



DataLab Community

Social Data Challenge 4.0



Modelo predictivo de sequías en la república mexicana.

Integrantes

Bernal Temores, Maximiliano
Cortés de la Torre, Alan Jesús
Hernández Mota, Daniel
López Zúñiga, Ana Sofía
Silva Ocegueda, Guillermo León

Jalisco, Octubre 2021

Índice

Descripción del problema	4
Contexto	6
Medición de Sequías	6
Modelos de Predicción	7
Modelo Estadístico	7
Propuesta de solución	9
Hipótesis	10
Objetivos	11
General	11
Particulares	11
Metodología	11
Bases de datos	14
Justificación de variables	15
Variable de respuesta	16
Sequía en México	16
Variables Independientes	17
Variables metereológicas	18
Temperatura	18
Índice ultravioleta	19
Cubrimiento de nubes	19
Precipitación	19
Presión atmosférica	19
Humedad	19
Punto de rocío	19
Variables geológicas	19
Hidrografía	20
Cuencas	20
Subcuencas	20
Índice topográfico de humedad	21
Hipsometría	21
Curvatura de perfil	21
Curvatura del plano	21
Curvatura del terreno	22
Pendiente	22
Hidrogeología	23
Edafología	23
Experimentos	24

Resultados	24
Conclusiones	28
Trabajo por hacer	28
Referencias	29
Anexos	32
Anexo 1	32
Selección de municipios	32
Anexo 2	33
Calidad de los datos de la variable de respuesta	33
Anexo 3	35
Variables de consumo para predecir la sequía el siguiente mes.	35
Anexo 4	50
Agrupación de score promedio respecto a la variable de respuesta por entidad federativa	51
Anexo 5	59
Explicabilidad de las primeras 10 variables más importantes del modelo simple mediante la técnica SHAP	59
Explicabilidad de las primeras 10 variables más importantes del modelo complejo mediante la técnica SHAP	64

Descripción del problema

Con más del **85% del territorio nacional** en estado de sequía, México está experimentando su segunda sequía registrada más fuerte en su historia (Arenas, 2021). La intensidad de esta crisis solo se encuentra debajo de la sequía del 2011, la cual causó alrededor de **8 mil millones de pesos de pérdida solo en el sector agropecuario** (Lobato-Sánchez, 2016).

Una sequía se define como la disminución de precipitaciones en una región respecto a su índice anual promedio. Esta escasez de agua afecta mayormente al sector agrícola, dificultando la producción de cultivos por la falta de lluvias. Sin embargo, otras consecuencias significantes incluyen costos incrementales en el sector salud, desestabilización económica generadas por la falta de agua, y el aumento de incendios forestales (Centro para el Control y Prevención de Enfermedades [CDC], 2019).

El 52% del territorio nacional mexicano está clasificado como región árida o semiárida, lo cual conlleva a mayor propensidad a sequías. En esta región se concentra el 77% de la población nacional y se produce el 85% del producto interno bruto (PIB) (Esparza, 2013). Esta concentración poblacional y económica genera una gran vulnerabilidad de seguridad nacional ante las sequías. El gobierno federal históricamente ha enfrentado las sequías sólo de manera reactiva, actuando solamente hasta que la escasez de agua se presentaba en la región. Fue hasta el 2013 cuando el gobierno estableció el Programa Nacional Contra la Sequía (PRONACOSE), el cual funciona como el principal desarrollador de metodologías a nivel nacional para la respuesta o la preparación para las sequías (CONAGUA, 2018). Estas metodologías buscan garantizar la disponibilidad de agua, minimizar los efectos de la sequía en caudales ecológicos, y minimizar el impacto económico de la sequía.

El éxito de este y otros programas similares alrededor del mundo depende en gran parte de qué tanto tiempo de anticipación se tenga para enfrentar a las sequías. Por lo tanto, ONGs y gobiernos de otros países han desarrollado distintos modelos para predecir el suceso e intensidad de estas crisis. El Banco Mundial ha desarrollado un modelo global que determina la probabilidad de que una región sufra ciertas intensidades de sequía (World Bank, 2021). Estados Unidos, por otra parte, cuenta con modelos de la Oficina Nacional de Administración Oceánica y Atmosférica (NOAA) que predicen con mayor certeza y localidad la ocurrencia, continuación, o disipación de una sequía (NOAA, 2021).

Actualmente, México carece de una herramienta con alcance local como NOAA que pueda ser fácilmente utilizada por organizaciones y trabajadores para desarrollar y mejorar las estrategias de mitigación a las sequías tales como los de la PRONACOSE. Por ende, el objetivo de este proyecto es crear un modelo estadístico supervisado de aprendizaje automático que prediga, con un mes de anticipación, la sucesión e intensidad de una sequía a nivel municipal en la República Mexicana. Esperamos que con este modelo, los tomadores de decisiones mexicanos puedan usar las predicciones locales y certeras de sucesión de sequías

para afinar sus estrategias de adaptación, reduciendo los impactos sociales y económicos de las sequías en México.

Contexto

Medición de Sequías

Para medir la sucesión e intensidad de una sequía, se han utilizado históricamente diversos índices que conjuntan distintas variables climatológicas y geobiológicas. Cada uno de estos índices tiene ventajas y desventajas dependiendo de las variables que decida incluir en su categorización. Por ejemplo, el *Índice estándar de Precipitación* (SPI) utiliza datos de precipitación para caracterizar simplificadaamente la humedad y disponibilidad de agua en una región. La facilidad de obtención y cálculo han hecho al SPI uno de los índices más utilizados globalmente, lo cual favorece comparaciones inter-regionales. Sin embargo, este no cuenta con adecuaciones a la evapotranspiración o escurrimiento de agua en la región, haciéndolo un índice más débil. Otro índice popular es el *Índice de Sequía de Palmer*, el cual utiliza datos de temperatura y nivel de precipitación para cuantificar en una escala de -10 a +10 la intensidad, sucesión, o ausencia de una sequía, donde -10 es la condición más seca posible (Centro Nacional de Investigación Atmosférica [NCAR], 2021). Este índice es efectivo en determinar la intensidad de sequía a largo plazo por su uso de datos precedentes de temperatura y precipitación. Sin embargo, justo como al SPI, le es difícil medir sequías de plazos de menos de 12 meses por no considerar factores como la evapotranspiración de las plantas o el almacenamiento de agua en nieve o hielo. Además, el *Índice de Palmer* no es utilizado a gran escala como el SPI, lo cual debilita comparaciones inter-regionales (WMO).

Para sobrepasar las carencias de información y uso global, la comunidad científica en Estados Unidos, Canadá, y México, han utilizado el índice de *United States Drought Monitor* (USDM). Este índice incluye los dos índices anteriores más el análisis de otras seis variables climáticas y geobiológicas:

- Índice de Severidad de Palmer
- Índice Estándar de Precipitación
- Humedad de Tierra USDA a ocho pulgadas
- Humedad de Tierra USDA a 20 pulgadas
- Unidades USGS de Escorrentía Hidrológica
- Equivalente de Pico de Aguanieve NRCS
- Almacenamiento de Reserva NRCS
- Índice de Respuesta Vegetal a Sequía

El índice categoriza a las sequías en una escala de D0 a D4, donde D0 no presenta sequía¹ y D1-D4 son niveles incrementales de la intensidad de esta. La **figura 1** muestra los cuatro niveles de intensidad con sus respectivos rangos de variables meteorológicas y geobiológicas

¹ D0 no denota sequía pero es un indicador de que la zona está anormalmente seca. Por ende no se contempla como sequía.

que ayudan a que un experto haga la determinación de categoría de intensidad (Universidad de Nebraska Lincoln [UNL], 2021). Esta combinación de observaciones interdisciplinarias locales, el juicio de expertos, y su amplio uso en la región hacen que el índice USDM sea preferible a otros para nuestro modelo de predicción.

Categoría	Descripción (severidad)	Índice de sequía de Palmer	Índice estandarizado de precipitación	Humedad de la tierra	Escorrentía hidrológica	Equivalente de nieve-agua	Almacenamiento de depósito	Índice de respuesta de vegetación a las sequías
D0	Anormalmente seco	[-1, -1.9]	[-0.5, -0.7]	[21, 30]	[25, 75]	[50, 62.5]	{100%}	Estrés pre-sequía
D1	Sequía moderada	[-2, -2.9]	[-0.8, -1.2]	[11, 20]	[10, 24]	[37.5, 50]	[75, 100]	Sequía moderada
D2	Sequía Severa	[-3, -3.9]	[-1.3, -1.5]	[6, 10]	[10, 24]	[25, 37.5]	[50, 75]	Sequía Severa
D3	Sequía Extrema	[-4, -4.9]	[-1.6, -1.9]	[3, 5]	menor a 10	[12.5, 25]	[25, 50]	Sequía extrema
D4	Sequía Excepcional	-5 o menos	-2.0 o menos	[0, 2]	mínimo	[0, 12.5]	[0, 25]	Sequía extrema

Figura 1: Tabla de clasificación de sequías según el índice USDM. Los climatólogos utilizan los rangos dados por esta tabla para determinar la categoría de intensidad de la sequía de una región en específico (UNL, 2021).

Modelos de Predicción

Debido a los distintos orígenes y escalas de las sequías, así como las variables meteorológicas relacionadas con estas, existen diversos métodos para la predicción de la sucesión e intensidad de estos fenómenos naturales. Actualmente se utilizan mayormente tres métodos de predicción: estadístico, dinámico, e híbrido. El método estadístico utiliza las correlaciones entre variables climatológicas y los registros históricos de sequía para determinar la probabilidad de la sucesión de una sequía. El método dinámico usa los modelos de circulación general climatológica para recrear los procesos atmosféricos y oceánicos y de esta manera generar una predicción de en qué regiones se darán las condiciones de sequía. Finalmente, el método híbrido utiliza una mezcla de las dos estrategias anteriores (Hao, 2018). Nuestra propuesta de solución maneja un modelo estadístico debido a los altos costos de obtener la tecnología y recolección de datos para desarrollar modelos de circulación general.

Modelo Estadístico

Históricamente, la mayor debilidad de la predicción estadística es que estos métodos normalmente utilizan modelos lineales para sus predicciones, los cuales, en nuestro caso, no representan certeramente las relaciones complejas del sistema climatológico. Por ello, con el reciente auge de la inteligencia artificial, los modelos estadísticos han comenzado a utilizar

una variedad de distintas técnicas tales como Redes Neuronales Artificiales (ANN), Máquinas Vectoriales de Soporte (SVM), Programación Genética (GP), entre otros (Hao, 2018).

Recientemente, el modelo más utilizado en predicciones climatológicas de AI ha sido el modelo ANN por su capacidad de modelar la dependencia no-lineal entre los índices de sequía y sus variables predictoras. Este modelo consiste en utilizar un conjunto de capas ocultas para aprender de las relaciones no lineales de los datos. Sin embargo, como se puede observar en la **figura 2**, las ANN y las SVM por sí solas carecen de las características apropiadas para los modelos de predicción, lo cual requiere que se realicen modificaciones al conjunto de datos, realizando una limpieza respecto a los tipos de datos a emplear. Por ejemplo, se tiene que contemplar el manejo de los valores nulos, remover datos atípicos, mantener las dimensiones de las variables similares para todo el conjunto de datos, y hacer una selección de variables. Este tipo de transformaciones y limpieza es lo que lleva la mayoría del trabajo y consumo de tiempo cuando se desean implementar estos algoritmos. Y así mismo las decisiones respecto a cambios y transformaciones pueden estar sesgadas por decisiones de los mismos desarrolladores.

Por otra parte, también se puede apreciar en la **figura 2** que los algoritmos basados en árboles cumplen con la mayoría de características de manera intrínseca. Es decir, manejan naturalmente datos tipo “mixto”, manejan los valores nulos, tienen robustez ante datos atípicos, son insensibles ante transformaciones monótonas, tienen escalabilidad computacional y dan menor importancia a variables irrelevantes. Sin embargo, por sí solos tienen muy poco poder predictivo y no son tan sencillos de interpretar, pero estos puntos se pueden mejorar drásticamente al aplicar técnicas de ensamble y algoritmos externos de explicabilidad (XGBoost, 2020; Lundberg, 2019).

Característica	ANN	SVM	Árboles
Manejo natural de datos de tipo “mixto”.	▲	▲	▲
Manejo de valores nulos.	▲	▲	▲
Robustez ante valores atípicos.	▲	▲	▲
Insensible a transformaciones monótonas de las variables.	▲	▲	▲
Escalabilidad computacional.	▲	▲	▲
Habilidad para tratar con variables irrelevantes.	▲	▲	▲
Habilidad de extraer combinaciones lineales de las variables.	▲	▲	▲
Interpretabilidad.	▲	▲	▲
Poder predictivo.	▲	▲	▲

Figura 2: Algunas características de distintos métodos de aprendizaje. El color verde significa buen desempeño; el color amarillo, desempeño justo; y el color rojo desempeño pobre. (Esta imagen fue traducida y modificada; Hastie, Tibshirani y Friedman, 2009).

Propuesta de solución

Se propone generar un modelo estadístico supervisado de aprendizaje automático de clasificación para realizar la predicción de la sucesión de sequías en municipios de la República Mexicana con un mes de anticipación. Para realizar el modelo, se tiene contemplado utilizar un algoritmo de ensamble de árboles conocido como *potenciación del gradiente*.

