



Predicción de estado fenológico de la oliva

Komorebi AI

KOMOREBI AI Technologies, S.L., es una startup inspirada en la tecnología, fundada en 2019 y con sede en Madrid, se especializa en el desarrollo de soluciones avanzadas de Inteligencia Artificial (IA) y Ciencia de Datos. La misión de la empresa es integrar avances de vanguardia en ciencia de datos e I.A. para proporcionar a los clientes una ventaja competitiva, lograda a través del desarrollo de productos escalables y efectivos que transforman los negocios.

KOMOREBI facilita la adopción de tecnologías de IA para los clientes mediante la identificación de oportunidades de innovación, la evaluación de posibles casos de uso de IA y la creación de soluciones de IA a medida. La empresa se distingue por:

- **Capital Humano Excepcional:** Con un equipo altamente calificado con antecedentes científicos y sólidas capacidades analíticas, KOMOREBI atrae a los mejores talentos y ofrece soluciones integrales en diversos ámbitos de la IA.
- **Vigilancia Tecnológica Constante:** A través de una vigilancia tecnológica activa, KOMOREBI se mantiene a la vanguardia de la innovación, asegurando el desarrollo de soluciones de vanguardia y holísticas para los clientes.
- **Historial Demostrado:** Con una historia comprobada de proyectos exitosos de IA en varios sectores, KOMOREBI aporta experiencia para garantizar el éxito del proyecto y mitigar los riesgos asociados con la implementación de IA.
- **Asociación Colaborativa:** KOMOREBI enfatiza la estrecha colaboración con los clientes, manteniendo una comunicación abierta y fomentando relaciones laborales sólidas para asegurar el éxito del proyecto y el crecimiento mutuo.

Objetivo

Tradicionalmente, la identificación del momento óptimo de recolección se ha realizado de manera manual, calculando el índice de madurez de los frutos y el rendimiento graso, lo que requiere muestreos aleatorios en las parcelas y conlleva desventajas como desplazamientos frecuentes a las fincas, consumo de tiempo, imprecisión en los cálculos y planificación incierta de la recolección.

El principal objetivo del proyecto es poder predecir con antelación la evolución de las características fenológicas de la oliva, dado el estado fenológico actual y la meteorología. Esta predicción permitiría anticipar la decisión de en qué momento se podrá realizar la recolección de la aceituna para obtener un aceite de oliva de mayor calidad.

La fenología es el estudio de los ciclos de desarrollo de las plantas a lo largo del tiempo. La observación cuidadosa de estos estados es esencial para maximizar la calidad y el rendimiento de la cosecha de aceitunas. En el caso del **olivo**, esto incluye varios estados clave en su ciclo de vida:

Código	Estado fenológico
0101 Fenología_ A	Yema de Invierno
0102 Fenología_ B	Yema Movida
0103 Fenología_ C	Formación Inflorescencia
0104 Fenología_ D1	Empieza Corola
0105 Fenología_ D2	Despliegue Corola
0106 Fenología_ D3	Corola Cambio Color
0107 Fenología_ F	Floración
0108 Fenología_ G1	Caída de Pétalos
0109 Fenología_ G2	Fruto Cuajado
0110 Fenología_ H	Endurecimiento Hueso
0111 Fenología_ I1	Envero (Amarilleo)
0112 Fenología_ I2	Envero (Manchas rojas)
0113 Fenología_ J1	Fruto Maduro (Pulpa Blanca)
0114 Fenología_ J2	Fruto Maduro (Pulpa Roja)



Descripción de los datos

Muestreos fenológicos y parcelas

Para poder obtener la información de los estados fenológicos se cuenta con datos históricos de muestreos realizados por la [RAIF](#) de la Junta de Andalucía en distintas parcelas. Estos muestreos contienen datos entre 2006 y 2021.

El fichero Parquet **muestreos_parcelas.parquet** es el resultado de un preprocesamiento de los datos en crudo de los muestreos mencionados y contiene las siguientes columnas:

1. **generated_muestreos**: Fecha y hora de generación de los muestreos.
2. **codparcela**: Código de la parcela.
3. **provincia**: Provincia donde se encuentra la plantación (string).
4. **municipio**: Municipio donde se encuentra la plantación (string).
5. **fecha**: Fecha del muestreo.
6. **campaña**: Año de la campaña.
7. **poligono**: Código de polígono.
8. **parcela**: Número de parcela.
9. **recinto**: Número de recinto.
10. **subrecinto**: Número de subrecinto.
11. **estado_fenologico_1 - estado_fenologico_14**: Variables que representan el estado fenológico de la oliva en diferentes momentos del ciclo de crecimiento
12. **porcentaje_floracion**: Porcentaje de floración.
13. **generated_parcelas**: Fecha y hora de generación de las parcelas.
14. **102_coordenada_x_(utm)**: Coordenada X UTM.
15. **103_coordenada_y_(utm)**: Coordenada Y UTM.
16. **104_altitud_(m)**: Altitud en metros (float32).
17. **105_comarca**: Comarca de la plantación.
18. **201_superf_cultivada_en_la_parcela_agrícola_(ha)**: Superficie cultivada en la parcela agrícola en hectáreas (float32).
19. **202_pendiente_(%)**: Pendiente del terreno en porcentaje (float32).
20. **203_orientación**: Orientación del terreno.
21. **204_textura_del_suelo**: Textura del suelo.
22. **206_secano_/regadío**: Indicador de si la plantación es de secano o regadío.
23. **211_utilización_de_cubierta_vegetal**: Utilización de cubierta vegetal.
24. **212_tipo_de_cubierta_vegetal**: Tipo de cubierta vegetal.

25. **214_cultivo_asociado/otro_aprovechamiento**: Cultivo asociado u otro aprovechamiento.
26. **301_marco_(m_x_m)**: Espaciado entre plantas en metros cuadrados.
27. **302_densidad_(plantas/ha)**: Densidad de plantas por hectárea (float32).
28. **303_nº_pies_por_árbol**: Número de pies por árbol (float32).
29. **304_formación**: Formación de los árboles.
30. **305_diámetro_de_copa_(m)**: Diámetro de la copa en metros (float32).
31. **308_variedad_principal**: Variedad principal de oliva.
32. **311_fecha_de_plantación_variedad_principal**: Fecha de plantación de la variedad principal.
33. **208_riego:_procedencia_del_agua**: Procedencia del agua para el riego.
34. **209_riego:_calidad_del_agua**: Calidad del agua para el riego.
35. **313_variedad_secundaria**: Variedad secundaria de oliva.
36. **107_zona_homogénea**: Zona homogénea.
37. **120_zona_biológica_raif**: Zona biológica RAIF.
38. **401_estación_climática_asociada**: Estación climática asociada.
39. **402_sensor_climático_asociado**: Sensor climático asociado.
40. **207_riego:_sistema_usual_de_riego**: Sistema usual de riego.
41. **108_u_h_c_a_la_que_pertenece**: Unidad de Homogeneización Climática y Agronómica a la que pertenece.
42. **316_fecha_de_plantación_variedad_secundaria**: Fecha de plantación de la variedad secundaria.
43. **315_patrón_variedad_secundaria**: Patrón de la variedad secundaria.
44. **317_%_superficie_ocupada_variedad_secundaria**: Porcentaje de superficie ocupada por la variedad secundaria (float32).
45. **306_altura_de_copa_(m)**: Altura de la copa en metros (float64).
46. **310_patrón_variedad_principal**: Patrón de la variedad principal.
47. **411_representa_a_la_u_h_c_(si/no)**: Indica si representa a la Unidad de Homogeneización Climática.
48. **109_sistema_para_el_cumplimiento_gestión_integrada**: Sistema para el cumplimiento de la gestión integrada.

Si bien es cierto que para la modelización de los estados fenológicos con la meteorología no harían falta la mayoría de las columnas del .parquet, os hacemos llegar el fichero completo para que se puedan observar cómo pueden llegar a ser los datos en un contexto aplicado a un caso real de implementación de soluciones de I.A.

Meteorología

Para la obtención de datos de meteorología, una de las mejores fuentes se trata de Copernicus, que es el programa de la Unión Europea de observación y monitorización de la Tierra, que analiza el planeta y su medio ambiente en beneficio de los ciudadanos europeos. El programa proporciona datos de manera operacional y servicios de información de forma gratuita sobre numerosas áreas de aplicación.

Los servicios de datos de Copernicus proporcionan acceso a una amplia gama de datos y productos derivados instrumentos satelitales, incluyendo imágenes de alta resolución, datos de RADAR y LiDAR, datos de temperatura y humedad, cobertura terrestre y la calidad del aire, entre otros. Estos datos pueden ser utilizados por empresas, organizaciones gubernamentales y científicos para llevar a cabo investigaciones, tomar decisiones y desarrollar aplicaciones innovadoras.

