

Escuela de negocios digital Kschool



EVOLUCIÓN AGRÍCOLA EN ESTADOS UNIDOS



MÁSTER EN CIENCIA DE DATOS

Trabajo Final de Máster

Autor: Alejandro Muñoz Molina

LinkedIn: <https://www.linkedin.com/in/alex245/>

ÍNDICE

INTRODUCCIÓN	1
OBJETIVOS.....	1
MATERIAL Y METODOS	2
WEB SCRAPING y APY	2
DATA CLEANING	4
DATA ANALYSIS	4
MODELING	4
FRONT END	5
RESULTADOS.....	5
Análisis univariante	5
Análisis multivariante.....	8
MODELOS PREDICTIVOS	14
CONCLUSIÓN.....	15
BIBLIOGRAFÍA	16

INTRODUCCIÓN

Estados Unidos es uno de los principales productores agrícolas a nivel global especialmente en productos como la soja, maíz y trigo. Esto es posible gracias a las vastas extensiones de superficies agrícolas que pueden llegar a encontrarse y también a la diversidad de climas que están presentes en este país.

A lo largo de la historia, esta superficie agrícola, se ha ido modificando. Actualmente el crecimiento de superficie agrícola no es una opción viable pero sí que lo es la modificación y adaptación de los cultivos a los nuevos factores a los que el sector agrario se enfrenta a día de hoy (*O'Brien et al., 2021*).

Algunos de estos factores pueden ser:

El cambio climático, el cual afecta directamente a la agricultura (*Weeler et al., 2013*).

El cambio en el rango de temperaturas y precipitaciones están generando modificaciones como por ejemplo reubicaciones del cultivo del maíz en el Oeste Medio de Estados Unidos (*Green et al., 2018; Lark et al., 2015; Wright et al., 2017*). También, los periodos prolongados de sequías están haciendo que los acuíferos y embalses reduzcan sus niveles de agua a lo largo de los años ya que son incapaces de volver a recuperar el agua extraída (*Barnett & Pierce, 2009; Russo & Lall, 2017; Zektser et al., 2005*).

Otro input en este cambio es la introducción de nuevas tecnologías en la agricultura, sistemas de información Geográfica que ayudan a la hora del uso de la maquinaria haciéndola más precisa en el trazado que realiza, interpretación de imágenes ya sean por satélite o drones para conocer el estado del cultivo, o sensores que ayudan a conocer las necesidades de los cultivos en tiempo real gracias a la transmisión continua de información (*Basso & Antle, 2020; Liu, 2020; Peters et al., 2020*).

También, los factores políticos han tenido una gran influencia en el desarrollo de este sector y en la evolución de la agricultura, aplicación y creación de nuevas políticas agrícolas, diversas crisis que han ido ocurriendo a lo largo de la historia han tenido un peso relevante en este aspecto (*Busch et al., 2019*).

Todos estos factores han hecho que la agricultura trate de adaptarse, ya sea modificando el tipo de cultivo de la región o cambiando a nuevas variedades de cultivos más resistentes.

OBJETIVOS

1. Este trabajo tiene como objetivo el estudio de los cambios y modificaciones que se han ido produciendo en las superficies agrícolas, concretamente ver la evolución de los principales cultivos de la región de Estados Unidos desde 1950 hasta 2021 y que efectos han tenido los elementos climáticos en este tránsito.
2. Evolución en los próximos años de los principales cultivos de Estados Unidos.

MATERIAL Y METODOS

Con el fin de poder desarrollar este proyecto y llegar a los objetivos descritos en el apartado anterior, este caso se diferenci3 principalmente en cinco puntos, los cuales se describen en los siguientes puntos.

Tabla 1: Notebooks y archivos destinados a cada apartado de estudio

Apartados	Notebooks
Web Scraping y API	1_WebScraping.ipynb Carpeta Rstudio
Data Cleaning	2_Transformaci3n_datos_Temp_y_PP.ipynb 3_Transformaci3n_datos_Sup_de_cultivo.ipynb
Data Analysis	4.1_An3lisis_PP.ipynb 4.2_An3lisis_T.ipynb 4.3_An3lisis_superficies.ipynb 4.4_An3lisis_multivariante.ipynb
Modeling	5_modelo.ipynb
Front end	6_streamlit.ipynb app.py

Todos estos apartados van dirigidos principalmente a la obtenci3n y an3lisis de las variables de temperatura, precipitaci3n y superficie, diferenciadas por estado y tambi3n por cultivo en el caso de la variable superficie, en un rango de a3os que va desde 1950 hasta 2021. Se ha seleccionado la variable superficie como variable objetivo (variable dependiente) y las variables de temperaturas y precipitaci3n como independientes con el fin de ver si existe alg3n tipo de influencia sobre la variable objetivo.

WEB SCRAPING y APY

Web Scraping:

Mediante esta t3cnica se obtuvieron todos los datos de precipitaciones y temperaturas procedentes la p3gina web NOAA (National Centers for Enviromental Information): <https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series>.

Mediante c3digo Python se estableci3 una estructura de datos con el que se pod3an descargar los dataframes necesarios, uno por cada estado con el rango de fechas seleccionado. Posteriormente, una vez obtenidos todos los datos necesarios se unir3an para obtener un 3nico dataframe de temperatura y otro de precipitaciones con todos los estados meses y a3os.

Las variables que aparecen en estos dos dataframes son:

Tabla 2: Variables Temperatura y Precipitaci3n

Variable	Descripci3n
State	Referencia al estado de EEUU
Date	A3o
Months	Meses
Precipitation	Datos de precipitaci3n en "mm" como unidad de medida
Temperature	Datos de temperatura en "eC" como unidad de medida

Todo este proceso junto con el c3digo que se ha creado para ello est3 descrito en el notebook:

- 1_WebScraping

API:

Para este caso la API que se utilizó proviene de la página web de USDA (National Agricultural Statistics Service) a la cual se puede acceder a través de este enlace: <https://www.robertdinterman.com/usdarnass/articles/usdarnass.html>.

Esta API está construida para ser utilizada con código "R" por lo que este fue el tipo de fuente que se utilizó para la descarga de los datos. En este caso la petición a la API fallaba constantemente y solo nos permitía hacer una petición con un valor total de 50.000 datos cada una, por lo que no se pudo automatizar todo lo que se pretendía. Para no sobrepasar el límite por petición lo que se hizo fue hacer una descarga de superficie por cada año (1950 - 2021) en la que aparecían todos los estados y cultivos. En algunos casos aun así el límite de valores por petición se sobrepasó, en estos casos se dividió por año y también por estado.