La predicción de sequías permitirá a varios niveles de tomadores de decisiones el minimizar los costos que presentará la crisis. Estos grupos incluyen, pero no se limitan a, políticos locales, administradores hídricos, ejidatarios, negocios privados, grupos indígenas, y principalmente agricultores. Por ejemplo, los administradores de recursos hídricos se verán favorecidos por el modelo de predicción local ya que podrán regular el consumo de agua en varios sectores, evitando la falta de disponibilidad de agua en sus comunidades cuando llegue la crisis. Similarmente, los agricultores podrán más eficientemente determinar su selección y administración de cultivos teniendo un mejor conocimiento de las condiciones climatológicas por venir. Finalmente, los administradores de tierras podrán desarrollar planes con mayor anticipación para evitar la propensión a incendios forestales debido a la sequía.

En la realización del modelo, se escogió el algoritmo de ensamble de árboles debido a que se cuenta con datos tabulares con variables que tienen distintos significados y escalas numéricas. Adicionalmente, se ha demostrado que con esta técnica se han obtenido resultados de estado-del-arte en muchos problemas de aprendizaje automático (Chen y Guestrin, 2016; Shwartz-Ziv y Armon, 2021). Aunado a esto, existen varias implementaciones del algoritmo con paquetes de código abierto, facilitando su uso y su implementación (XGBoost, 2020; LightGBM, 2020; Scikit-Learn, 2019).

Los algoritmos de ensamble basados en árboles son técnicas de aprendizaje automático supervisado no paramétricas que son usados para proporcionar soluciones a problemas de clasificación y regresión. Estos se basan en la combinación de varios algoritmos débiles conocidos como *árboles de decisión*². Al implementar las técnicas de ensamble, el poder predictivo de los árboles se ve impulsado ya que los métodos consisten en combinar de manera estratégica y secuencial estos árboles. Esto hace que, a lo largo de varias iteraciones, el error de aprendizaje se vaya reduciendo y se converja a un valor de predicción.

El algoritmo de potenciación del gradiente funciona de tal manera que se realiza un aprendizaje de los datos a través de la generación continua de *árboles de decisión*. Cada árbol de decisión se añade individualmente a lo largo de muchas iteraciones, ajustándose para

² Los árboles de decisión fueron introducidos desde el punto de vista estadístico en el año 1984 (Breiman *et al.*, 1984). Algorítmicamente, lo que se hace para generar un árbol es una partición recursiva en el espacio multidimensional de las variables mediante ciertos criterios de ganancia de información buscando siempre predecir un valor (ya sea una clasificación o un valor continuo). Por sí solos, los árboles de decisión tienen un desempeño relativamente pobre pero conforme se van agregando más mediante técnicas de ensamble, la predicción se hace mejor (Zhang y Hagami, 2015).

corregir los errores de predicción de todos los árboles previamente añadidos. reduciendo el error del valor de predicción. La versatilidad del algoritmo respecto a la función de costo le permite emplear distintas configuraciones para poder realizar las predicciones deseadas (XGBoost, 2020). En el caso de clasificación binaria, se intenta predecir una variable de respuesta categórica representada por un 0 para la clase negativa y un 1 para la clase positiva³.

Las ventajas de implementar este tipo de algoritmos es que puede abordar distintos tipos de relaciones complejas y no lineales. No existe un problema respecto a la alta dimensionalidad de los datos y cuenta con el beneficio de que requieren un mínimo preprocesamiento y limpieza de datos. Además, son robustos ante datos aberrantes, no tienen sensibilidad ante transformaciones monotónicas, y tienen métricas internas para determinar la importancia de las variables (Zhang y Hagami, 2015; Cutler, Cutler y Stevens, 2008). Finalmente, existen otros algoritmos que tienen una excelente integración para brindar la explicabilidad a estos modelos y por lo tanto facilitan el entendimiento del impacto que tiene cada una de las variables en el desempeño general del algoritmo y la resolución del problema (Lundberg, 2019; Lundberg, Erio, y Lee, 2019). Todo este conjunto de características hacen que el algoritmo de *potenciación del gradiente* sea un candidato ideal para abordar el problema planteado.

Hipótesis

Con un conjunto de variables temporales se puede predecir con una métrica de área bajo la curva de precisión-exhaustividad del 80%, el suceso de una sequía con un mes de anticipación en la república mexicana.

- a) La sequía actual es indispensable para realizar una predicción de la sequía en el futuro.
- b) Las condiciones climatológicas son de las variables más importantes a considerar.

Objetivos

General

Desarrollar un modelo estadístico de aprendizaje automático en una plataforma, donde se pueda consumir de manera sencilla, para realizar predicciones con un mes de anticipación de las sequías en la república mexicana con una métrica de 0.80 de área bajo la curva de precisión-exhaustividad.

³ En el caso específico de este proyecto, 0 hace referencia a “No-sequía” y 1 hace referencia a “Sequía”.

Particulares

Obtener datos adecuados, pertinentes y orientados a la predicción de la variable de respuesta.

Entrenamiento de un modelo de aprendizaje automático con una ventana temporal suficiente para generalizar el fenómeno de sequías a través del tiempo y espacio geográfico mediante un conjunto específico de variables.

Desarrollar una plataforma en la cual se pueda montar el modelo para el consumo del mismo de una manera sencilla y amigable.

Metodología

Para llevar a cabo este proyecto, se tuvieron que contemplar muchas situaciones que lo harían más factible y adecuado de desarrollar dado los recursos temporales y monetarios que se tenían disponibles. Primero, se buscaron las bases de datos correspondientes respecto a las sequías. Una vez teniendo estos datos, se realizó una disminución a la cantidad de registros que se utilizarían para realizar el entrenamiento del modelo debido a que se tiene contemplado utilizar un API que después de cierto número de llamadas, tiene un costo el seguir consumiéndola. Por lo tanto se decidió reducir el número de municipios que se iban a analizar (en lugar de utilizar los 2468 distintos municipios de la república, el número se redujo a 1200, cuidando mantener la mayor representatividad⁴ posible).

Una vez que se contaba con la reducción de los datos, se procedió a hacer un análisis de calidad de datos (revisar **Anexo 2**), dónde se buscaba que no hubiera alguna inconsistencia o tendencias atípicas en la naturaleza de los datos, donde en caso de encontrarlos, se omitiría su uso. Después de realizar esto, se procedió a calcular la variable de respuesta teniendo como objetivo binarizar el nivel de sequía del siguiente mes. Una vez que se tenía toda la información disponible. Se procedió a buscar la información correspondiente en distintas bases de datos. Se buscaron variables geológicas en el geoportal del sistema nacional de información sobre biodiversidad (SNIB)⁵. También se buscaron las variables meteorológicas históricas en la API de DarkSky.

Esto abrió paso a que se calcularan un conjunto de variables que toman información de las condiciones meteorológicas, geológicas y de sequías considerando, por municipio, información respecto a los municipios colindantes o el mismo municipio en el pasado. Generando así todas nuestras variables predictoras y poder llevar a cabo un modelo de aprendizaje automático supervisado de clasificación.

⁴ Para tener la mejor representatividad de los datos de la República Mexicana, se realizó un análisis que contemplaba la disponibilidad de datos de manera histórica, revisar **Anexo 1**.

⁵ Esta información se encuentra en el apartado de bases de datos.

Para llevar a cabo el modelo fue necesario contar con distintos conjuntos para entrenar el modelo y validar el desempeño del mismo. Para comenzar con la asignación de datos a los distintos conjuntos se contempló la totalidad de los datos con los que se contaba en general, denotado por el nombre T. Este conjunto de datos contiene toda la información histórica a lo largo de los municipios de la República Mexicana desde el año 2003 hasta la fecha del 17 de agosto del 2021. Partiendo de aquí se generaron los demás conjuntos de datos correspondientes para el entrenamiento y la prueba.

Se desarrolló un conjunto denotado por M, Este conjunto está definido por una ventana temporal que va desde 2004/01/01 hasta 2019/01/01 y tiene la función de generar todos los conjuntos de entrenamiento y uno de prueba. Respecto a los conjuntos de entrenamiento, el primer conjunto que se generó, está denotado por ME refiriéndonos al conjunto muestreado para realizar el aprendizaje. Este es el conjunto más grande de los tres conjuntos de entrenamiento y está definido por el 65% de las instancias de M seleccionadas de manera aleatoria. De manera similar, el segundo conjunto generado de entrenamiento, es obtenido de manera aleatoria del conjunto M, pero sólo contemplando 10% de las instancias, este será denotado por MH, con el uso específico de la realización de la optimización de hiperparámetros. Así mismo, se generó otro conjunto de entrenamiento denotado por el nombre MP, el cuál tiene el propósito de realizar una parada anticipada a la cantidad total de iteraciones que se requieren para el entrenamiento del modelo⁶. Para este conjunto se utiliza la cantidad de 5% de las instancias del conjunto M.⁷

Adicionalmente a esto, se generaron otros 3 conjuntos de prueba para monitorear el desempeño. El primer conjunto de prueba es el conjunto FDME, haciendo referencia a *fuera del conjunto de muestreo*. Este conjunto de prueba contiene los datos dentro de la misma ventana temporal con los cuales se realiza el entrenamiento, para este conjunto se tomó de manera aleatoria, la cantidad sobrante de datos, que en este caso representa el 20% de la totalidad de M. El segundo conjunto de prueba es el conjunto FDTP haciendo referencia a *fuera de la ventana temporal en el pasado*, y tal como su nombre lo indica, tiene datos anteriores a la ventana temporal del entrenamiento. De manera similar, el tercer conjunto de prueba será el conjunto FDTF haciendo referencia a *fuera de la ventana temporal en el futuro*, pero en este caso, tiene datos posteriores a la ventana temporal del entrenamiento. En la **Tabla 1** se puede apreciar de manera resumida la incorporación de los conjuntos de datos así como su uso.

El modelo que se utiliza para realizar las predicciones está basado en árboles y es llamado *árboles de potenciación del gradiente*, el cuál fue seleccionado debido a las características de las diferencias existentes respecto a la naturaleza de las variables previamente mencionadas. Sin embargo, se codificaron las variables categóricas mediante la técnica de CatBoost-Encoder para cambiar de valores categóricos a numéricos.

⁶ Esto se debe contemplar debido a que el algoritmo empleado es iterativo.

⁷ Es importante recalcar que el mismo municipio no se ve repetido en más de un conjunto.

Se buscó realizar una optimización de hiperparámetros para determinar las características que el modelo tendrá utilizando el conjunto de datos MH. Posteriormente se realizó el entrenamiento con el conjunto de datos ME utilizando el conjunto MP para detener el aprendizaje cuando las iteraciones dejaran de ser significativas. Finalmente, se evaluó con los conjuntos de datos:

- FDME para tener un entendimiento del desempeño en la misma ventana temporal.
- FDTP para ver cómo le habría ido al modelo en el pasado.
- FDTF para ver el desempeño del modelo con los datos más actuales.

A la par, se realizó lo mismo con el modelo base, pero ahora contemplando los mismos conjuntos de entrenamiento y prueba para tener contra qué comparar el desempeño del modelo. Una vez que se tenía el desempeño final de este modelo, se aplicaron distintas métricas para evaluar su desempeño comparándolo con el modelo base y con el valor predicción de la sequía respecto al presente. Esto se realizó con la finalidad de asegurar que la evaluación del modelo fuera mejor un mes hacia el futuro que en el presente.

Finalmente, para validar los resultados del modelo se analizaron las variables más importantes mediante la técnica de explicabilidad de modelos conocida como SHAP para determinar si el modelo en efecto tenía sentido y apego a la realidad.

Tabla 1. Conjuntos de datos: características y uso.

Conjunto de datos	Característica	Uso
T	Conjunto total de información	Generación de todos los conjuntos
M	Conjunto de datos definido por una ventana temporal.	Generación de conjuntos de entrenamiento y prueba.
MH	10% de M	Optimización de hiperparámetros.
MP	5% de M	Parada anticipada del algoritmo
ME	65% de M	Conjunto de aprendizaje del modelo.
FDME	20% de M	Conjunto de prueba del modelo.
FDTP	Conjunto de datos con ventana temporal anterior a M.	Conjunto de prueba en el pasado.
FDTF	Conjunto de datos con ventana temporal posterior a M.	Conjunto de prueba en el futuro.

Después de haber entendido el funcionamiento del modelo y validado su desempeño, el modelo se montó en una página web donde su acceso es sencillo y amigable.

Bases de datos

El corazón de este proyecto se basa primordialmente en el *Monitor de Sequía de México*. Esos datos son proporcionados por el Servicio Meteorológico Nacional (SMN) utilizando distintas redes de observación⁸. Este conjunto de datos consiste en información municipal de cada entidad de la república mexicana y contempla datos desde el 31 de enero del 2003 hasta la actualidad⁹, donde se tiene el correspondiente nivel de sequía de acuerdo al índice USDM.

Adicionalmente a estos datos, se cuenta con una API ¹⁰llamada Dark Sky que proporciona datos meteorológicos al proporcionarle una geolocalización y una fecha en particular. En este caso en específico se hará uso de la funcionalidad llamada “Time Machine Request”, donde

⁸Las redes de observación que se utilizan son: estaciones automáticas, observatorios sinópticos, radares, estaciones de radiosondeo y estaciones receptoras de imágenes de satélite.

⁹ La “actualidad” se considera como el 17 agosto 2021.

¹⁰ Para realizar una llamada a la API, se tiene que contar con la siguiente estructura:
<https://api.darksky.net/forecast/{llave}/{lat},{lon},{aaaa-mm-ddThh:mm:ss}?units=si>

la API otorga las condiciones climáticas en algún punto del pasado. Debido a que los datos del *Monitor de Sequía de México* son datos históricos, esta funcionalidad es precisamente lo que se requiere para obtener las condiciones meteorológicas que se requieren.