Copernicus cuenta con múltiples servicios, pero los dos más importantes para este proyecto son:

[Copernicus Land Monitoring Service](#), este servicio contiene [múltiples productos](#), de los cuales destacan los siguientes para este proyecto :

- Surface Soil Moisture (SSM): Frecuencia (aprox) diaria, cobertura del globo terráqueo.
- Land Surface Temperature (LTS): Frecuencia (aprox) de 10 días, cobertura del globo terráqueo.
- Fraction of Photosynthetically Active Radiation Absorbed by the Vegetation (FAPAR): Frecuencia (aprox) 10 días, cobertura del globo terráqueo.

Y [Copernicus Browser](#) para los los siguientes índices:

- Normalized Difference Vegetation Index (NDVI): Frecuencia (aprox) 5 días, coordenadas Jaen, Andalucía.
- Normalized Difference Water Index (NDWI): Frecuencia (aprox) 5 días, coordenadas Jaen, Andalucía.
- Structure Insensitive Pigment Index (SIPI): Frecuencia (aprox) 5 días, coordenadas Jaen, Andalucía.

- Soil Adjusted Vegetation Index (SAVI): Frecuencia (aprox) 5 días, coordenadas Jaen, Andalucía.
- Green Normalized Difference Vegetation Index (GNDVI): Frecuencia (aprox) 5 días, coordenadas Jaen, Andalucía.

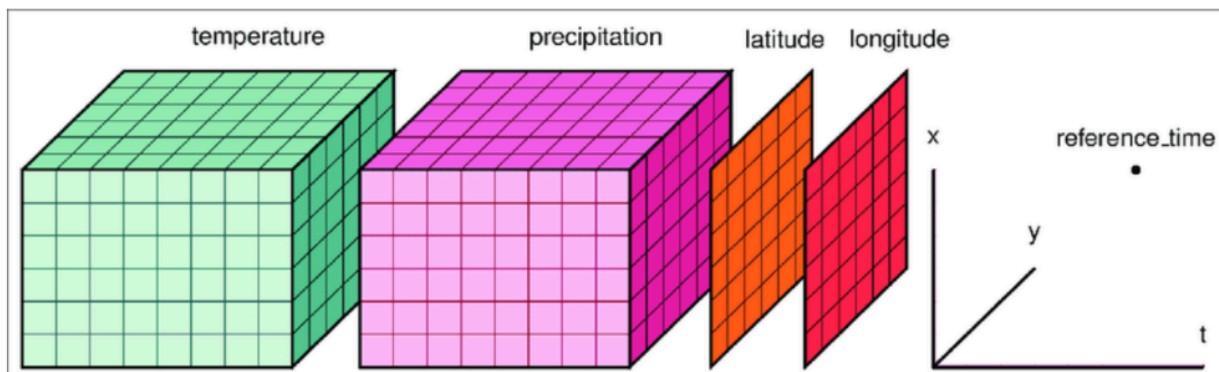
Se os compartirá un fichero .parquet con los ficheros de cada uno de los índices citados ya procesados, lo que una vez fueron ficheros en formato network Common Data Form (NetCDF) han sido procesados para vosotros en una base de datos. Este último tendrá una estructura de: FECHA / CODPARCELA / LAT/ LON / INDICE / VALOR

Calculo del número de filas aprox, suponiendo que todos los índices tuvieran una frecuencia de 5 días:

- 72 (6 valores al mes) * 6 años (2018 a 2023) * 8 índices (NDVI repetido) * 4848 parcelas $\approx 17M$. Este cálculo es aproximado debido a que hay algún índice que es diario o con frecuencia 10 días.

Los archivos NetCDF son un formato de archivo binario utilizado para almacenar datos científicos y ambientales, como información climática, oceanográfica y atmosférica, entre otros. Los archivos NetCDF son diseñados para ser independientes de la plataforma y pueden ser leídos por una amplia variedad de programas de software científico, como MATLAB, R, Python y otros.

Pueden contener información en forma de variables, dimensiones y atributos. Las variables representan los datos almacenados en el archivo, mientras que las dimensiones se refieren a las dimensiones de la matriz que contiene los datos. Los atributos contienen información adicional sobre las variables, como unidades, rangos o unidades de tiempo.



La principal desventaja de los ficheros NetCDF es que posteriormente se requiere un proceso para acceder a las coordenadas de interés. En Python esto se puede realizar con la librería [xarray](#).

Por último, también existen otras formas de acceder directamente a los datos para una serie de coordenadas. Un ejemplo en Python se puede ver en el siguiente [notebook](#).