Finalmente, lo que se obtuvo tras unir todos los dataframes descargados fue una tabla con las siguientes variables:

Tabla 3: Descripción de las variables API

Variable	Descripción
source_desc	"Programa": Incluye el Censo de Agricultura y el programa de encuestas encuestas nacionales, estatales y del condado.
sector_desc	"Sector": Variable que se divide en cinco categorías: "ANIMALES Y PRODUCTOS", "CULTIVOS", "DEMOGRAFÍA", "ECONOMÍA" o "AMBIENTAL"
group_desc	"Grupo": subconjuntos dentro del sector (p. ej., en sector_desc = "CULTIVOS", los grupos son "CULTIVOS DE CAMPO", "FRUTAS Y NUECES DE ÁRBOL", "HORTICULTURA" y "HORTALIZAS").
commodity_desc	"Producto básico": el tema principal de interés (p. ej., "MAÍZ", "GANADO", "MANO DE OBRA", "TRACTORES", "OPERADORES").
short_desc	"Elemento de datos": complemento de las columnas: commodities_desc, class_desc, prodn_practice_desc, util_practice_desc, statisticcat_desc y unit_desc.
domain_desc	"Dominio": Característica de las operaciones que producen un bien en particular (p. ej., "CLASE ECONÓMICA", "ÁREA OPERADA", "CLASIFICACIÓN NAICS", "VENTAS").
domaincat_desc	"Categoría de dominio": categorías o particiones dentro de un dominio (p. ej., en domain_desc = "SALES", las categorías de dominio incluyen \$1,000 A \$9,999, \$10,000 A \$19,999, etc.).
agg_level_desc	"Nivel Geográfico": Nivel de agregación o granularidad geográfica de los datos. ("DISTRITO AGRÍCOLA", "CONDADO", "INTERNACIONAL", "NACIONAL", "REGIÓN: MULTIESTATAL", "REGIÓN: SUBESTADO", "ESTADO", "CUENCA" o "CÓDIGO POSTAL")
statisticcat_desc	"Categoría": el aspecto de un producto que se mide (p. ej., "ÁREA COSECHADA", "PRECIO RECIBIDO", "INVENTARIO", "VENTAS").
state_name	"Estado" - Nombre completo del estado.
asd_desc	"Distrito agrícola": nombre del distrito de estadísticas agrícolas.
county_name	"Condado" - Nombre del condado.
region_desc	"Región": entidades geográficas definidas por NASS que no se definen fácilmente por otros niveles geográficos estándar.
zip_5	"Código postal": código postal de EE. UU.

watershed_desc	“Cuenca” - Nombre asignado al HUC.
year	“Año”
reference_period_desc	“Período”: el marco de tiempo específico, dentro de un freq_desc.

DATA CLEANING

Una vez descargado los datos el siguiente paso fue hacer un estudio con el fin de obtener los datos de interés para un posterior análisis de estos. Para ello se crearon dos notebooks:

- 2_Transformación_datos_Temp_y_PP
- 3_Transformación-datos_Sup_de_cultivo

En el primer caso (2_Transformación_datos_Temp_y_PP) las dos acciones principales que se realizaron fueron un concatenado de todos los dataframes de temperaturas divididos por estados en uno solo, esta misma acción se hizo también con los dataframes de precipitaciones. Seguidamente se modificaron las posiciones de los valores para facilitar posteriormente el análisis de las variables.

En el segundo caso (3_Transformación-datos_Sup_de_cultivo) se realizó, como en el caso anterior, una unificación de todos los dataframes procedentes de la descarga a través de la API. Otros pasos que se realizaron fue el filtrado y eliminación de datos que no eran necesarios para este estudio. Todo ello que explicado de manera detallada en notebook.

DATA ANALYSIS

Los notebooks creados para este apartado han sido:

- 4.1_Análisis_PP
- 4.2_Análisis_T
- 4.3_Análisis_superficies
- 4.4_Análisis_multivariante

Todos ellos fueron destinados al análisis las variables de estudio con la finalidad de mostrar y entender la evolución de las variables dentro del rango de años analizados y determinar si existe alguna correlación entre ellas. Para ello se realizó un primer análisis univariante seguido de un análisis conjunto de todas las variables de estudio.

Todos estos resultados se muestran en el apartado 4. RESULTADOS de este documento.

MODELING

En este caso el nombre del notebook donde aparece todo el proceso de modelos es:

- 5_modelo

El objetivo de este notebook fue intentar realizar un análisis predictivo sobre la superficie de los cultivos. Primero se estudió el peso de las variables independientes (temperatura y precipitación) sobre la variable dependiente (superficie de cultivos) para ello se utilizaron los algoritmos que se utilizan en los modelos de árboles de decisiones y bosques aleatorios. Posteriormente se aplicaron varios modelos y se analizaron cuáles fueron los más acertados en la predicción. Los modelos utilizados han sido:

- Regresión lineal
- K Neares Neighbour Regressor
- Random forest
- XGBoost

FRONT END

El notebook donde se encuentra todo el contenido de este apartado es el siguiente:

- 6_streamlit

La librería que se utilizó fue streamlit, a partir del cual se hizo un resumen esquematizado y visual de todos los resultados obtenidos en el análisis de las variables.

Para este apartado del proyecto, con el fin de poder visualizar la página web se debe acceder a la terminal y ubicarnos en el archivo NOTEBOOKS (../AgriculturalEvolution/NOTEBOOKS), que es donde aparece el archivo “app.py” con el contenido de la web. Una vez en este archivo ejecutamos el comando “streamlit run app.py” para visualizar nuestra página web.

Aquí se deja un enlace a un pequeño video donde se visualiza el inicio de la aplicación:

<https://youtu.be/GftS6Tyrsv4>

RESULTADOS

En este apartado se pretende exponer los principales resultados obtenidos tanto en el análisis univariante como multivariante:

En los puntos 4.1. Análisis univariante y 4.2. Análisis multivariante existen numerosas gráficas interactivas las cuales provienen de los análisis realizados en los “notebook”, en esta memoria se presentan únicamente como imagen. Estas se pueden visualizar también en los “notebook” relativos a cada uno de los apartados.

Para todos los casos, los resultados que se muestran en esta memoria son un resumen de los principales análisis realizados. Si se quisiese completar la información, en los notebooks de análisis de datos aparecen todos los estudios realizados.

Los principales análisis que se han hecho para las variables han sido análisis de tendencias, distribución y correlaciones.

Análisis univariante

Variable Precipitación:

Estudiando la variable de precipitación se observa que su tendencia desde la fecha de inicio (1950) hasta el último año de estudio (2021) se ha ido incrementando (figura 1).

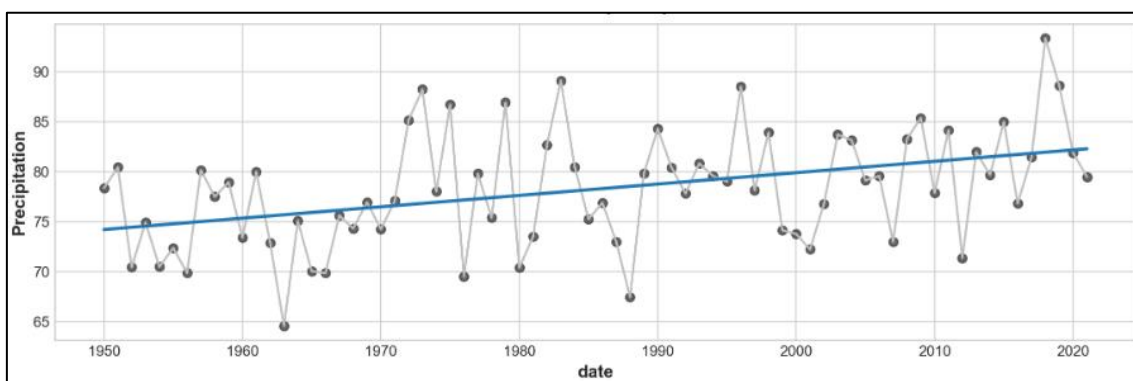


Figura 1: Tendencia de la precipitación media desde 1950 hasta 2021. (Fuente: 4.1_Análisis_PP)

Las épocas de mayor precipitación se dan normalmente en los meses de verano, mientras que el mes con menor precipitación de manera generalizada es febrero (figura 2).