También se incorporan datos del Portal de Geoinformación 2021, del SNIB. Debido a que no tienen información de manera temporal, se obtuvieron aquellos datos que se mantienen relativamente estáticos a través de la escala temporal que se está utilizando y en su mayoría son de índole geológico. En específico, se obtuvo información acerca de: cuencas, subcuencas, curvatura de perfil, curvatura del plano, curvatura del terreno, edafología, hidrología, hidrografía, hipsometría, índice topográfico de humedad y pendiente.

Tabla 2.

Fuentes de datos.

Datos	Fuente	Formato
Sequía en México.	Monitor de sequía de México	xlsx
Metereológicos	Dark Sky API	JSON
Geológicos	Portal de Geoinformación	shp, raster

Justificación de variables

Una sequía es un fenómeno complejo debido a la alta cantidad de variables que determinan su existencia e intensidad (NOAA, 2021a). Por lo tanto, el modelo requiere contemplar distintos factores, así como sus interacciones complejas, para explicar y predecir el comportamiento de este fenómeno.

Para entender en su totalidad la problemática, primero se tiene que definir una unidad de análisis, de la cual, es más sencillo determinar todos estos factores que contribuyen a la explicación. Estos factores se pueden observar en la **figura 3**.

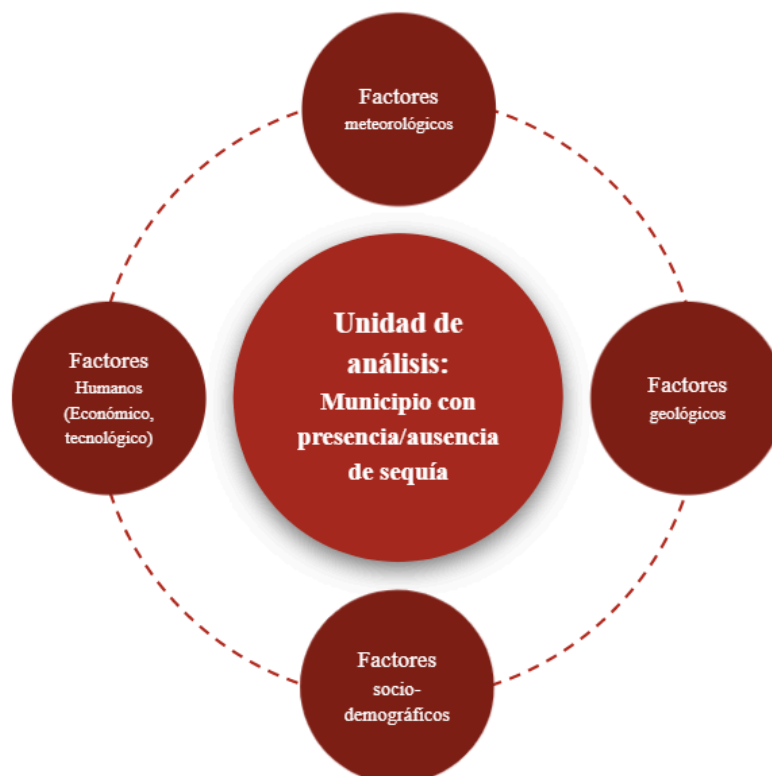


Figura 3: Factores que componen a la unidad de análisis.

En este caso, la unidad de análisis es un municipio con presencia o ausencia de sequía. De esta unidad, sólo se analizan los factores meteorológicos y geológicos altamente relacionados con la presencia de sequías ya que es posible encontrar la información en bases de datos y APIs. Así mismo, se descartan los factores humanos y sociodemográficos debido a la falta de datos históricos y a que lo que se busca predecir en específico es una sequía natural y no tanto una humana o económica.

Variable de respuesta

Sequía en México

Se cuenta con información histórica de la presencia e intensidad de sequía en México medida por el índice USDM. Sin embargo, para poder simplificar la presencia/ausencia de esta variable, se pretende binarizar el índice tal que el valor de la variable de respuesta sea 0 si el valor del índice es Nulo o D0. De manera consecuente, el valor de la variable de respuesta es 1 si el valor del índice es D1, D2, D3 o D4. No obstante, debido a que se quiere predecir la sequía con un mes de anticipación, es imperativo contemplar un desfase en el tiempo de la variable de respuesta deseada, tal que se tenga un este valor de la variable en el futuro. Lo que se busca responder con esta variable es prácticamente la siguiente pregunta: *¿Existe sequía el siguiente mes?*. En la **figura 4 y 5** se puede observar, para una fecha en específica en cada municipio, el nivel de sequía y la binarización correspondiente. Esta da una idea más clara de qué es lo que se busca tener como variable de respuesta.

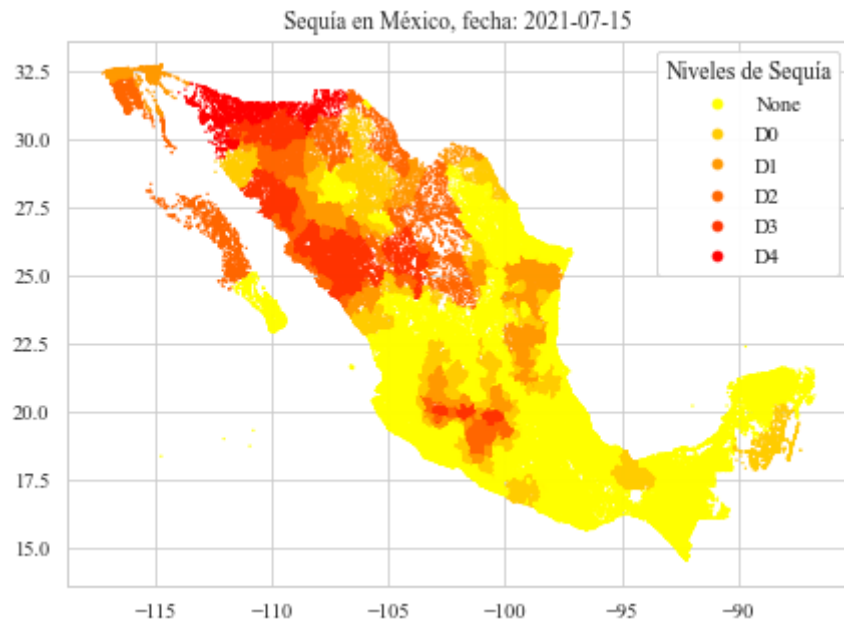


Figura 4: Nivel de sequía en los municipios de la República Mexicana para la fecha de 15 de agosto del 2021.

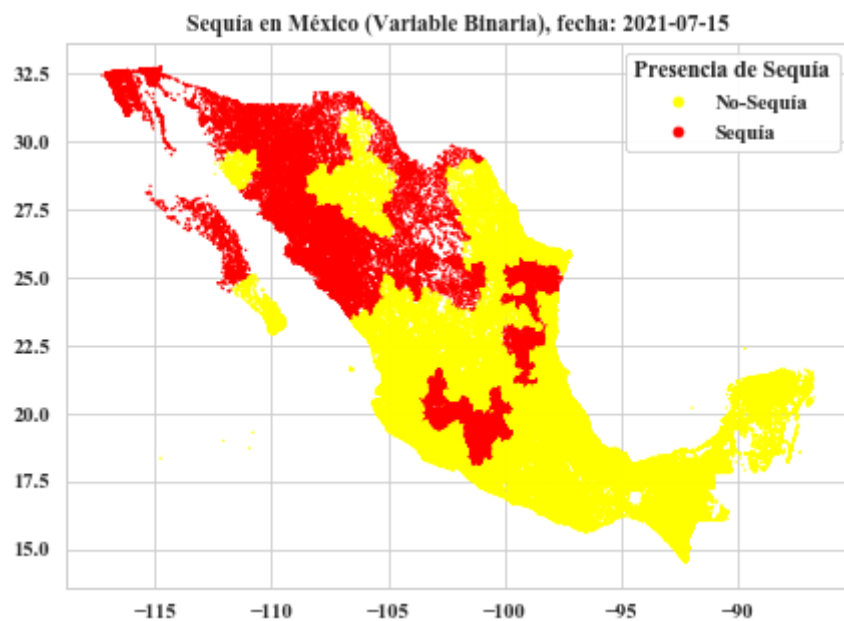


Figura 5: Binarización del Nivel de sequía en los municipios de la República Mexicana para la fecha de 15 de agosto del 2021.

Variables Independientes

Para determinar qué variables independientes se utilizan en nuestro modelo, se consideró la disponibilidad de datos así como su relevancia predictora para la existencia de una sequía. Por ejemplo, hay un fácil acceso a variables socioeconómicas de nuestra unidad de análisis tales como población, PIB, o consumo de recursos. Sin embargo, la predicción de una sequía meteorológica, en comparación a una sequía agrícola, no es afectada por estas variables, por

lo que se decidió no incluir esta categoría de información. Similarmente, variables con alta correlación a la existencia de una sequía, tal como el desarrollo de fenómenos climatológicos globales tales como el niño o la niña, no pudieron ser incluidos en el modelo ya que estos datos son altamente difíciles de obtener sin el uso de modelos climatológicos dinámicos globales, los cuales presentan varias barreras económicas, técnicas, y logísticas para su uso en nuestro proyecto.

Por lo tanto, este modelo estará conformado por variables locales de fácil obtención con alta adecuación histórica a la existencia de una sequía. Con base en este criterio, se clasificaron las variables predictoras en dos categorías: variables meteorológicas temporales, y variables geológicas atemporales. Las variables meteorológicas temporales son mediciones atmosféricas que varían históricamente mientras que las variables geológicas atemporales son mediciones del suelo, relieve, y terreno que son constantes en nuestra escala temporal. En los siguientes apartados se discutirá la justificación individual de cada una de nuestras variables independientes, estableciendo su relación con la sucesión de sequías determinada por los factores de clasificación del índice USDM.

Variables meteorológicas

Las variables meteorológicas locales, específicamente temperatura y precipitación, son unos de los predictores más comunes en los modelos de predicción de sequía debido a su alta correlación con la sucesión de este fenómeno climatológico. Estos modelos normalmente consideran la relación temporal entre el índice de sequía y los factores meteorológicos en un periodo antecedente de un mes i.e NOAA, IEEE, Instituto Americano de Geociencias (AGU). Además de la inclusión de temperatura y precipitación, se decidió incluir variables secundarias altamente correlacionadas a estas medidas primarias para obtener una predicción más certera de las condiciones climatológicas locales. Este conjunto de variables busca cubrir el grupo uno de los factores de intensidad de sequía USDM.

Estas variables están representadas a través de un *punto de dato*; indicando que cada punto representa un evento meteorológico en cierta coordenada y cierto tiempo con diversas características. En específico, las características son las variables donde cada variable representa el promedio de un fenómeno climático ocurriendo en el periodo de tiempo deseado.

Para las variables meteorológicas calcularon varias métricas de agrupación temporal como sus valores estadísticos en los últimos 180 días. Esto con la finalidad de tener más información.

Temperatura

Los datos de temperatura miden el valor máximo, mínimo y la temperatura aparente del día especificado en unidades de grados celsius. La temperatura, al ser una medida de la energía térmica en la región, está altamente correlacionada con la mayoría de los factores USDM.

Esta variable está involucrada en el Índice de sequía de Palmer al ser uno de los factores del mismo índice. La temperatura también ayuda a determinar el nivel de precipitación de una región al permitir la condensación de agua en la atmósfera. Esta afecta la humedad de la tierra y reservas subterráneas al incrementar o disminuir la evaporación de agua en el sistema y similarmente afecta el nivel de nieve en la región al controlar la cantidad de derretimiento. Finalmente, la temperatura genera respuestas de la flora local medidas por el VegDRI.

Índice ultravioleta

El Índice ultravioleta es la medición de la fuerza de radiación UV solar. Esta medición complementa a la temperatura regional ya que esta depende directa y positivamente con la cantidad de energía que la región absorbe del sol.

Cubrimiento de nubes

Indica el porcentaje de cielo que fue ocluido por las nubes. Es un coeficiente con valores entre 0 y 1 (inclusivo). El cubrimiento de nubes sirve como complemento al Índice UV y a la temperatura ya que este dicta altamente el albedo de la región, afectando la cantidad de radiación solar que entra al sistema meteorológico regional.

Precipitación

Mide la intensidad y la probabilidad de ocurrencia de precipitación. La intensidad está medida en mililitros por hora, mientras que la probabilidad es un coeficiente con valores entre 0 y 1 (inclusivo). La precipitación está directamente representada en los factores USDM, por lo que es uno de los predictores climatológicos locales más relevantes del modelo.

Presión atmosférica

Se tiene el dato de la presión del aire en hectopascales. La presión atmosférica está relacionada, junto con la temperatura y humedad del aire, al nivel de precipitación de una región.

Humedad

Es el valor de humedad relativa. Es un coeficiente con valores entre 0 y 1 (inclusivo). La humedad relativa es la proporción de la cantidad de agua en el aire al límite que el aire puede sostener a cierta temperatura sin que el agua se precipite. Por ende, está altamente relacionado con la precipitación.

Punto de rocío

Las unidades son grados celsius. El punto de rocío es la temperatura a la cual se empieza condensar el agua de la atmósfera, y este está determinado por la humedad relativa y presión atmosférica. Por lo tanto, está altamente relacionado con la precipitación local.

