858	3.576
4	Royal	0.05	6	female	FALSE	21	3	white	solid	blue	FALSE	wet	24	8	France	45.812	3.878
5	Royal	0.05	6	male	FALSE	21	3	red/orange	solid	green	FALSE	wet	51	10	France	45.805	3.848
6	Royal	0.05	6	female	FALSE	21	3	white	solid	orange	FALSE	wet	33	10	France	45.812	3.878

Preprocesamiento de Datos

Pasos realizados:

1. Carga del archivo CSV con el nodo **File Reader**
2. Transformación de variables categóricas a numéricas con **One to Many**
3. Relleno de valores faltantes con **Missing Value**
4. Eliminación de columnas no relevantes con **Column Filter**
5. Análisis de estadísticas básicas con **Statistics**

Estos pasos aseguran que los datos estén preparados correctamente para ser utilizados por los modelos de clasificación.



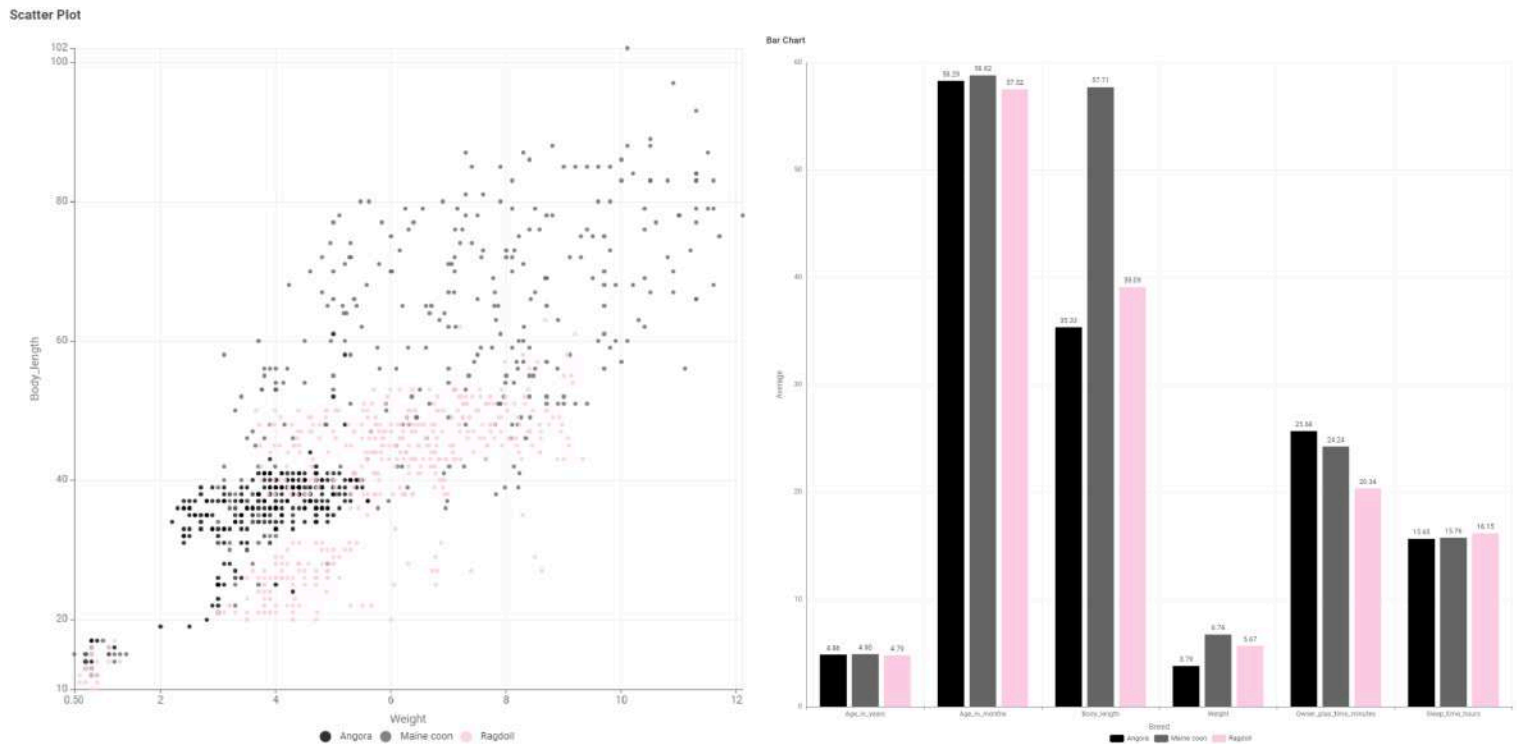
Análisis Exploratorio y Visualizaciones

Se realizaron visualizaciones para identificar patrones y distribuciones según la raza del gato.