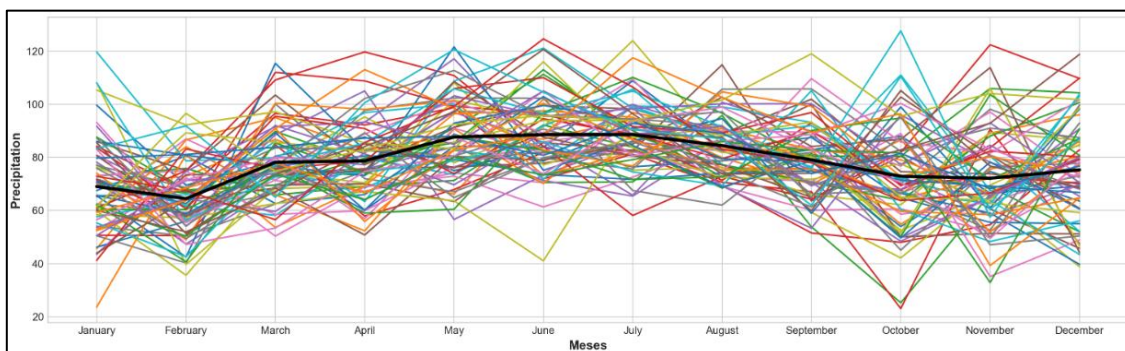


Figura 2: Evolución media de la precipitación mensual (mm). (Fuente: 4.1_Análisis_PP)

Finalmente, en la figura 3 se observa la distribución que tienen las precipitaciones, concentrándose en rangos de entre 50 y 100 mm.

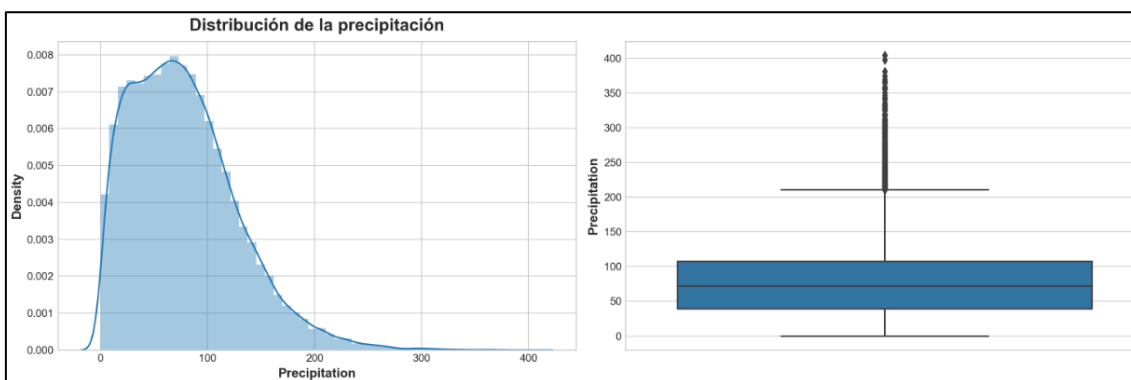


Figura 3: Distribución de la precipitación. (Fuente: 4.1_Análisis_PP)

Variable Temperatura:

Igual que en la variable anterior, la temperatura ha tenido un incremento a lo largo de los años (figura 4).

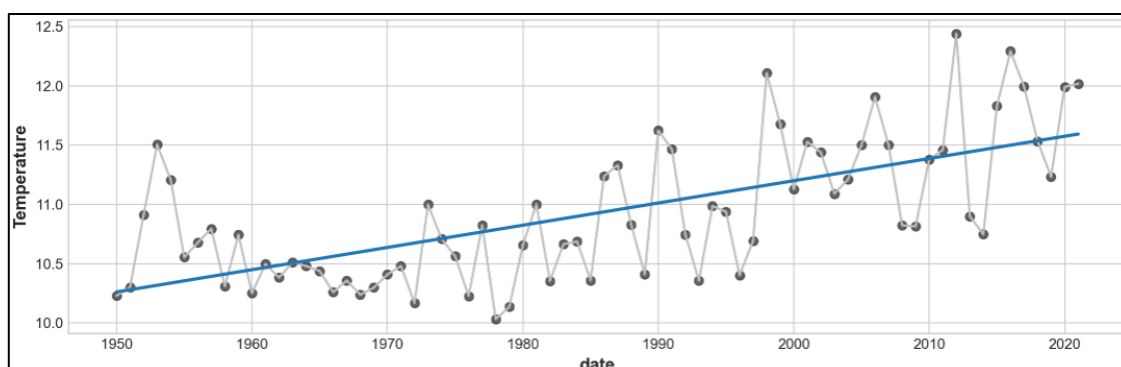


Figura 4: Evolución media de la temperatura anual (°C). (Fuente: 4.2_Análisis_T)

Como era de esperar, la media con el mes más caluroso, reflejados en la figura 5, es Julio, mientras que los meses más fríos aparecen en diciembre y enero con una temperatura media aproximada de 0°C.

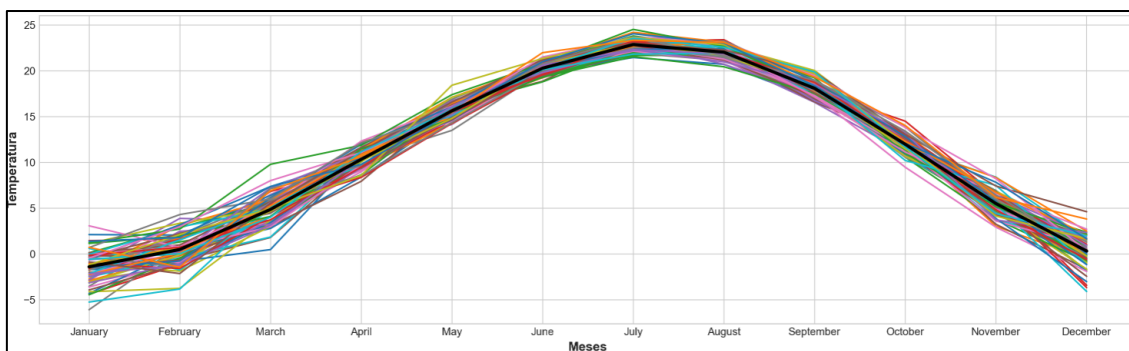


Figura 51: Evolución media de la temperatura mensual (°C) .(Fuente: 4.2_Análisis_T)

La figura 6 muestra la distribución que tienen las temperaturas, concentrándose en rangos de entre 3 y 20 °C.

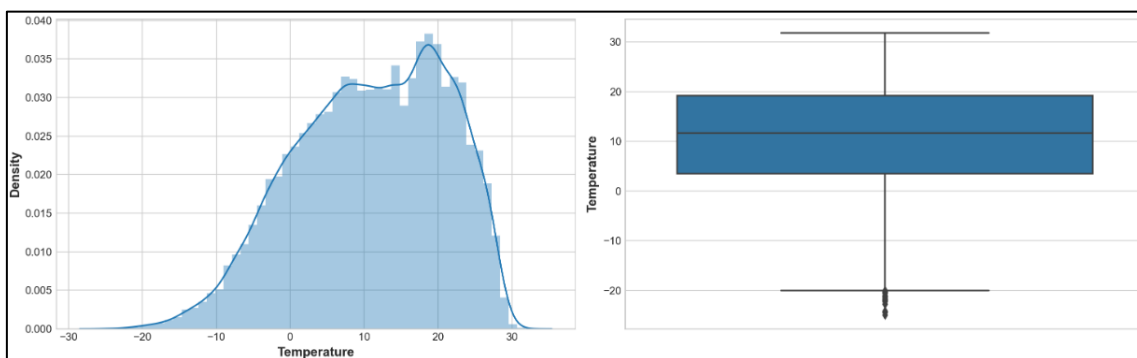


Figura 6: Distribución de la temperatura. (Fuente: 4.2_Análisis_T)

Variable Superficie:

La figura 7 muestra la evolución de la superficie agrícola a lo largo de los años en Estados Unidos, se diferencian dos grandes épocas, desde 1950 hasta 1990 las cuales están marcadas por una bajada y subida de esta superficie y otra desde 1990 hasta la actualidad la cual esta superficie se ha mantenido estable hasta hoy en día.