Variables geológicas

Las variables geológicas escogidas buscan cubrir el rango teórico establecido por el grupo dos, tres, y cuatro de los factores USDM. Las variables de humedad de suelo, escorrentía hidrológica, pico de aguanieve, almacenamiento hidrológico, y el índice de respuesta vegetal presentan altas barreras económicas para la recolección de datos debido a que estas no se pueden ser medidas satelitalmente como las variables meteorológicas. Por lo tanto, es escasa la disponibilidad de datos a nivel local en toda la república. Como solución, se escogieron variables geobiológicas atemporales de fácil acceso que presentan una alta correlación a las variables deseadas para ser usadas como sustitutos en la predicción. A continuación se discute la relación de las variables atemporales utilizadas con las variables óptimas deseadas como justificación de su elección.

Hidrografía

Para los datos de hidrografía, se obtuvo un mapa de la red hidrológica a nivel nacional. Este mapa sólo representa la red hidrográfica y no el tipo de flujo ni el caudal del mismo, se cuenta con un dominio, que representa la longitud o distancia de cada río. Se decidió utilizar la distancia de un punto terrestre al sistema fluvial más cercano como uno de los predictores de humedad de suelo, escorrentía hidrológica, y almacenamiento hidrológico. Se cuenta con un dominio espacial por coordenadas extremas de la siguiente manera: Oeste: -117.131639, Este: -88.27261, Norte: 32.60798 y Sur: 14.566166. Se completó con la nomenclatura de ríos escala 1:1000000 bajo el propósito de conocer las principales corrientes superficiales del país contemplando un modelo geodésico WGS84. Este conjunto de datos fue publicado el 03 de agosto de 1998, en formato de representación geográfica Shapefile (Maderrey-R, y Torres-Ruata, 1990).

Cuencas

Una cuenca hidrológica es una unidad hidrogeomórfica que representa el área terrestre que es drenado a un mismo sistema fluvial y sus tributarios. Se utilizará la afiliación del punto terrestre a su cuenca como predictor de datos hidrológicos. Este predictor se escogió debido a nuestra estimación de que los puntos terrestres en una misma cuenca hidrológica compartirán una alta correlación en condiciones hidrológicas. Para los datos de cuencas, se obtuvo un mapa que representa las cuencas hidrológicas de la república mexicana a escala 1:1250000. Se cuenta con dominio de un área de polígono asociado a una cuenca hidrológica. Se cuenta con un dominio espacial por coordenadas extremas de la siguiente manera: Oeste: -118.400821, Este: -86.711271, Norte: 32.718968 y Sur: 14.537673. Para este conjunto de datos, se realizó una digitalización, contemplando un modelo geodésico WGS84. Este conjunto de datos fue publicado el 11 de mayo de 2001, en formato de representación geográfica Shapefile (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad [CONABIO], 2001).

Subcuencas

Para los datos de subcuencas, se obtuvo un mapa que representa las subcuencas hidrológicas del país a escala 1:1000000. Se cuenta con dominio de un área de polígono asociado a una subcuenca hidrológica. Se cuenta con un dominio espacial por coordenadas extremas de la

siguiente manera: Oeste: -118.400821, Este: -86.711271, Norte: 32.718968 y Sur: 14.537673. Para este conjunto de datos, se realizó una conversión del medio analógico al digital mediante el uso de una tableta digitalizadora contemplando un modelo geodésico WGS84. Este conjunto de datos fue publicado el 18 de marzo de 2001, en formato de representación geográfica Shapefile (Comisión Nacional del Agua [CONAGUA], 2001).

Índice topográfico de humedad

El índice topográfico de humedad resume la posición y configuración topográfica de un punto terrestre para determinar la influencia de estos factores en la humedad relativa de la región. Mientras más alto sea el índice topográfico, más propenso es la región a la acumulación de humedad en la tierra. Por lo tanto, se utiliza este índice topográfico de humedad aunado a otras medidas topográficas como topografía, curvas de nivel, curvas de perfil, curvatura de plano y de terreno, y la pendiente, como sustituto a las variables de humedad de suelo, escorrentía hidrológica, y almacenamiento hidrológico. Se obtuvo información de un mapa que indica el área de captación y la pendiente para identificar y delimitar las zonas receptoras, indica la tendencia de la topografía en la ubicación y el tamaño de las zonas saturadas con transición a zonas no saturadas. Se cuenta con un dominio espacial por coordenadas extremas de la siguiente manera: Oeste: -119.351260954228, Este: -84.5030540088871, Norte: 33.4034819577191 y Sur: 13.7526878979966. El modelo geodésico utilizado es WGS84. Este conjunto de datos fue publicado el 31 de mayo del 2016 en formato GEOTIFF, formato raster con resolución (1000,1000) metros (Guevara, Arroyo-Cruz, 2016b).

Hipsometría

Para la hipsometría, se obtuvo un mapa con rangos hipsométricos a nivel nacional que oscilan entre 200 a más de 5000 metros sobre el nivel del mar (msnm), en una escala 1:4000000. Se cuenta con un dominio que representa el área de los polígonos asociados a una franja altimétrica mediante un intervalo de altitud. Se cuenta con un dominio espacial por coordenadas extremas de la siguiente manera: Oeste: -118.390358, Este: -86.659914, Norte: 32.73177 y Sur: 14.566159. El modelo geodésico utilizado es WGS84. Este conjunto de datos fue publicado el 10 de mayo de 2001 en formato de representación geográfica Shapefile (Lugo-Hupb, Vidal Zepeda, *et al*, 2001).

Curvatura de perfil

Para los datos de curvas de perfil, se obtuvo un mapa que tiene informaciones que indican la convexidad o concavidad del terreno en un perfil vertical al plano de una línea de flujo de agua superficial, en radianes por unidad lineal. El conjunto de datos se desarrolló con el propósito de generar atributos topográficos (primarios y secundarios) utilizando el modelo digital de elevaciones generado por el INEGI. Se cuenta con un dominio espacial por coordenadas extremas de la siguiente manera: Oeste: -119.351260954228, Este: -84.5030540088871, Norte: 33.4034819577191 y Sur: 13.7526878979966. Se cuenta con un dominio tipo raster donde se tienen valores de cada pixel. Este conjunto de datos contempla un modelo geodésico WGS84 y fue publicado el 31 de mayo del 2016, en formato

de representación geográfica GEOTIFF raster Resolución (1000,1000) metros (Guevara, Arroyo-Cruz, 2016).

Curvatura del plano

Para los datos de curvatura del plano, se obtuvo un mapa que tiene informaciones que indican la convexidad o concavidad del terreno dado el plano horizontal de una línea de flujo de agua superficial, en radianes por unidad lineal. El conjunto de datos se desarrolló con el propósito de generar atributos topográficos (primarios y secundarios) utilizando el modelo digital de elevaciones generado por el INEGI. Se cuenta con un dominio espacial por coordenadas extremas de la siguiente manera: Oeste: -119.351260954228, Este:-84.5030540088871, Norte: 33.4034819577191 y Sur: 13.7526878979966. Se cuenta con un dominio tipo raster donde se tienen valores de cada pixel. Este conjunto de datos contempla un modelo geodésico WGS84 y fue publicado el 31 de mayo del 2016, en formato de representación geográfica GEOTIFF raster Resolución (1000,1000) metros (Guevara, Arroyo-Cruz, 2016a).

Curvatura del terreno

Para los datos de curvatura del plano, se obtuvo un mapa que tiene informaciones que indican si el terreno tiene una curvatura positiva y convexa o negativa cóncava. El conjunto de datos se desarrolló con el propósito de generar atributos topográficos (primarios y secundarios) utilizando el modelo digital de elevaciones generado por el INEGI. Se cuenta con un dominio espacial por coordenadas extremas de la siguiente manera: Oeste: -119.351260954228, Este:-84.5030540088871, Norte: 33.4034819577191 y Sur: 13.7526878979966. Se cuenta con un dominio tipo raster donde se tienen valores de cada pixel. Este conjunto de datos contempla un modelo geodésico WGS84 y fue publicado el 31 de mayo del 2016, en formato de representación geográfica GEOTIFF raster Resolución (1000,1000) metros (Guevara, Arroyo-Cruz, 2016b).

Pendiente

Para la pendiente, se obtuvo información de un mapa que indica la tasa de cambio de la elevación del terreno con la distancia a una dirección dada. Se cuenta con un dominio espacial por coordenadas extremas de la siguiente manera: Oeste: -119.351260954228, Este: -84.5030540088871, Norte: 33.4034819577191 y Sur: 13.7526878979966. El modelo geodésico utilizado es WGS84. Este conjunto de datos fue publicado el 31 de mayo del 2016 en formato GEOTIFF, formato raster con resolución (1000,1000) metros (Guevara, Arroyo-Cruz, 2016c).

Hidrogeología

Para los datos de hidrogeología, se muestran las unidades hidrogeológicas¹¹ de México caracterizando la porosidad y permeabilidad del subsuelo. Similarmente al uso de cuencas hidrológicas, se utilizará la afiliación de un punto terrestre a su unidad hidrológica y a su categoría de suelo como predictor de factores hidrogeológicos en la región. Se escogieron estos predictores debido a que se estima una alta correlación en condiciones hidrogeológicas en puntos terrestres que comparten unidades hidrogeológicas y edafológicas. Se cuenta con un dominio que representa el área de los polígonos que indican el nombre e identificador numérico para cada unidad hidrogeológica en un dominio espacial por coordenadas extremas localizado de la siguiente manera: Oeste: -118.390358, Este: -86.659914, Norte: 32.736646, Sur: 14.566168. Este conjunto de datos contempla un modelo geodésico WGS84 y fue publicado el 29 de enero de 2002 en formato de representación geográfica Shapefile (Marin-C y Torres-Ruata, 2002).

Edafología

Para los datos de edafología, se tiene un mapa que muestra los distintos tipos de suelo encontrados a nivel nacional a partir de la unión de 32 coberturas: 17 a escala 1:250000 y 15 a 1:1000000. Se maneja información respecto al tipo de suelo, textura, fase física y fase química a través de un dominio donde se tiene el área del polígono asociado al tipo de suelo, que fue obtenidos a través del proyecto P147: Enriquecimiento y uso de la base de datos geográficos del Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP) apoyado por CONABIO. Se cuenta con un dominio espacial por coordenadas extremas de la siguiente manera: Oeste: -117.124754, Este: -86.709521, Norte: 32.715208, Sur: 14.53172. Este conjunto de datos contempla un modelo geodésico WGS84 y fue publicado el 10 de julio del 2001 en formato de representación geográfica Shapefile (INIFAP, 2001).

¹¹ Las unidades que se describen son las siguientes:

- Pleistoceno y reciente: Terrazas marinas, gravas, arenas y limos, depósitos aluviales y lacustres, permeabilidad media a alta.
- Terciario marino: lutitas, limolitas, areniscas y conglomerados, permeabilidad baja a alta.
- Terciario continental, cenozoico superior e inferior clástico, areniscas y conglomerados predominantes; permeabilidad media a alta.
- Cenozoico superior volcánico, rocas volcánicas principalmente basálticas y andesíticas; Permeabilidad media a alta.
- Cenozoico medio volcánico, rocas volcánicas, predominantemente riolitas, permeabilidad baja a media.
- Cenozoico mesozoico y paleozoico intrusivo, rocas intrusivas graníticas granodioritas y doleritas.
- Cretácico superior, rocas sedimentarias marinas predominantemente arcillosas, permeabilidad baja
- Cretácico medio e inferior, terciario, rocas sedimentarias marinas predominantemente calcáreas, permeabilidad alta
- Jurásico lutitas, limolitas, arenosas areniscas y conglomerados limoarcillosos. Permeabilidad baja
- Mesozoico paleozoico y precámbrico, rocas metamórficas esquistos, cuarcitas y gneis, permeabilidad baja.
- Principales cuerpos de aguas tanto naturales (lagos) como artificiales que influyen a los acuíferos locales.

Experimentos

Se realizaron dos modelos para poder generar una comparación válida respecto a las métricas que se utilizarían.

Modelo Simple:

El primer caso (Modelo Simple) fungió de un modelo que contenía pocas variables para predecir la presencia de sequía el siguiente mes las cuales se pueden apreciar en el **Anexo 3**. Este modelo es un clasificador de potenciación de gradiente utilizando la librería de XGBOOST. Los siguientes hiperparámetros fueron elegidos para realizar el entrenamiento.

```
XGBClassifier(base_score=0.99, booster='gbtree', colsample_bylevel=0.8,
              colsample_bynode=0.8, colsample_bytree=0.8, gamma=5, gpu_id=-1,
              importance_type='gain', interaction_constraints="",
              learning_rate=0.05, max_delta_step=0, max_depth=4,
              min_child_weight=1, missing=nan, monotone_constraints=()),
              n_estimators=10000, n_jobs=0, num_parallel_tree=1, random_state=0,
              reg_alpha=3, reg_lambda=10, scale_pos_weight=1.3217302452316075,
              subsample=0.7, tree_method='exact', validate_parameters=1,
              verbosity=None)
```

Modelo Complejo

El segundo caso (Modelo complejo) fue un modelo donde se incorporaron todas las variables calculadas para realizar la misma predicción. Las variables se pueden apreciar en el **Anexo 3**. Este modelo al igual que el anterior es un clasificador de potenciación de gradiente utilizando la librería de XGBOOST. Lo único que cambia, fundamentalmente y con el fin de generar una comparación justa es la cantidad de variables predictoras que se utilizan, sin embargo se entrenaron con el mismo conjunto de entrenamiento.

```
XGBClassifier(base_score=0.99, booster='gbtree', colsample_bylevel=0.8,
              colsample_bynode=0.8, colsample_bytree=0.8, gamma=5, gpu_id=-1,
              importance_type='gain', interaction_constraints="",
              learning_rate=0.05, max_delta_step=0, max_depth=4,
              min_child_weight=1, missing=nan, monotone_constraints=()),
              n_estimators=10000, n_jobs=0, num_parallel_tree=1, random_state=0,
              reg_alpha=3, reg_lambda=10, scale_pos_weight=1.3217302452316075,
              subsample=0.7, tree_method='exact', validate_parameters=1,
              verbosity=None)
```

Resultados

Después de realizar el entrenamiento de los modelos, se obtuvieron las curvas de aprendizaje en el conjunto FDTF, el cual es el más riguroso debido a que contiene a los valores más

actuales (posteriores al 2019). Se puede apreciar en la **Figura 6** que el desempeño del modelo complejo está siempre por arriba del modelo simple en dos métricas distintas.