Gráficos utilizados:

- **Scatter Plot:** Comparación entre peso y longitud del cuerpo, coloreado por raza.
- **Box Plot:** Distribución de peso por raza.
- **Bar Chart:** Visualización gráfica de la distribución promedio de características numéricas por raza.

Estas gráficas permiten observar la separación entre clases en función de atributos clave.

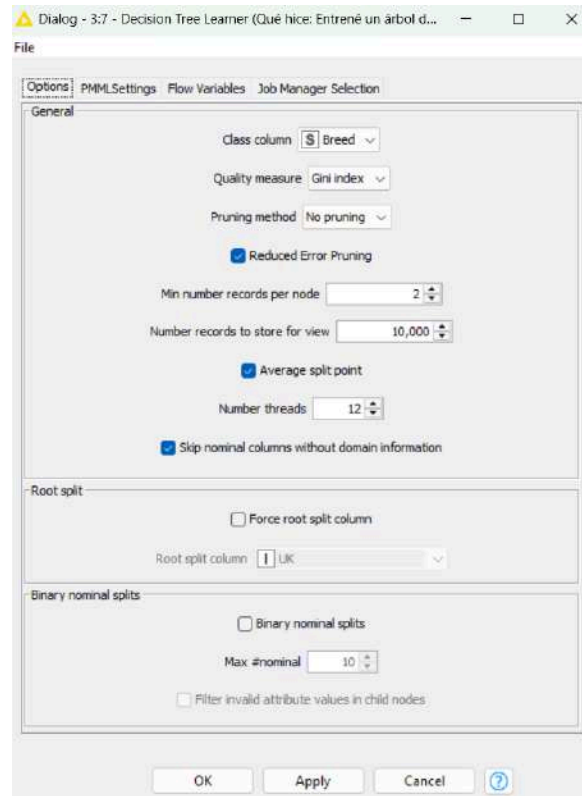


Modelado Predictivo

Se eligió un modelo de Árbol de Decisión por su interpretabilidad y facilidad de uso en KNIME.

Configuración del modelo:

- Nodo: **Decision Tree Learner**
- Partición: 70% para entrenamiento, 30% para prueba (**Partitioning** node)
- Target variable: **Breed**



Evaluación del Modelo

El rendimiento del modelo se evaluó utilizando el nodo **Scorer**, que generó las siguientes métricas:

Matriz de confusión:

- Angora: 86 correctamente clasificados.
- Maine Coon: 100 correctamente clasificados.
- Ragdoll: 131 correctamente clasificados.

► 1: Confusion matrix ► 2: Accuracy statistics ⚙ Flow Variables

Rows: 3 | Columns: 3

<input type="checkbox"/>	#	RowID	Angora Number (integer)	Maine coon Number (integer)	Ragdoll Number (integer)
<input type="checkbox"/>	1	Angora	86	2	0
<input type="checkbox"/>	2	Maine	2	100	1
<input type="checkbox"/>	3	Ragdol	0	0	131

Métricas de desempeño:

- **Precisión general (Accuracy):** 0.984
- **F-measure promedio:** 0.984
- **Cohen's Kappa:** 0.976
- **Recall mínimo por clase:** 0.971 (Maine Coon)



	Recall	Precision	F1 Score	Specificity	Accuracy	Cohen's Kappa
American Shorthair	0.971	0.984	0.977	0.984	0.984	0.976
Maine Coon	0.971	0.984	0.977	0.984	0.984	0.976
Russian Blue	0.971	0.984	0.977	0.984	0.984	0.976
British Shorthair	0.971	0.984	0.977	0.984	0.984	0.976

Conclusiones

El modelo desarrollado en KNIME mostró un excelente desempeño, con una alta precisión y métricas equilibradas para todas las clases. Las variables físicas como **Weight** y **Body_length** fueron claves para la diferenciación de razas.

La herramienta KNIME resultó ser efectiva para el desarrollo de proyectos de ciencia de datos sin necesidad de programar, facilitando tanto la preparación de datos como la construcción y evaluación de modelos.

Reflexiones Finales

- Este proyecto demuestra el potencial de las herramientas no-code como KNIME para resolver problemas reales de clasificación.
- La trazabilidad del flujo de trabajo y la facilidad de exploración visual hacen de KNIME una opción ideal para proyectos educativos y profesionales.
- El mismo enfoque puede extenderse a otros contextos: clasificación de razas de perros, tipos de plantas, etc.



Anexos y Recomendaciones

Recomendaciones futuras:

- Probar con modelos más complejos como Random Forest o Gradient Boosting
- Realizar validación cruzada para evitar overfitting.
- Incluir más razas o aumentar el número de muestras.

Anexos:

- Enlace al dataset original:
<https://www.kaggle.com/datasets/joannanplkrk/its-raining-cats>
- Enlace al proyecto:
<https://github.com/LynxPardelle/KNIMECatBreedsMachineLearning>