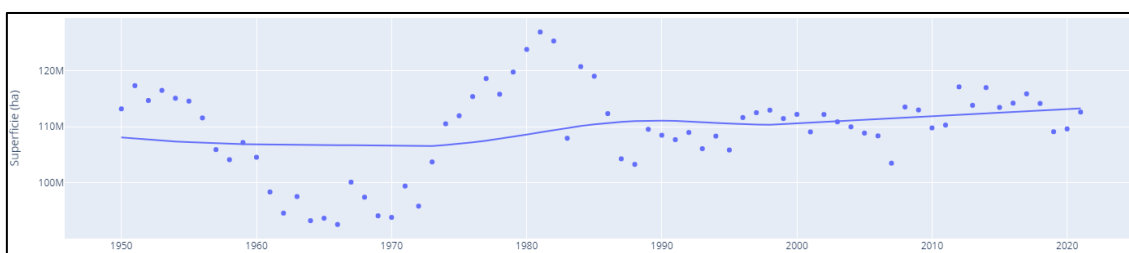


Figura 7: Evolución de la superficie (ha) media para cada año de estudio. (Fuente: 4.3_Análisis_superficies)

En la figura 8 se observa cuáles son los estados con mayor superficie agraria. En 2021 Iowa, Illinois, Kansas y el Norte de Dakota fueron los que mayor superficie de cultivo tuvieron con un porcentaje del 25% y un total de 27898490 hectáreas.

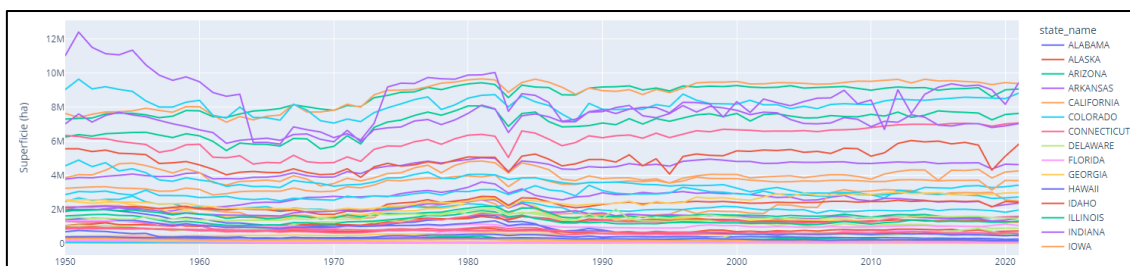


Figura 8: Superficie (ha) por cada estado de EEUU. (Fuente: 4.3_Análisis_superficies)

Haciendo referencia a la superficie de cada tipo de cultivo (figura 9) se observa que claramente, lo largo de las décadas, tres de ellos destacan sobre el resto, en este caso se hace referencia al maíz, la soja y trigo, de los cuales este último ha ido decayendo en las últimas décadas, mientras que la soja y el maíz siguen incrementando la superficie de cultivo. En 2021 estos tres cultivos representaron el 82 % de superficie cultivada respecto al total en EEUU.

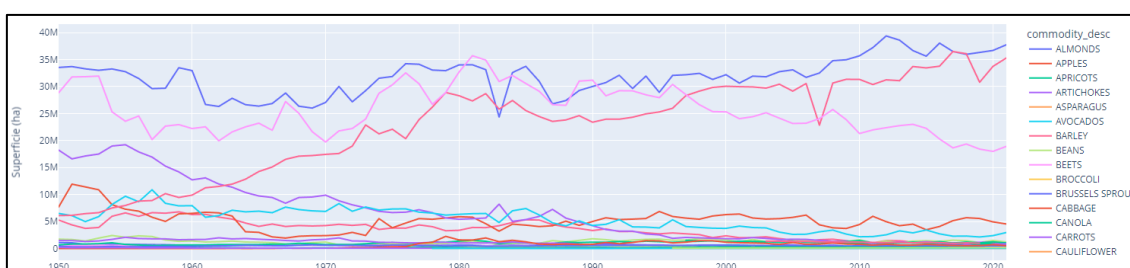


Figura 9: Superficie (ha) de cada cultivo de estudio. (Fuente: 4.3_Análisis_superficies)

Otro análisis realizado (figura 10) es la distribución de los datos. La mayoría de las hectáreas para cada cultivo por estado y año se sitúan entre 0 y 100.000 hectáreas. En cuanto a los años se ve como a partir del año 2000 el número de datos aumenta en comparación a los años anteriores.

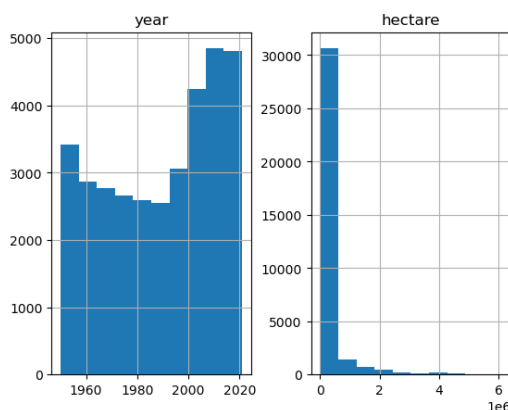


Figura 10: Distribución de los datos para cada año y superficie (ha) de los cultivos. (Fuente: 4.3_Análisis_superficies)

Análisis multivariante

Estudio de las variables agrupadas

Para este caso, se hizo, primero, un análisis multivariante entre las variables independientes (temperatura y precipitación) con el fin de conocer la relación que había entre ellas y posteriormente se introdujeron los datos la variable superficie.

Una vez incluida la variable dependiente, primero se realizó un estudio de correlaciones de la media total de todos los estados, superficies y meses para cada uno de los años y segundo, un análisis para cada uno de los cultivos con mayor superficie.

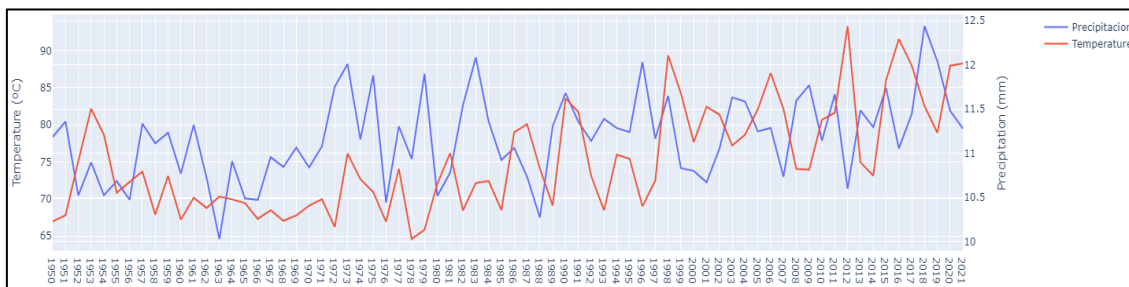


Figura 11: Evolución de las variables temperatura (°C) y precipitación (mm). (Fuente: 4.4_Análisis_multivariante)

La figura 11 muestra la evolución de las variables independientes a lo largo de los años y son comparadas en la figura 12 a través de una correlación móvil comparando cada año con sus cinco años anteriores. De este modo se pueden ver picos considerablemente altos de correlaciones que oscilan entre 0.9 y -0.9, tomando una correlación negativa en los últimos 20 años indicando que cuando una variable disminuye la otra aumenta.