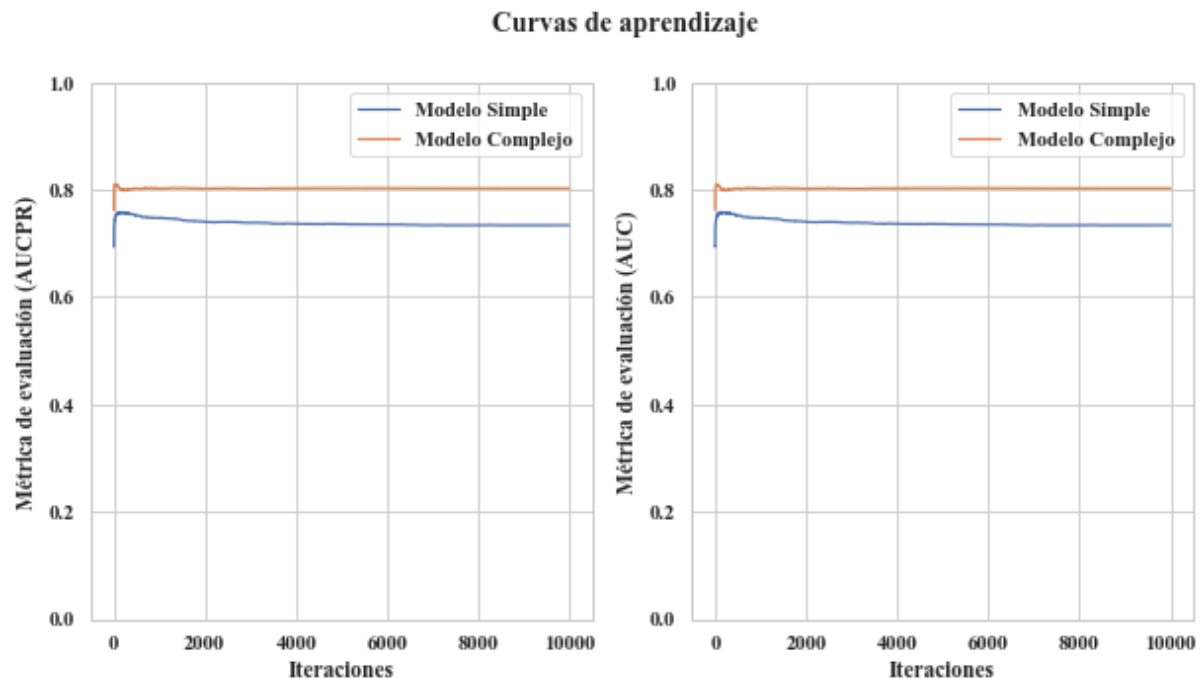


Figura 6: Comparación de curvas de aprendizaje para el modelo simple y el modelo complejo contemplando dos métricas distintas, AUCPR (izquierda) y AUC ROC (derecha).

Esto se puede corroborar de manera visual en la **figura 7**, dónde se alcanza a apreciar que en la distribución del modelo complejo existe una mejor separabilidad de la variable de respuesta contemplando el score otorgado por el mismo. Esto quiere decir que la mayoría de las instancias que contienen sequía fueron clasificados de una mejor manera. Esto es similar para aquellas instancias donde su variable de respuesta es no sequías.

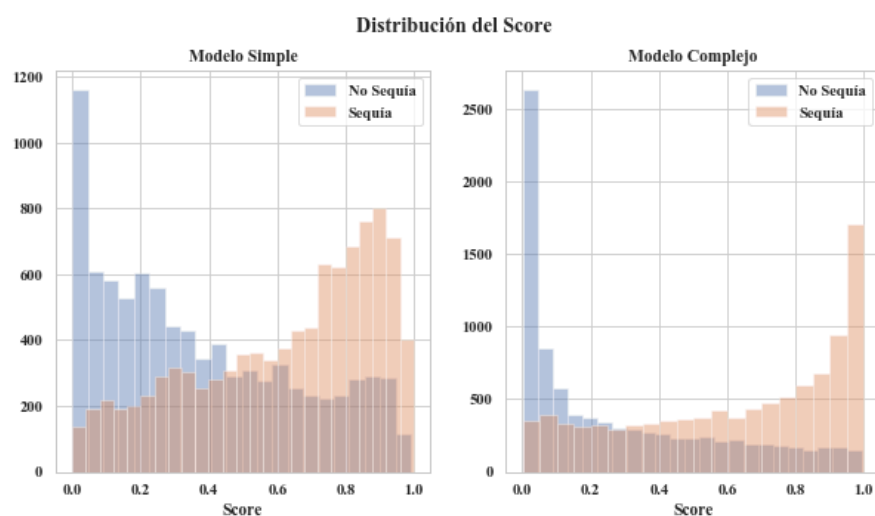


Figura 7: Comparación de distribución del score para el modelo simple y el modelo complejo.

Debido a estos resultados, es relativamente trivial definir que el modelo complejo tiene un mejor desempeño que el modelo simple. De igual manera esto se corroboró de manera cuantitativa utilizando la métrica del área bajo la curva de Precisión-Exhaustividad (referirse a la **Figura 8**).

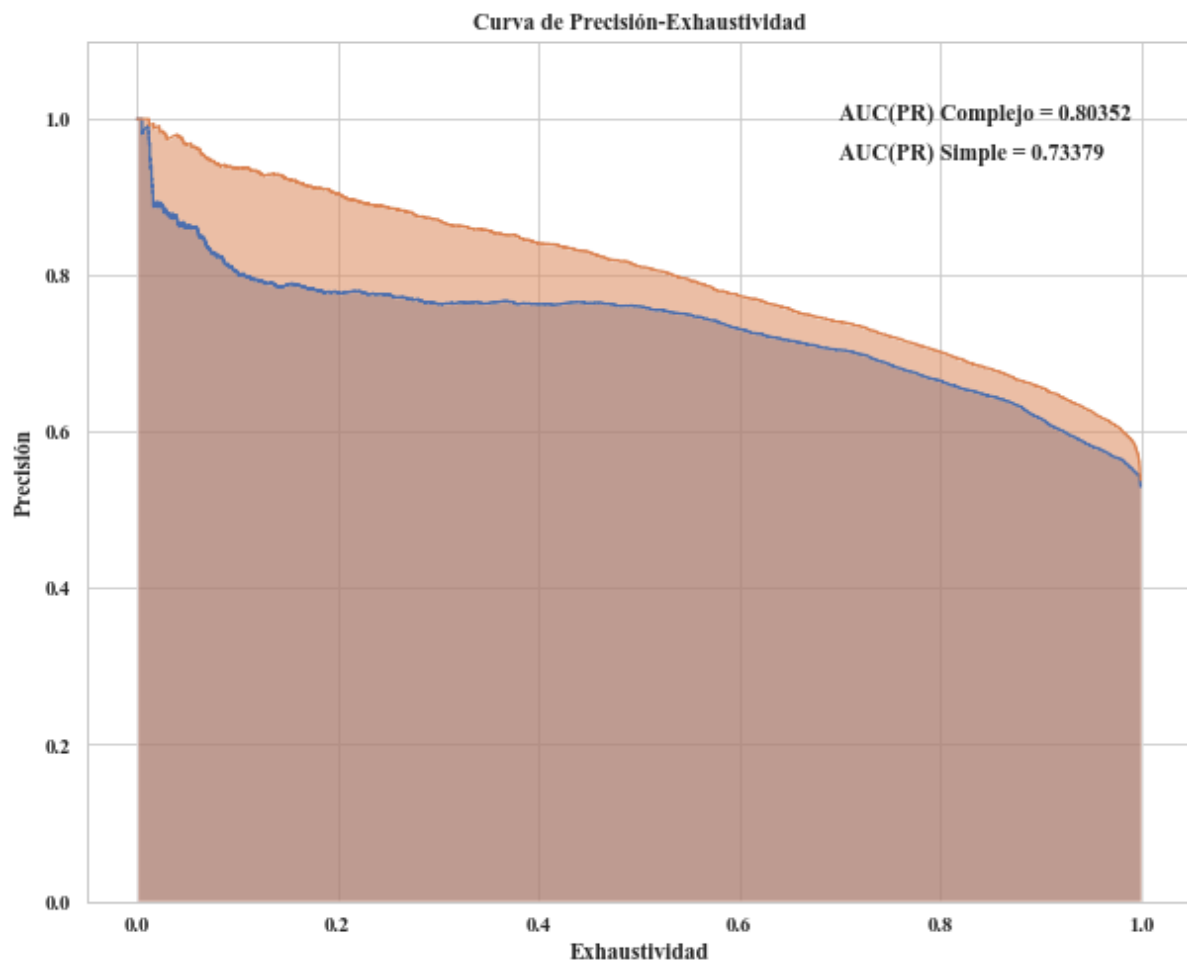


Figura 8: Comparación de curvas de Precisión-Exhaustividad y sus áreas correspondientes para el modelo simple y el modelo complejo. Se puede identificar claramente que el modelo complejo (naranja) tiene un área mayor que el modelo simple (azul).

Gracias a todas estas características, se procedió a hacer un análisis del modelo complejo respecto a su desempeño individual con la variable de interés y respecto a las variables más importantes (explicabilidad). En la **figura 9** se puede apreciar el desempeño del modelo para predecir el comportamiento de la sequía en un estado de la república Mexicana¹². Se puede ver cualitativamente que el valor numérico promedio de la predicción de todos los municipios se acerca mucho al valor promedio de presencia de sequía en los mismos. Incluso, en varias ocasiones llega a subir al mismo tiempo. Esto nos indica que en ciertos casos la presencia de sequía se determina de manera correcta por lo que podemos decir que el modelo está calibrado respecto a la variable de respuesta. Sin embargo hay casos donde existe un retraso respecto al valor promedio (la predicción sube o baja, después del valor real).

¹² Para ver los resultados de todos los Estados, referirse al **anexo 4**.

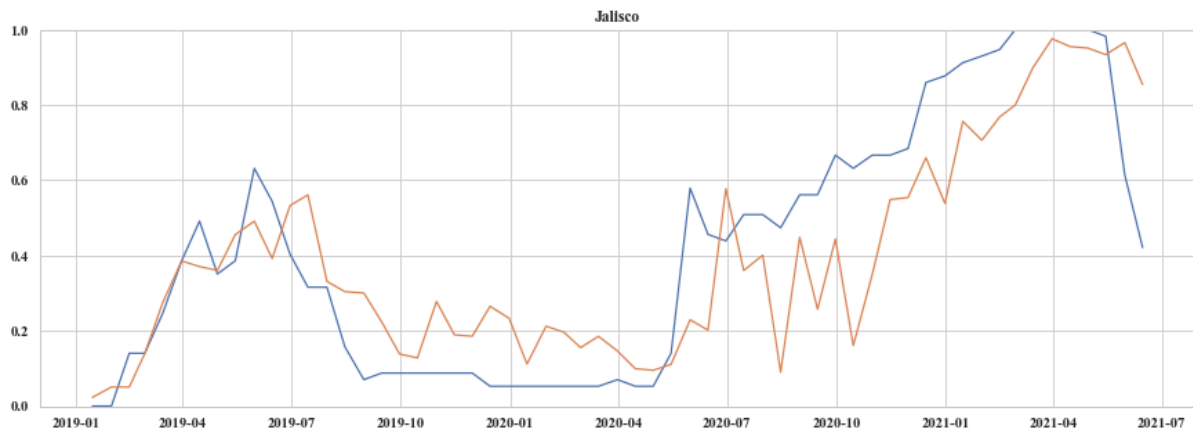


Figura 9: Desempeño visual de predicción (línea naranja) respecto a la realidad (línea azul) para el estado de Jalisco.

Finalmente, para validar que el desempeño del modelo tenga sentido y apego con la realidad, se realizó la técnica de explicabilidad SHAP y se determinaron las variables más importantes. Estas se pueden ver en la **figura 10**.

