

Naturaleza del proyecto

Una empresa agrícola líder se dedica al cultivo y comercialización de diversas variedades de arroz premium, incluyendo Cammeo y Osmancik. La calidad del producto se evalúa en base a atributos clave, y para asegurar la excelencia y eficiencia operativa, la empresa requiere un sistema de clasificación automática de granos de arroz de alta precisión.



Objetivo:

Implementar algoritmos de Machine Learning basado en el lenguaje de programación Python, con el fin de desarrollar modelos que utilice las características físicas de los granos de arroz (área, perímetro, longitud del eje mayor, longitud del eje menor, excentricidad, área convexa, medida, clase) para identificar si pertenecen a la variedad Cammeo o Osmancik. Este sistema mejorará la eficiencia en el proceso de clasificación, reducirá errores humanos y ayudará a mantener la calidad del producto.





Aspectos claves del negocio a resolver

- 1.- Mejorar el proceso operativo de la empresa
- 2.- Aumentar la precisión en la clasificación del grano de arroz
- 3.- Reducir el desperdicio
- 4.- Aumentar la calidad del producto
- 5.- Ahorro en costos y recursos
- 6.- Mejora en la toma de decisiones

Modelos empleados y análisis

Conjunto de Datos

- El conjunto de datos utilizado en este proyecto se encuentra disponible en el repositorio de [UCI Machine Learning](#). Posee características morfológicas de un total de 3810 granos de arroz de la variedad Cammeo y Osmancik.

	Area	Perimeter	Major_Axis_Length	Minor_Axis_Length	Eccentricity	Convex_Area	Extent	Class
0	15231.00	525.58	229.75	85.09	0.93	15617.00	0.57	b'Cammeo'
1	14656.00	494.31	206.02	91.73	0.90	15072.00	0.62	b'Cammeo'
2	14634.00	501.12	214.11	87.77	0.91	14954.00	0.69	b'Cammeo'
3	13176.00	458.34	193.34	87.45	0.89	13368.00	0.64	b'Cammeo'
4	14688.00	507.17	211.74	89.31	0.91	15262.00	0.65	b'Cammeo'
...
3805	11441.00	415.86	170.49	85.76	0.86	11628.00	0.68	b'Osmancik'
3806	11625.00	421.39	167.71	89.46	0.85	11904.00	0.69	b'Osmancik'
3807	12437.00	442.50	183.57	86.80	0.88	12645.00	0.63	b'Osmancik'
3808	9882.00	392.30	161.19	78.21	0.87	10097.00	0.66	b'Osmancik'
3809	11434.00	404.71	161.08	90.87	0.83	11591.00	0.80	b'Osmancik'

3810 rows × 8 columns

Modelos

Aprendizaje Supervisado

- Regresión logística
- SVM (Support Vector Machines)
- Random Forest
- K-NN
- Agrupación de Modelos (Ensembles)

Aprendizaje No supervisado

- PCA

Analysis

K-vecinos

	precision	recall	f1-score	support
0	0.90	0.84	0.87	430
1	0.87	0.92	0.90	523
accuracy			0.88	953
macro avg	0.89	0.88	0.88	953
weighted avg	0.89	0.88	0.88	953

Random Forest

	precision	recall	f1-score	support
0	0.93	0.89	0.91	501
1	0.92	0.94	0.93	642
accuracy			0.92	1143
macro avg	0.92	0.92	0.92	1143
weighted avg	0.92	0.92	0.92	1143

Analysis

Regresion Logistica

	precision	recall	f1-score	support
0	0.92	0.93	0.92	350
1	0.94	0.93	0.93	412
accuracy			0.93	762
macro avg	0.93	0.93	0.93	762
weighted avg	0.93	0.93	0.93	762

SVC

	precision	recall	f1-score	support
0	0.92	0.91	0.92	475
1	0.94	0.95	0.94	668
accuracy			0.93	1143
macro avg	0.93	0.93	0.93	1143
weighted avg	0.93	0.93	0.93	1143

Potencia final del Proyecto

El modelo de Regresión Logística utilizado, se posiciona como la mejor opción de clasificación, alcanzando una precisión del 93% y un ajuste superior al azar en la clasificación de granos de arroz.

El desarrollo de este trabajo permite llegar a la conclusión de que el uso de técnicas de ML en la industria arrocera, permite optimizar procesos, reducir el tiempo de clasificación y minimizar los costos que se general por esta actividad.