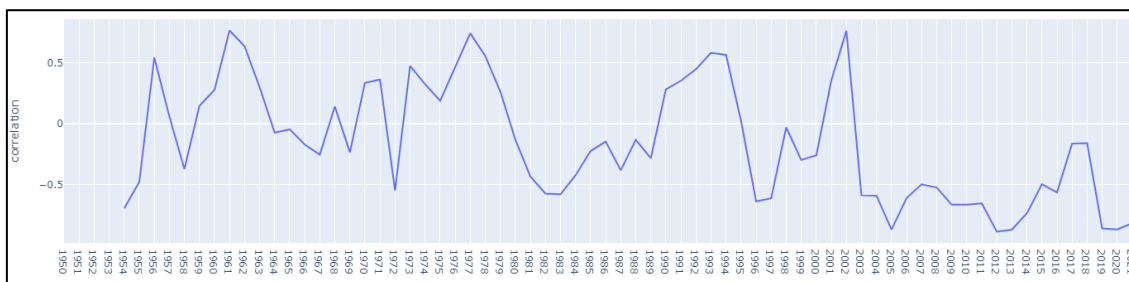


Figura 12: Correlación móvil (5 años) entre las variables independientes (temperatura y precipitación). (Fuente: 4.4_Análisis_multivariante)

Otro tipo de correlación que se estudió fue la correlación lineal (Pearson correlation) y la correlación no lineal (MIC), en ningún caso estas correlaciones dieron una correlación significativa de las variables independientes sobre la variable dependiente.

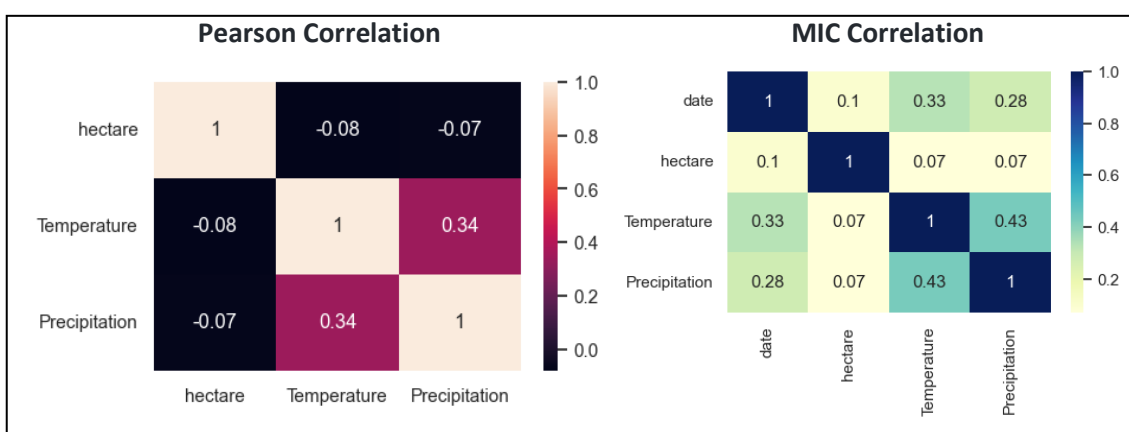


Figura 13: Correlación lineal (Pearson) y no lineal (MIC). (Fuente: 4.4_Análisis_multivariante)

Siguiendo con el estudio de correlaciones se intentó ver si había alguna correlación (en este caso correlación lineal) desfasada en el tiempo entre las variables independientes y las variables dependientes. En todos los casos la mayor correlación que salía era en el momento cero (figura 14).

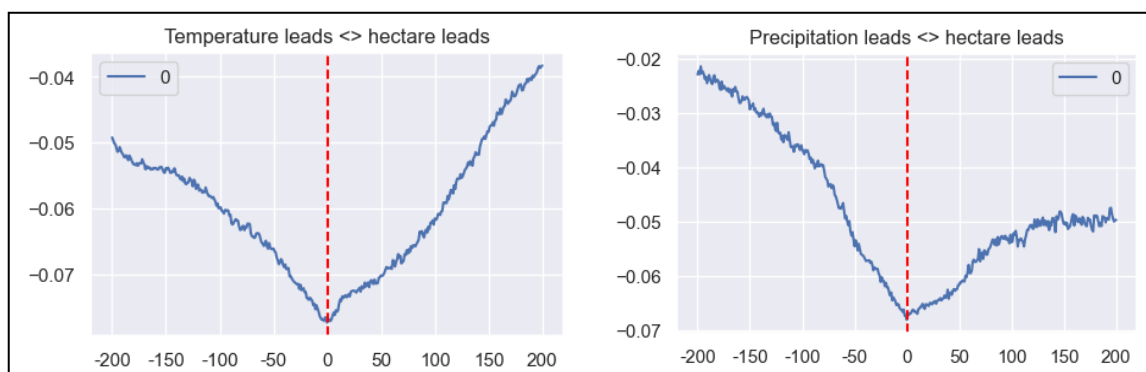


Figura 14: Correlación lagueada entre las variables independiente y la variable dependiente. (Fuente: 4.4_Análisis_multivariante)

Igual que en la figura 12, las figuras 15 y 16 representan correlaciones móviles, pero en este caso correlaciones móviles entre la superficie y las variables temperatura y precipitación respectivamente. Igual que en la figura 12 la correlación móvil en estos casos oscila entre -0.9 y 0.9, aproximadamente.

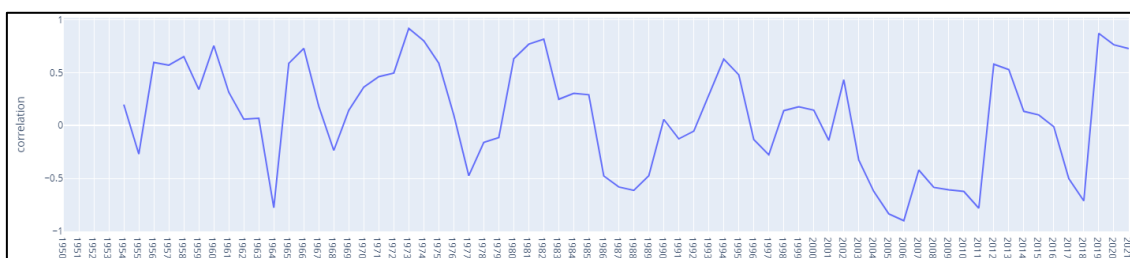


Figura 15: Correlación móvil (5 años) variable superficie (ha) vs temperatura (°C). (Fuente: 4.4_Análisis_multivariante)

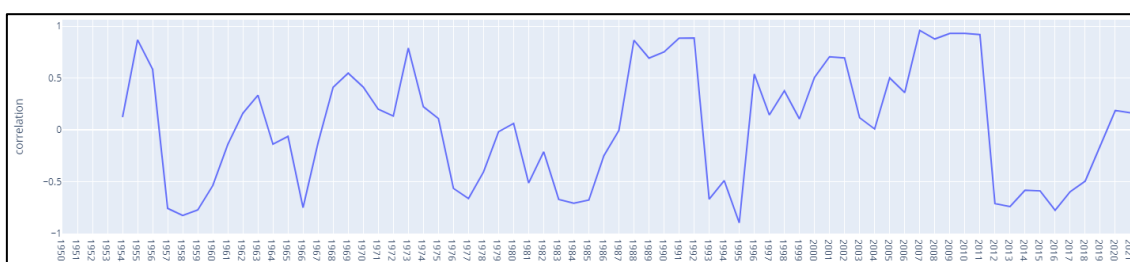


Figura 16: Correlación móvil (5 años) variable superficie (ha) vs precipitación (mm). (Fuente: 4.4_Análisis_multivariante)

Hasta ahora, los análisis que se han mostrado han sido una agrupación mediante el cálculo de la media de todos los estados y meses para las variables independientes (temperatura y precipitación) y en el caso de la superficie una media de todos los cultivos.