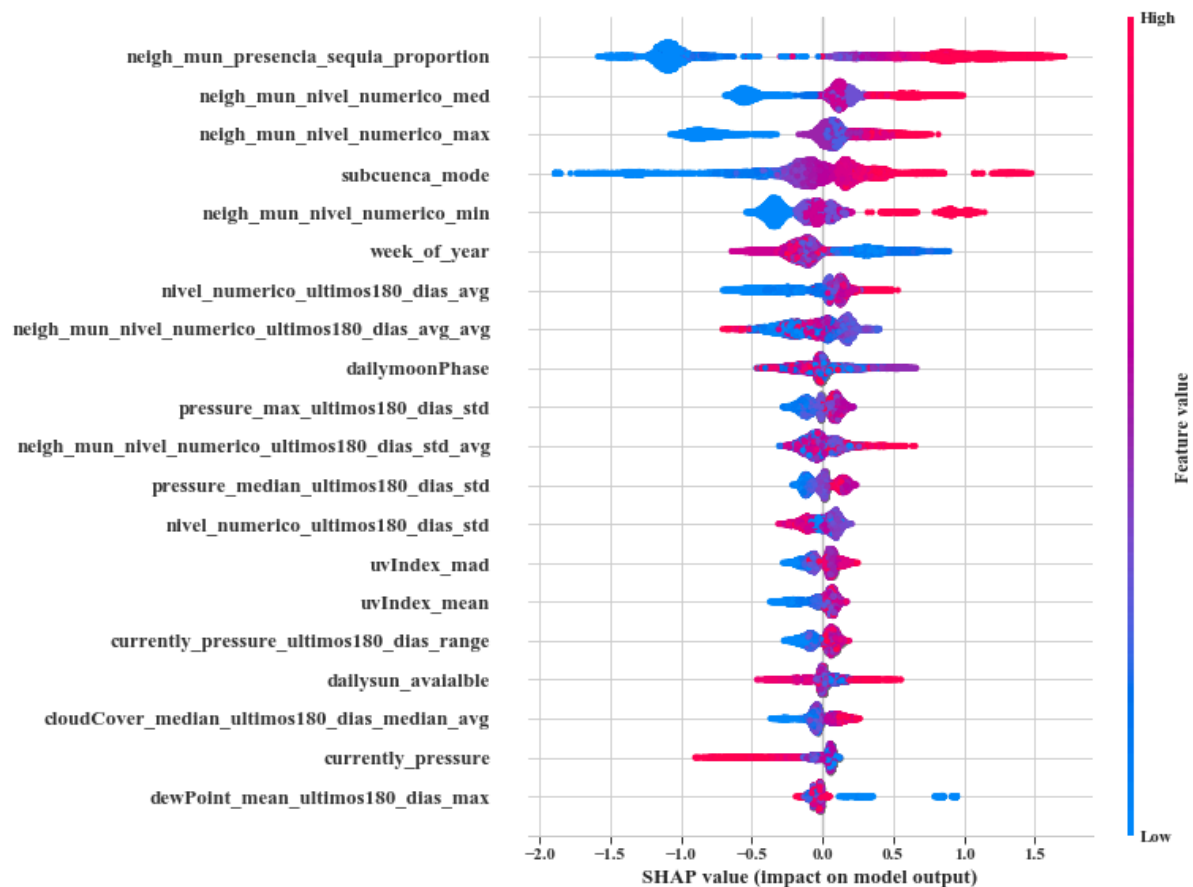


Figura 10: Mejores 20 variables predictoras para la presencia de sequía del siguiente mes utilizando la técnica de explicabilidad SHAP.

Se puede apreciar que las variables más importantes hacen referencia a la proporción de sequías en los municipios colindantes al municipio de interés, al nivel numérico de sequía de estos mismos municipios colindantes (los valores de la mediana, el valor máximo y el valor mínimo, respectivamente). También se considera la subcuenca perteneciente, la semana del año en la cual se está intentando predecir y el nivel numérico promedio de los últimos 180 días de ese mismo municipio.

Conclusiones

Se logró desarrollar un modelo estadístico de aprendizaje automático para realizar predicciones con un mes de anticipación de las sequías en la república mexicana con una métrica de 0.80 de área bajo la curva de precisión-exhaustividad mediante la incorporación de datos variables meteorológicas, climatológicas y geográficas.

Fue posible entrenar este modelo con una ventana temporal suficiente que ayudó a generalizar el fenómeno de sequías a través del tiempo.

Los resultados de las predicciones del modelo complejo mostraron que la métrica deseada fue alcanzada exitosamente, y aunado a esto se realizaron experimentos que nos indica que tener sólo información de sequía tiene poder predictivo pero este se puede mejorar al incluir más variables predictoras.

Trabajo por hacer

El desempeño del modelo es lo mínimo esperado sin embargo se puede observar el potencial predictivo que tiene a través del tiempo. Para lograr hacer esto, todavía queda trabajo por hacer.

Se requiere:

- Incorporar variables de índole global para ayudar a tener un contexto generalizado. Por ejemplo, se planea incluir variables como el fenómeno de “El niño y La niña”.
- Incorporar más fuentes de datos para complementar la información (datos meteorológicos de la NASA).
- Reemplazar nuestro consumo de datos meteorológico (DarkSky dejará de operar en el 2022).
- Búsqueda de información más granular: en lugar de tener datos meteorológicos a nivel mensual, tener esta información cada día.
- Realizar un análisis más exhaustivo para encontrar la ventana de entrenamiento óptima.
- Realizar optimización de hiperparámetros correcta.
- Mejorar la creación de variables adicionales.
- Implementar este proyecto con predicciones o entrenamientos en otros países, como Estados Unidos o Canadá.

- Predicción certera de cambios repentinos en la presencia/ausencia de sequías y evaluaciones más rigurosas.

Referencias

1. **Arenas, Ivan. (2021).** *Sequía 2020-2021: La segunda más severa del registro reciente*. Centro de Ciencias de la Atmósfera. Recuperado el 22 de agosto del 2021 de: <https://www.atmosfera.unam.mx/sequia-2020-2021-la-segunda-mas-severa-del-registro-reciente/>
2. **Breiman, L., Friedman, J., Olshen, R., y Stone, J. (1984).** *Classification and Regression Trees*. Series: The Wadsworth Statistics/Probability Series. Chapman and Hall CRC. Boca Raton, London, New York, Washington D.C. ISBN: 0-412-04841-8.
3. **Centro para el Control y la Prevención de Enfermedades. (2019).** Las Sequías y su Salud. Centro Nacional de Salud Ambiental. Recuperado (en línea) el 21 de agosto del 2021 de: <https://www.cdc.gov/spanish/nceh/especiales/sequias/index.html>
4. **Chen T., y Guestrin, C. (2016)** *XGBoost: A Scalable Tree Boosting System*. Procedimientos de la 22ava ACM SIGKDD Conferencia Internacional en Descubrimiento y Minería de Datos. Páginas 685-794 DOI: <https://doi.org/10.1145/2939672.2939785>. Recuperado (en línea) el 30 de julio del 2021 de: <https://dl.acm.org/doi/10.1145/2939672.2939785>
5. **CONAGUA (2001).** *Cuencas Hidrológicas*. Escala 1:250000. México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
6. **CONAGUA (2018).** *Programa Nacional Contra las Sequías*. Acciones y Programas. México. Recuperado (en línea) el 21 de Septiembre del 2021 de: <https://www.gob.mx/conagua/acciones-y-programas/programa-nacional-contra-la-sequia-pronacose-programas-de-medidas-preventivas-y-de-mitigacion-a-la-sequia-pmpms-para-ciudades>
7. **CONABIO (2001).** *Subcuencas Hidrológicas*. Extraído de Boletín hidrológico (1970). Subcuencas hidrológicas en mapas de regiones hidrológicas. Escala más común 1:1000000. Secretaría de Recursos Hidráulicos, Jefatura de Irrigación y Control de Ríos, Dirección de Hidrología, México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
8. **CONABIO (2001a).** *Curvas de nivel para la República Mexicana*. Escala 1:250000 Extraído del Modelo Digital del Terreno. Instituto Nacional de Estadística Geografía e Informática (INEGI). México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
9. **Cutler, A., Cutler, D., y Stevens, J. (2008).** *Tree-based Methods*. Department of Mathematics and Statistics, High-Dimensional Data Analysis in Cancer Research. 1-19. Utah State University. doi: 10.1007/978-0-387-69765-9_5

10. **Esparza, Miguel.** (2014). *La sequía y la escasez de agua en México: Situación actual y perspectivas futuras*. Secuencia, (89), 193-219. Recuperado en 22 de agosto de 2021, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0186-03482014000200008&lng=es&tlng=es.
11. **Guevara, M., y Arroyo-Cruz, C.E** (2016). *Curvatura de perfil*. Escala 1:4000000. Edición 1. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. Ciudad de México, México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
12. **Guevara, M., y Arroyo-Cruz, C.E** (2016a). *Curvatura del plano*. Escala 1:4000000. Edición 1. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. Ciudad de México, México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
13. **Guevara, M., y Arroyo-Cruz, C.E** (2016b). *Índice topográfico de humedad*. Escala 1:4000000. Edición 1. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. Ciudad de México, México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
14. **Guevara, M., y Arroyo-Cruz, C.E** (2016c). *Pendiente*. Escala 1:4000000. Edición 1. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. Ciudad de México, México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
15. **Guevara, M., y Arroyo-Cruz, C.E** (2016a). *Curvatura del terreno*. Escala 1:4000000. Edición 1. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. Ciudad de México, México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
16. **Hao, Z., Singh, V. P., & Xia, Y.** (2018). *Seasonal drought prediction: Advances, challenges, and future prospects*. Reviews of Geophysics, 56, 108– 141. <https://doi.org/10.1002/2016RG000549>
17. **Hastie, T., Tibshirani, R., y Friedman, J.** (2009). *The elements of Statistical Learning, Data Mining, Inference, and Prediction*. Springer Series in Statistics, Second Edition.
18. **INIFAP, CONABIO** (2001) *Edafología*. Escalas 1:250000 y 1:1000000, México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
19. **LightGBM.** (2020). *Welcome to LightGBM's documentation!* Microsoft Corporation Revision e83042f2. Recuperado el 14 de agosto del 2021 de: <https://lightgbm.readthedocs.io/en/latest/index.html>
20. **Lobato-Sánchez, René.** (2021). *Perspectiva sobre la sequía actual en México*. Instituto Mexicano de Tecnología del Agua. Recuperado el 14 de agosto del 2021 de: <https://www.gob.mx/imta/articulos/perspectiva-sobre-la-sequia-actual-en-mexico?idiom=es>
21. **Lobato-Sánchez, René.** (2016). *El monitor de la sequía en México. Tecnología y ciencias del agua*, 7(5), 197-211. Recuperado en 22 de agosto de 2021, de

http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-24222016000500197&lng=es&tlng=es.

22. **Lugo-Hupb, J., Vidal Zepeda, R., Fernández-Equiarte, A., Gallegos-García, A., y Zavala-H., J.** (2001) *Hipsometría*. Extraído de Hipsometría y Batimetría, I.1.1. Atlas Nacional de México, Vol. I. Escala 1:4000000. Instituto de Geografía UNAM. México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
23. **Lundberg, S.** (2019). *SHAP*. Github. Recuperado el 14 de agosto del 2021 de: <https://github.com/slundberg/shap>.
24. **Lundberg, S., Erion, G., y Lee, S.** (2019). *Consistent Individualized Feature Attribution for Tree Ensembles*. University of Washington. ArXiv, Recuperado el 14 de agosto del 2021 de: <https://arxiv.org/abs/1802.03888v3>
25. **Maderey-R, L. E. y Torres-Ruata, C.** (1998) *Hidrografía*. Extraído de Hidrografía e hidrometría, IV.6.1 (A) Atlas Nacional de México. Vol. II. Escala 1: 4000000. Instituto de Geografía UNAM. México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
26. **Marín-C, S. y Torres-Ruata, C.** (2002) *Hidrogeología*. IV. 6.4 Atlas nacional de México Vol II Escala 1: 4000000 Instituto de Geografía UNAM, México. Recuperado (en línea) el 21 de agosto del 2021 de: http://www.conabio.gob.mx/informacion/gis/#dt_dlds
27. **NCAR.** (2021). *Palmer Drought Severity Index*. Climate Data. Climate Data Guide. Recuperado (en línea) el 21 de agosto del 2021 de: <https://climatedataguide.ucar.edu/climate-data/palmer-drought-severity-index-pdsi>
28. **NOAA.** (2021). *Outlooks and Forecasts. Data & Maps*. National Integrated Drought Information System. Recuperado (en línea) el 21 de agosto del 2021 de: <https://www.drought.gov/forecasts>
29. **Scikit-Learn.** (2019). *Ensemble Methods*. BSD Licence. Recuperado el 14 de agosto del 2021 de: <https://scikit-learn.org/stable/modules/ensemble.html#gradient-boosting>
30. **Shwartz-zIV, R. y Armon, A.** (2021) *Tabular Data: Deep Learning is Not All You Need*. IT AI Group, Intel. Recuperado (en línea) el 30 de julio del 2021 de: <https://arxiv.org/pdf/2106.03253.pdf>
31. **TWC.** (2021). *Keetch-Byram Drought Index*. Recuperado el 14 de agosto del 2021 de: <https://twc.tamu.edu/kbdi>
32. **University of Nebraska Lincoln.** (2021). *Drought Classification*. Recuperado (en línea) el 21 de agosto del 2021 de: <https://droughtmonitor.unl.edu/About/AboutData/DroughtClassification.aspx>
33. **World Bank** (2021). *Global Forecasting Tool*. International Research Institute for Climate and Society. Recuperado el 14 de agosto del 2021 de: https://iridl.ldeo.columbia.edu/maproom/Global/World_Bank/Drought_Monitor/index_3.html?gmap=%5B-56.37591133025235%2C-3.1117076884074386%2C3%5D
34. **World Meteorological Organization** (2016). *Handbook of Drought Indicators and Indices*. Recuperado el 28 de agosto del 2021 de: https://www.droughtmanagement.info/literature/GWP_Handbook_of_Drought_Indicators_and_Indices_2016.pdf

35. **XGBoost.** (2020). *Introduction to Boosted Trees*. Recuperado el 14 de agosto del 2021 de: <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>
36. **Zhang, Y., y Haghani, A.** (2015). *A gradient boosting method to improve travel time prediction*. *Transportation Research Part C: Emerging Technologies*, 58, 308–324. doi:10.1016/j.trc.2015.02.019

Anexos

Anexo 1

Selección de municipios

En un inicio, se quería obtener una muestra de 1200 municipios de los 2463 que son parte de la república mexicana. Esto con la finalidad de poder duplicar la cantidad de datos que teníamos la capacidad de obtener por medio de la API del clima.