Estudio de los cultivos con mayor superficie

Los análisis que se muestran a continuación hacen referencia a los principales cultivos de EEUU: trigo, maíz, soja, algodón, sorgo, cebada y avena; los cuales representan el 93 % del total sobre la superficie cultivada en este país.

Para este caso de estudio se creó un dataframe donde aparecen reflejados los valores temperatura precipitación y superficie que percibe cada cultivo de estudio.

Del mismo modo que en el caso anterior, se ha estudiado la correlación de los principales cultivos en función de la temperatura, precipitación y superficie (figura 17). Se observa que existe una correlación, en general, alta entre valores de los propios parámetros. Comparando los parámetros de superficie con el resto de las variables (precipitación y temperatura) se ve que la temperatura sí que tiene, al menos en algunos cultivos, una correlación que nos pueda indicar que existe influencia entre las variables ya sea influencia positiva o negativa, por el contrario, al comparar la correlación entre superficie y precipitación en general la mayoría de los valores son cercanos a cero.

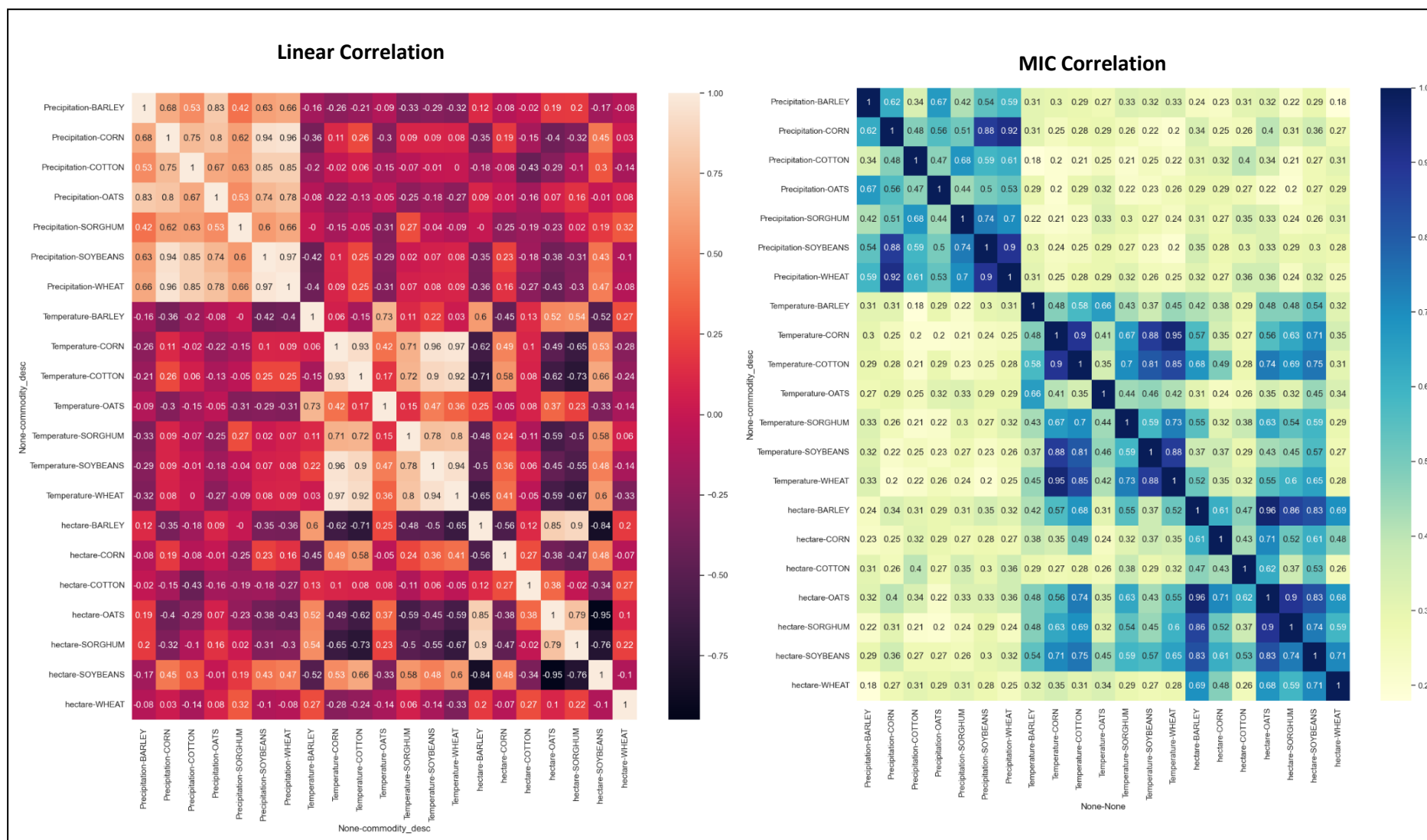


Figura 17: Correlación lineal (Pearson) y no lineal (MIC). (Fuente: 4.4_Análisis_multivariante)

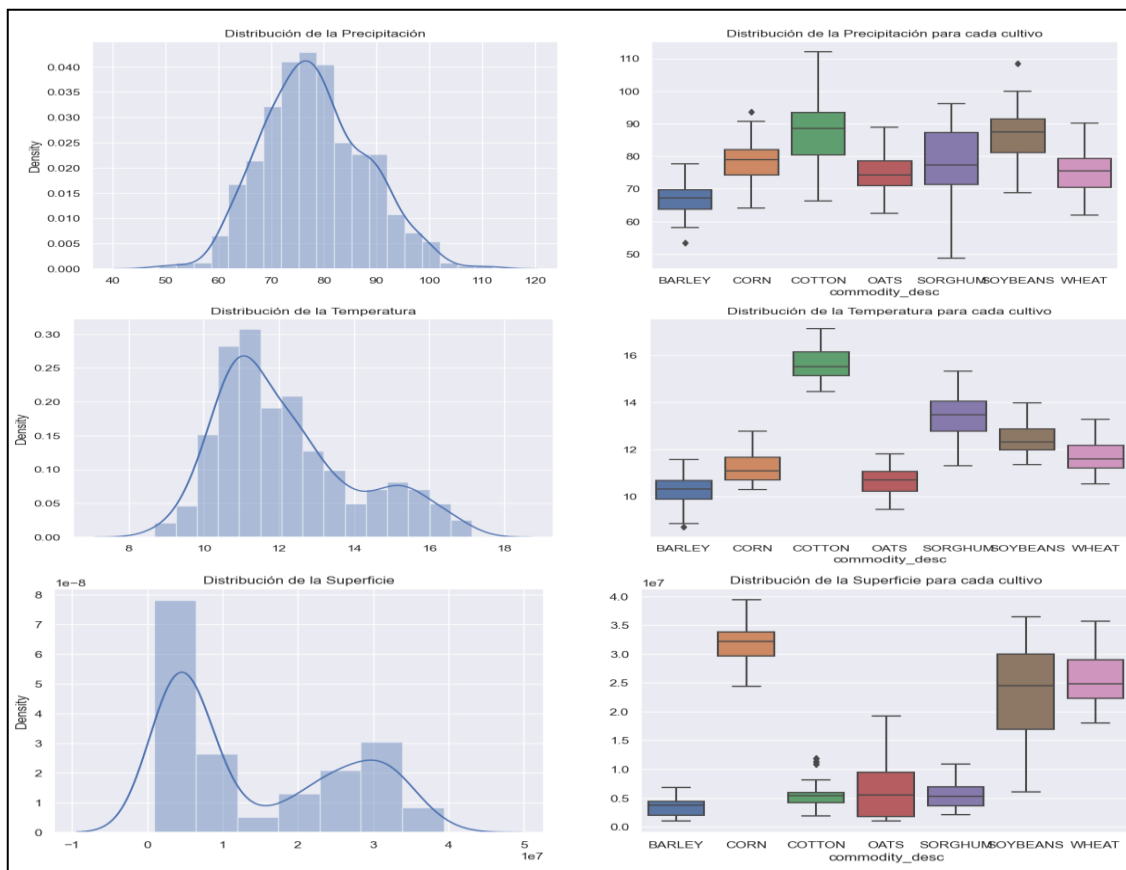


Figura 18: Distribución de las variables de estudio para los principales cultivos. (Fuente: 4.4_Análisis_multivariante)