Esta estrategia se llevó a cabo por medio de un algoritmo de clustering jerárquico, en el cual se tomaba en cuenta la distancia geoespacial entre los municipios (utilizando los datos de latitud y longitud), así como el comportamiento histórico de sequías en cada municipio.

Al final, se obtuvieron 1200 clusters optimizando estos dos parámetros para después seleccionar aleatoriamente un municipio dentro del cluster con el cual haríamos las llamadas de la API.

Sin embargo, al empezar con el proceso de la obtención de los datos meteorológicos nos dimos cuenta que muchos municipios para fechas específicas traían muy poca o nula información. Esto representaba un grave problema por lo que se decidió realizar un análisis más profundo para determinar cuáles eran los municipios afectados. Cabe mencionar que para fechas recientes (el año en curso) todos los municipios presentaban la información completa.

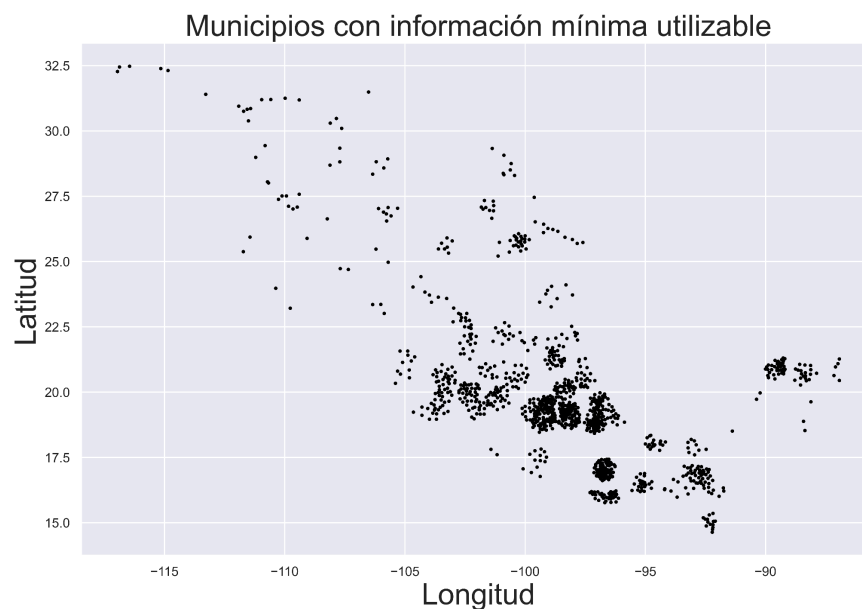
En primer lugar, se escogieron fechas específicas para realizar las llamadas de los 2463 municipios:

- 2016-01-31
- 2014-01-31
- 2012-01-31
- 2010-01-31
- 2008-01-31
- 2006-01-31

Con cada una de las llamadas se obtenía un archivo .json el cual contiene toda la información de las variables meteorológicas. Lo que se hizo a continuación fue analizar cómo fluctúa el

tamaño de los archivos de cada municipio en las diferentes fechas, y se definió un umbral de información mínima utilizable de 300 bytes.

A partir de este análisis se encontraron 1212 municipios que, desde el 2006, sobrepasan este umbral, por lo que se decidió trabajar con ellos. La distribución geográfica de dichos municipios se muestra en la figura #.



Anexo 2

Calidad de los datos de la variable de respuesta

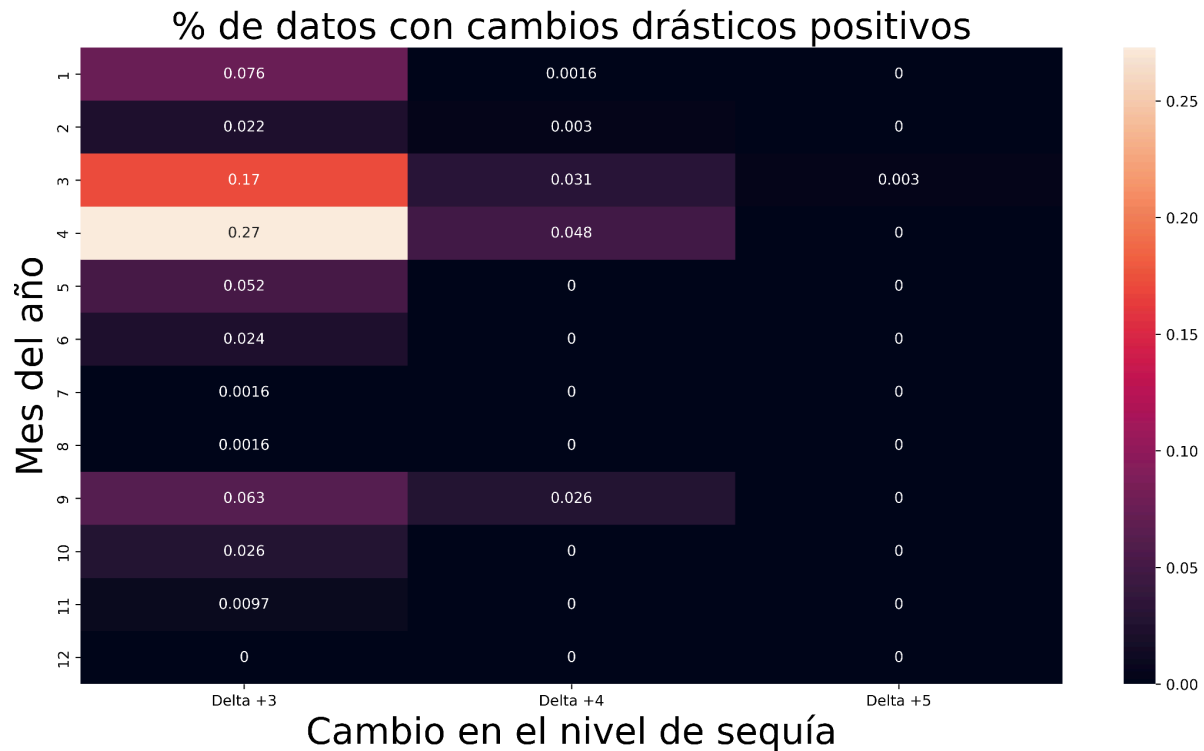
Para verificar la veracidad de los datos sobre el nivel de sequía para cada municipio, se realizó un análisis basado en los cambios de estos niveles a través del tiempo para cada una de las entidades.

La premisa principal fue: no deberían de haber muchos datos con cambios abruptos en nivel de sequía de una medición a otra. Esto principalmente se debe cumplir para cambios positivos en la variable de respuesta, esto quiere decir un cambio de “no sequía” a “sequía” muy abrupto. Mientras que es más probable que existan cambios negativos pronunciados debido a que una lluvia puede acabar con la sequía rápidamente, lo cuál, si cierto, debería mostrarse con cierta estacionalidad.

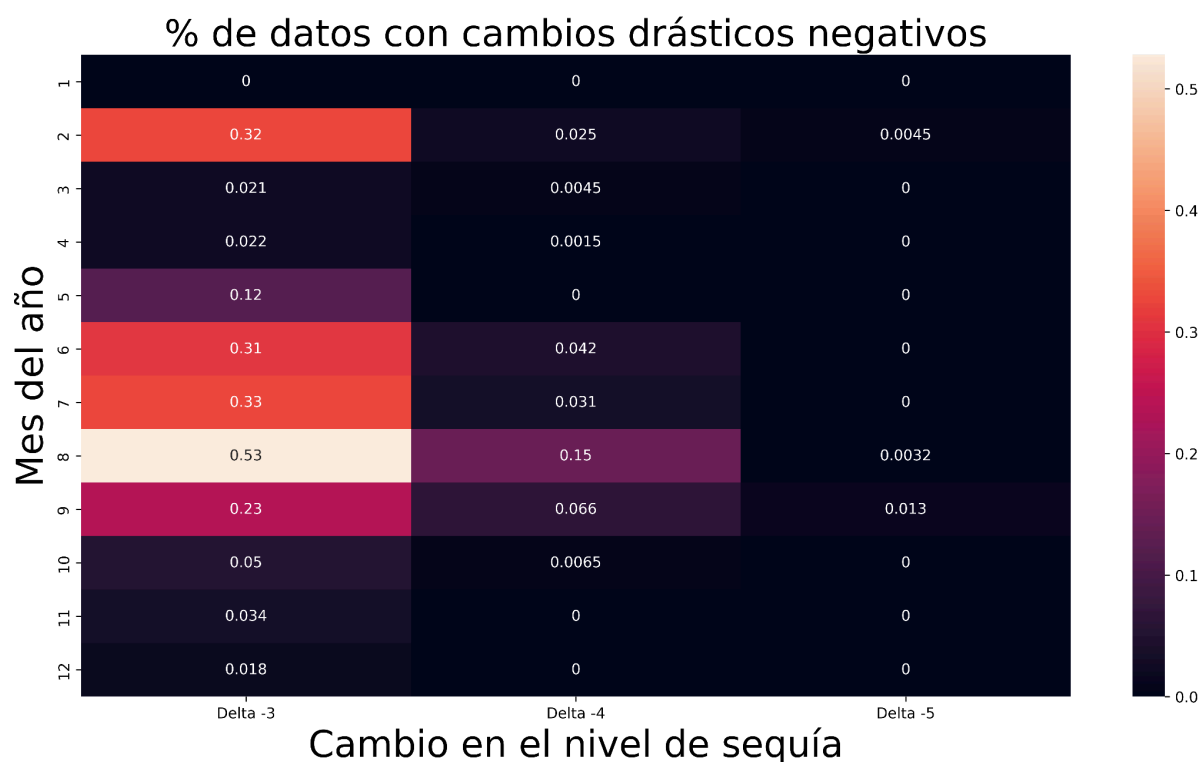
Para ello, una vez obtenidas las deltas para cada municipio, se elaboraron unos mapas de calor en donde se puede apreciar el porcentaje de datos que presentan este tipo de cambios drásticos con respecto a la totalidad de los datos disponibles.

En la figura #, se muestran el porcentaje de datos con delta mayor a 3 por mes. De esta figura podemos apreciar patrones de estacionalidad con un pico importante en el mes de abril. Sin

embargo, este pico solo representa el 0.27% de todos los datos que tenemos para ese mes, y la estacionalidad nos da indicios de que el comportamiento es correcto. Además, podemos notar como bajan drásticamente los casos para delta mayor a 4 y delta mayor a 5, que de igual forma es el comportamiento esperado.



Por otra parte, la figura # nos muestra los porcentajes de datos con deltas negativas. De la misma manera que en la figura pasada podemos notar cierta estacionalidad que coincide con el temporal de lluvias, teniendo picos importantes en Junio, Julio y Agosto. Así mismo, el pico más alto se mantiene en 0.53% de los datos totales y vemos una disminución en este tipo de casos con deltas menores a -4 y a -5.



Analizando estos resultados y tomando en cuenta:

- La baja recurrencia de este tipo de datos
- La estacionalidad aparente
- La disminución de casos con el aumento del valor absoluto de la delta

Podemos tener cierta confianza en que los datos de la variable de respuesta no presentan situaciones que pueden poner en riesgo el desarrollo del modelo de aprendizaje automático.

Anexo 3

Variables de consumo para predecir la sequía el siguiente mes.