En la figura 18 se muestra que la variable de precipitación tiene una distribución gaussiana, mientras que los histogramas de superficie y temperatura tiene una distribución no uniforme. En cuanto al tipo de distribución que tienen los cultivos, el cultivo del algodón es el que mayor temperatura recibe en comparación al resto de cultivos. También la precipitación media es de las más elevadas junto con el cultivo de la soja, y como ya se había visto anteriormente el cultivo del maíz es el que mayor superficie tiene.

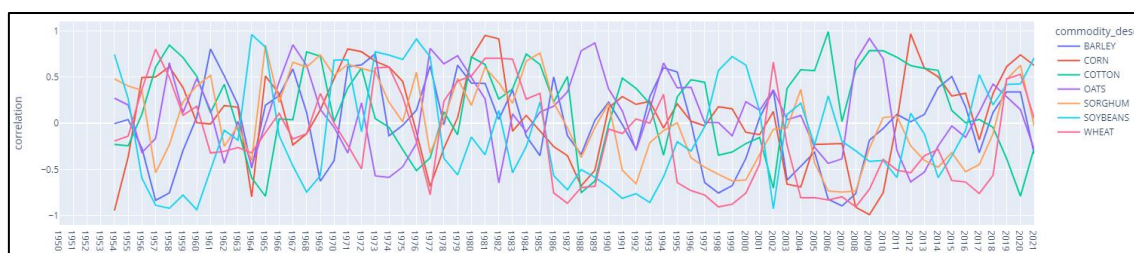


Figura 19: Correlación móvil (5 años) superficie (ha) de los principales cultivos) vs temperatura (°C). (Fuente: 4.4_Análisis_multivariante)

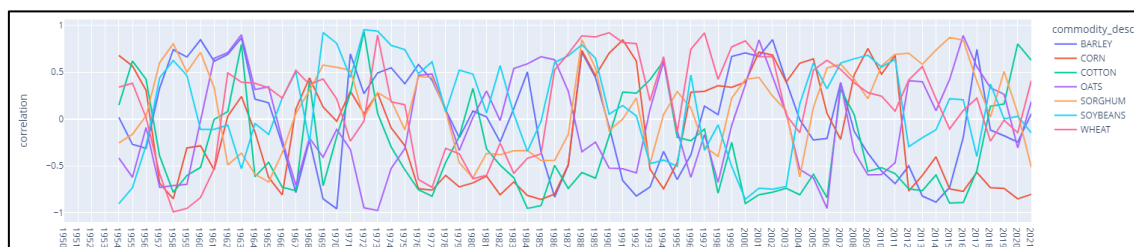


Figura 20: Correlación móvil (5 años) superficie (ha) de los principales cultivos vs precipitación (mm). (Fuente: 4.4_Análisis_multivariante)

La figura 19 y 20 muestran las correlaciones lineales móviles en rangos de cinco años para cada año. La figura 19 muestra la correlación móvil entre la superficie de los cultivos y la variable temperatura y la figura 20 muestra esta misma correlación, pero para la precipitación.

Como se puede ver, existe una gran variación de correlaciones a lo largo de los años.

MODELOS PREDICTIVOS

El objetivo de los modelos era el de intentar predecir de la manera más acertada posible y evitando el sobreajuste, la evolución de la superficie de los principales cultivos de estudio en EEUU. Para ello, primero se estudiaron mediante un estudio de árboles de decisiones, concretamente “*Decisión Tree*” y “*Random Forest*”, que parámetros podrían ser los más influyentes en la predicción, en la mayoría de los casos apareció la temperatura como parámetro más influyente que la precipitación (figura 21).

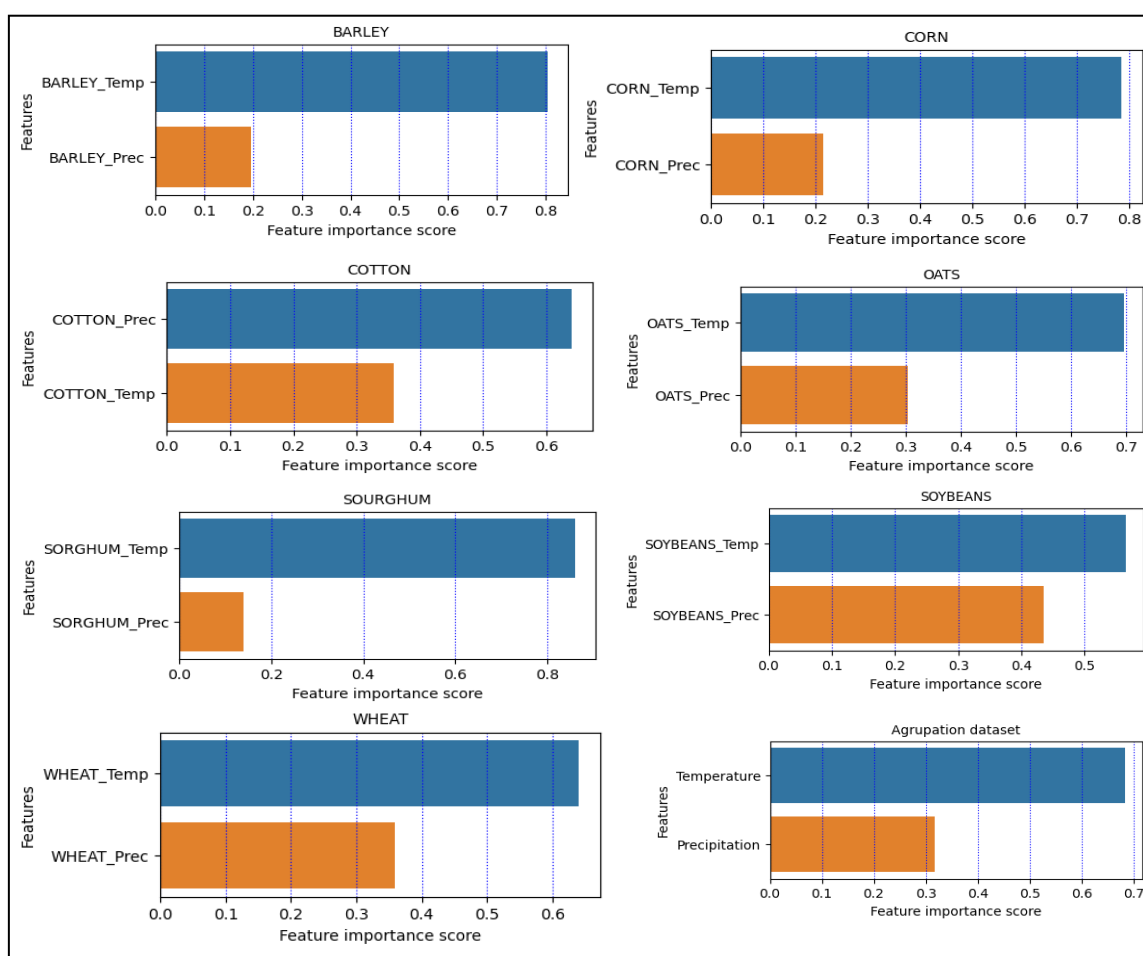


Figura 21: Influencia de las variables independientes sobre la variable dependiente en cada cultivo de estudio y para la agrupación de todos los cultivos. (Fuente: 4.4_Análisis_multivariante)

Para los modelos predictivos se emplearon varios modelos: regresión lineal (Lineal Regression), modelo basado en agrupaciones (K Nearest Neighbour Regressor) y dos modelos de árboles de decisiones (Random Forest y XGBoost).

Para los modelos de árboles de decisión (Random Forest y XGBoost) se realizó un estudio para ver los mejores parámetros que podrían funcionar con nuestros datos, los modelos que mejores

resultados dieron, comparando las distintas metricas (MAE, MSE, R2) fueron el modelo de regresión lineal y el modelo de Random Forest. En la tabla 4 se muestran los resultados.

Tabla 4: Resultados MAE, MSE, R2, de los modelos de estudio. (Fuente: 4_modelo)

Model Name	Measures	BARLEY	CORN	COTTON	OATS	SORHGHUM	OYBEANS	WEATH	GLOBAL
Regresión Lineal	MAE	2027386,17	4403033,84	1364205,27	6667478,72	3034744,53	7165692,18	5633283,66	2840585,39
Regresión Lineal	MSE	1423,86	2098,34	1167,99	2582,15	1742,05	2676,88	2373,45	1685,40
Regresión Lineal	R^2	-71,08	-4,31	-3,11	-1004,40	-64,53	-5,72	-6,58	-0,29
K Nearest Neighbour Regressor	MAE	2500925,76	4457144,16	1585540,64	7427319,07	3613711,08	7079767,99	6023267,74	5947172,99
K Nearest Neighbour Regressor	MSE	1581,43	2111,19	1259,18	2725,31	1900,98	2660,78	2454,23	2438,68
K Nearest Neighbour Regressor	R^2	-106,91	-4,76	-4,35	-1468,72	-94,33	-5,97	-8,55	-3,66
Random Forest	MAE	1869400,84	4611726,79	1231186,13	6645011,97	2767642,83	9020431,38	5409996,68	3659391,00
Random Forest	MSE	1367,26	2147,49	1109,59	2577,79	1663,62	3003,40	2325,94	1912,95
Random Forest	R^2	-63,45	-5,04	-2,94	-1082,72	-53,06	-9,78	-5,86	-1,01
XGBoost	MAE	1736751,39	4642398,55	1341184,63	6657613,15	2761327,69	9645011,32	4700554,18	5379760,98
XGBoost	MSE	1317,86	2154,62	1158,10	2580,24	1661,72	3105,64	2168,08	2319,43
XGBoost	R^2	-60,61	-5,38	-4,61	-1182,69	-52,39	-12,13	-4,53	-3,24

CONCLUSIÓN

Como conclusiones de este trabajo se podría decir que la superficie de los cultivos agrícolas en Estados Unidos se ha mantenido estable a lo largo de las décadas, pero sí que se han encontrado diferencias entre las superficies de algunos cultivos, por ejemplo: mientras que los cultivos de soja y maíz han ido ganando superficie agrícola, el trigo ha ido disminuyendo. En los diferentes estados la superficie agrícola se ha mantenido estable a lo largo de las últimas décadas.

Sobre las variables independientes estudiadas, sí se ha visto una clara correlación en alguno de los cultivos con la temperatura, mientras que la precipitación no ha tenido la misma influencia, esto pueda deberse a que existen técnicas de riego lo suficientemente desarrolladas que pueden sustituir las lluvias y cubrir las necesidades de la planta eficientemente.

Por otro lado, los modelos predictivos han dado unos resultados muy poco acertados por lo que estos modelos no servirían para hacer una buena predicción, se necesitarían una mayor cantidad de datos o variables influyentes sobre la superficie que pudiesen hacer mejorar los modelos predictivos.

Algunos de los próximos pasos que se podrían realizar para mejorar este estudio podrían ser:

- Aumentar el número de posibles variables de interés, ya que la variable de superficie puede estar mucho más influenciadas por otro tipo de variables, (temperatura y precipitación máxima y mínima, rendimientos de los cultivos, humedades relativas, exportaciones e importaciones, evolución de la población, precios de productos, etc.)
- Nuevos análisis, como la evolución de los principales cultivos a lo largo de los años en cada uno de los estados.

BIBLIOGRAFÍA

BARNETT, Tim P.; PIERCE, David W. Sustainable water deliveries from the Colorado River in a changing climate. *Proceedings of the National Academy of Sciences*, 2009, vol. 106, no 18, p. 7334-7338.

BASSO, Bruno; ANTLE, John. Digital agriculture to design sustainable agricultural systems. *Nature Sustainability*, 2020, vol. 3, no 4, p. 254-256.

BUSCH, Lawrence M.; LACY, William B. *Science, agriculture, and the politics of research*. Routledge, 2019.

CABRERA, Abraham Aparicio. Historia Económica Mundial 1950–1990. *Economía informa*, 2014, vol. 385, p. 70-83.

GREEN, Timothy R., et al. Where is the USA Corn Belt, and how is it changing?. *Science of the Total Environment*, 2018, vol. 618, p. 1613-1618.

LARK, Tyler J.; SALMON, J. Meghan; GIBBS, Holly K. Cropland expansion outpaces agricultural and biofuel policies in the United States. *Environmental Research Letters*, 2015, vol. 10, no 4, p. 044003.

LIU, Simon Y. Artificial intelligence (AI) in agriculture. *IT Professional*, 2020, vol. 22, no 3, p. 14-15.

O'BRIEN, Peter; KRAL-O'BRIEN, Katherine; HATFIELD, Jerry L. Agronomic approach to understanding climate change and food security. *Agronomy Journal*, 2021, vol. 113, no 6, p. 4616-4626.

PETERS, Debra PC, et al. Harnessing AI to transform agriculture and inform agricultural research. *IT Professional*, 2020, vol. 22, no 3, p. 16-21.

RUSSO, T. A.; LALL, U. Depletion and response of deep groundwater to climate-induced pumping variability. *Nat Geosci* 10: 105–108. 2017.

WHEELER, Tim; VON BRAUN, Joachim. Climate change impacts on global food security. *Science*, 2013, vol. 341, no 6145, p. 508-513.

Wright, Christopher K., et al. "Recent grassland losses are concentrated around US ethanol refineries." *Environmental Research Letters* 12.4 (2017): 044001.

ZEKTSER, S.; LOÁICIGA, Hugo A.; WOLF, J. T. Environmental impacts of groundwater overdraft: selected case studies in the southwestern United States. *Environmental Geology*, 2005, vol. 47, no 3, p. 396-404.

PAGINAS WEB:

<https://cnvrg.io/random-forest-regression/>

<https://scikit-learn.org/stable/>

<https://stackoverflow.com/>

<https://www.kaggle.com/>

<https://plotly.com/>

https://pandas.pydata.org/pandas-docs/stable/reference/general_functions.html

<https://matplotlib.org/>

<https://seaborn.pydata.org/>

<https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series>

<https://www.robertdinterman.com/usdarnass/articles/usdarnass>

<https://quickstats.nass.usda.gov/#60CB39DB-E74C-3619-8851-E0EAD520AACA>

<https://plotly.com/python/choropleth-maps/>