Variables para el modelo simple:

- 'presencia_sequia',
- 'nivel_numerico_ultimos180_dias_avg',
- 'nivel_numerico_ultimos180_dias_std',
- 'presencia_sequia_ultimos180_dias_avg',
- 'month_of_year',
- 'week_of_year',
- 'Lon',
- 'Lat',
- 'Lon_euclidean',
- 'Lat_euclidean',

Variables para el modelo complejo:

- 'Lon'
- 'Lat'
- 'Lon_euclidean'
- 'Lat_euclidean'
- 'presencia_sequia'
- 'loc_count'
- 'nivel_numerico_ultimos180_dias_avg'
- 'nivel_numerico_ultimos180_dias_std'
- 'nivel_numerico_ultimos180_dias_max'
- 'nivel_numerico_ultimos180_dias_min'
- 'presencia_sequia_ultimos180_dias_avg'
- 'neigh_mun_nivel_numerico_std'
- 'neigh_mun_nivel_numerico_min'
- 'neigh_mun_nivel_numerico_max'
- 'neigh_mun_nivel_numerico_med'
- 'neigh_mun_presencia_sequia_proportion'
- 'neigh_mun_nivel_numerico_ultimos180_dias_avg_avg'
- 'neigh_mun_nivel_numerico_ultimos180_dias_avg_std'
- 'neigh_mun_nivel_numerico_ultimos180_dias_std_avg'
- 'neigh_mun_nivel_numerico_ultimos180_dias_std_std'
- 'tex_edafologia_avg'
- 'tex_edafologia_std'
- 'tex_edafologia_min'
- 'tex_edafologia_max'
- 'tex_edafologia_kur'
- 'tex_edafologia_skew'
- 'fasfis_edafologia_avg'
- 'fasfis_edafologia_std'
- 'fasfis_edafologia_min'
- 'fasfis_edafologia_max'
- 'fasfis_edafologia_kur'
- 'fasfis_edafologia_skew'
- 'fasquim_edafologia_avg'
- 'fasquim_edafologia_std'
- 'fasquim_edafologia_min'
- 'fasquim_edafologia_max'
- 'fasquim_edafologia_kur'
- 'fasquim_edafologia_skew'
- 'descripcio_edafologia_mode'
- 'descripcio_edafologia_prop'
- 'descripcio_edafologia_diff'
- 'desc_tex_edafologia_mode'
- 'desc_tex_edafologia_prop'

- 'desc_tex_edafologia_diff'
- 'desc_fasfi_edafologia_mode'
- 'desc_fasfi_edafologia_prop'
- 'desc_fasfi_edafologia_diff'
- 'desc_fasqui_edafologia_mode'
- 'desc_fasqui_edafologia_prop'
- 'desc_fasqui_edafologia_diff'
- 'descripcion_hidrogeologia_mode'
- 'descripcion_hidrogeologia_prop'
- 'descripcion_hidrogeologia_diff'
- 'indice_topografico_de_humedad_avg'
- 'indice_topografico_de_humedad_std'
- 'indice_topografico_de_humedad_min'
- 'indice_topografico_de_humedad_max'
- 'indice_topografico_de_humedad_kur'
- 'indice_topografico_de_humedad_skew'
- 'curvatura_del_perfil_avg'
- 'curvatura_del_perfil_std'
- 'curvatura_del_perfil_min'
- 'curvatura_del_perfil_max'
- 'curvatura_del_perfil_kur'
- 'curvatura_del_perfil_skew'
- 'curvatura_del_perfil_neg_prop'
- 'curvatura_del_perfil_pos_prop'
- 'curvatura_del_plano_avg'
- 'curvatura_del_plano_std'
- 'curvatura_del_plano_min'
- 'curvatura_del_plano_max'
- 'curvatura_del_plano_kur'
- 'curvatura_del_plano_skew'
- 'curvatura_del_plano_neg_prop'
- 'curvatura_del_plano_pos_prop'
- 'curvatura_del_terreno_avg'
- 'curvatura_del_terreno_std'
- 'curvatura_del_terreno_min'
- 'curvatura_del_terreno_max'
- 'curvatura_del_terreno_kur'
- 'curvatura_del_terreno_skew'
- 'curvatura_del_terreno_neg_prop'
- 'curvatura_del_terreno_pos_prop'
- 'pendiente_avg'
- 'pendiente_std'
- 'pendiente_min'
- 'pendiente_max'

- 'pendiente_kur'
- 'pendiente_skew'
- 'hipsometria_numeric_avg'
- 'hipsometria_numeric_std'
- 'hipsometria_numeric_min'
- 'hipsometria_numeric_max'
- 'hipsometria_numeric_kur'
- 'hipsometria_numeric_skew'
- 'cuenca_mode'
- 'cuenca_prop'
- 'cuenca_diff'
- 'subcuenca_mode'
- 'subcuenca_prop'
- 'subcuenca_diff'
- 'total_count'
- 'dailymoonPhase'
- 'dailytemperatureHigh'
- 'dailytemperatureLow'
- 'dailycloudCover'
- 'dailyuvIndex'
- 'dailytemperatureMin'
- 'dailytemperatureMax'
- 'dailydewPoint'
- 'dailyhumidity'
- 'dailyicon'
- 'dailyprecipIntensity'
- 'dailyprecipIntensityMax'
- 'dailyprecipProbability'
- 'dailyozone'
- 'dailysun_avaialble'
- 'dailytime_between_min_max_temp'
- 'dailytime_between_low_high_temp'
- 'dailytime_between_sunrise_and_max_temp'
- 'dailytime_between_sunset_and_max_temp'
- 'dailytime_between_sunrise_and_min_temp'
- 'dailytime_between_sunset_and_min_temp'
- 'dailytime_between_sunrise_and_high_temp'
- 'dailytime_between_sunset_and_high_temp'
- 'dailytime_between_sunrise_and_low_temp'
- 'dailytime_between_sunset_and_low_temp'
- 'pressure_mean_ultimos180_dias_avg'
- 'pressure_mean_ultimos180_dias_std'
- 'pressure_mean_ultimos180_dias_max'
- 'pressure_mean_ultimos180_dias_min'

- 'pressure_mean_ultimos180_dias_median'
- 'pressure_mean_ultimos180_dias_median_avg'
- 'pressure_mean_ultimos180_dias_range'
- 'pressure_std_ultimos180_dias_avg'
- 'pressure_std_ultimos180_dias_std'
- 'pressure_std_ultimos180_dias_max'
- 'pressure_std_ultimos180_dias_min'
- 'pressure_std_ultimos180_dias_median'
- 'pressure_std_ultimos180_dias_median_avg'
- 'pressure_std_ultimos180_dias_range'
- 'pressure_min_ultimos180_dias_avg'
- 'pressure_min_ultimos180_dias_std'
- 'pressure_min_ultimos180_dias_max'
- 'pressure_min_ultimos180_dias_min'
- 'pressure_min_ultimos180_dias_median'
- 'pressure_min_ultimos180_dias_median_avg'
- 'pressure_min_ultimos180_dias_range'
- 'pressure_max_ultimos180_dias_avg'
- 'pressure_max_ultimos180_dias_std'
- 'pressure_max_ultimos180_dias_max'
- 'pressure_max_ultimos180_dias_min'
- 'pressure_max_ultimos180_dias_median'
- 'pressure_max_ultimos180_dias_median_avg'
- 'pressure_max_ultimos180_dias_range'
- 'pressure_median_ultimos180_dias_avg'
- 'pressure_median_ultimos180_dias_std'
- 'pressure_median_ultimos180_dias_max'
- 'pressure_median_ultimos180_dias_min'
- 'pressure_median_ultimos180_dias_median'
- 'pressure_median_ultimos180_dias_median_avg'
- 'pressure_median_ultimos180_dias_range'
- 'pressure_mad_ultimos180_dias_avg'
- 'pressure_mad_ultimos180_dias_std'
- 'pressure_mad_ultimos180_dias_max'
- 'pressure_mad_ultimos180_dias_min'
- 'pressure_mad_ultimos180_dias_median'
- 'pressure_mad_ultimos180_dias_median_avg'
- 'pressure_mad_ultimos180_dias_range'
- 'pressure_count_ultimos180_dias_avg'
- 'pressure_count_ultimos180_dias_std'
- 'pressure_count_ultimos180_dias_max'
- 'pressure_count_ultimos180_dias_min'
- 'pressure_count_ultimos180_dias_median'
- 'pressure_count_ultimos180_dias_median_avg'

- 'pressure_count_ultimos180_dias_range'
- 'humidity_mean_ultimos180_dias_avg'
- 'humidity_mean_ultimos180_dias_std'
- 'humidity_mean_ultimos180_dias_max'
- 'humidity_mean_ultimos180_dias_min'
- 'humidity_mean_ultimos180_dias_median'
- 'humidity_mean_ultimos180_dias_median_avg'
- 'humidity_mean_ultimos180_dias_range'
- 'humidity_std_ultimos180_dias_avg'
- 'humidity_std_ultimos180_dias_std'
- 'humidity_std_ultimos180_dias_max'
- 'humidity_std_ultimos180_dias_min'
- 'humidity_std_ultimos180_dias_median'
- 'humidity_std_ultimos180_dias_median_avg'
- 'humidity_std_ultimos180_dias_range'
- 'humidity_min_ultimos180_dias_avg'
- 'humidity_min_ultimos180_dias_std'
- 'humidity_min_ultimos180_dias_max'
- 'humidity_min_ultimos180_dias_min'
- 'humidity_min_ultimos180_dias_median'
- 'humidity_min_ultimos180_dias_median_avg'
- 'humidity_min_ultimos180_dias_range'
- 'humidity_max_ultimos180_dias_avg'
- 'humidity_max_ultimos180_dias_std'
- 'humidity_max_ultimos180_dias_max'
- 'humidity_max_ultimos180_dias_min'
- 'humidity_max_ultimos180_dias_median'
- 'humidity_max_ultimos180_dias_median_avg'
- 'humidity_max_ultimos180_dias_range'
- 'humidity_median_ultimos180_dias_avg'
- 'humidity_median_ultimos180_dias_std'
- 'humidity_median_ultimos180_dias_max'
- 'humidity_median_ultimos180_dias_min'
- 'humidity_median_ultimos180_dias_median'
- 'humidity_median_ultimos180_dias_median_avg'
- 'humidity_median_ultimos180_dias_range'
- 'humidity_mad_ultimos180_dias_avg'
- 'humidity_mad_ultimos180_dias_std'
- 'humidity_mad_ultimos180_dias_max'
- 'humidity_mad_ultimos180_dias_min'
- 'humidity_mad_ultimos180_dias_median'
- 'humidity_mad_ultimos180_dias_median_avg'
- 'humidity_mad_ultimos180_dias_range'
- 'humidity_count_ultimos180_dias_avg'

- 'humidity_count_ultimos180_dias_std'
- 'humidity_count_ultimos180_dias_max'
- 'humidity_count_ultimos180_dias_min'
- 'humidity_count_ultimos180_dias_median'
- 'humidity_count_ultimos180_dias_median_avg'
- 'humidity_count_ultimos180_dias_range'
- 'dewPoint_mean_ultimos180_dias_avg'
- 'dewPoint_mean_ultimos180_dias_std'
- 'dewPoint_mean_ultimos180_dias_max'
- 'dewPoint_mean_ultimos180_dias_min'
- 'dewPoint_mean_ultimos180_dias_median'
- 'dewPoint_mean_ultimos180_dias_median_avg'
- 'dewPoint_mean_ultimos180_dias_range'
- 'dewPoint_std_ultimos180_dias_avg'
- 'dewPoint_std_ultimos180_dias_std'
- 'dewPoint_std_ultimos180_dias_max'
- 'dewPoint_std_ultimos180_dias_min'
- 'dewPoint_std_ultimos180_dias_median'
- 'dewPoint_std_ultimos180_dias_median_avg'
- 'dewPoint_std_ultimos180_dias_range'
- 'dewPoint_min_ultimos180_dias_avg'
- 'dewPoint_min_ultimos180_dias_std'
- 'dewPoint_min_ultimos180_dias_max'
- 'dewPoint_min_ultimos180_dias_min'
- 'dewPoint_min_ultimos180_dias_median'
- 'dewPoint_min_ultimos180_dias_median_avg'
- 'dewPoint_min_ultimos180_dias_range'
- 'dewPoint_max_ultimos180_dias_avg'
- 'dewPoint_max_ultimos180_dias_std'
- 'dewPoint_max_ultimos180_dias_max'
- 'dewPoint_max_ultimos180_dias_min'
- 'dewPoint_max_ultimos180_dias_median'
- 'dewPoint_max_ultimos180_dias_median_avg'
- 'dewPoint_max_ultimos180_dias_range'
- 'dewPoint_median_ultimos180_dias_avg'
- 'dewPoint_median_ultimos180_dias_std'
- 'dewPoint_median_ultimos180_dias_max'
- 'dewPoint_median_ultimos180_dias_min'
- 'dewPoint_median_ultimos180_dias_median'
- 'dewPoint_median_ultimos180_dias_median_avg'
- 'dewPoint_median_ultimos180_dias_range'
- 'dewPoint_mad_ultimos180_dias_avg'
- 'dewPoint_mad_ultimos180_dias_std'
- 'dewPoint_mad_ultimos180_dias_max'

- 'dewPoint_mad_ultimos180_dias_min'
- 'dewPoint_mad_ultimos180_dias_median'
- 'dewPoint_mad_ultimos180_dias_median_avg'
- 'dewPoint_mad_ultimos180_dias_range'
- 'dewPoint_count_ultimos180_dias_avg'
- 'dewPoint_count_ultimos180_dias_std'
- 'dewPoint_count_ultimos180_dias_max'
- 'dewPoint_count_ultimos180_dias_min'
- 'dewPoint_count_ultimos180_dias_median'
- 'dewPoint_count_ultimos180_dias_median_avg'
- 'dewPoint_count_ultimos180_dias_range'
- 'temperature_mean_ultimos180_dias_avg'
- 'temperature_mean_ultimos180_dias_std'
- 'temperature_mean_ultimos180_dias_max'
- 'temperature_mean_ultimos180_dias_min'
- 'temperature_mean_ultimos180_dias_median'
- 'temperature_mean_ultimos180_dias_median_avg'
- 'temperature_mean_ultimos180_dias_range'
- 'temperature_std_ultimos180_dias_avg'
- 'temperature_std_ultimos180_dias_std'
- 'temperature_std_ultimos180_dias_max'
- 'temperature_std_ultimos180_dias_min'
- 'temperature_std_ultimos180_dias_median'
- 'temperature_std_ultimos180_dias_median_avg'
- 'temperature_std_ultimos180_dias_range'
- 'temperature_min_ultimos180_dias_avg'
- 'temperature_min_ultimos180_dias_std'
- 'temperature_min_ultimos180_dias_max'
- 'temperature_min_ultimos180_dias_min'
- 'temperature_min_ultimos180_dias_median'
- 'temperature_min_ultimos180_dias_median_avg'
- 'temperature_min_ultimos180_dias_range'
- 'temperature_max_ultimos180_dias_avg'
- 'temperature_max_ultimos180_dias_std'
- 'temperature_max_ultimos180_dias_max'
- 'temperature_max_ultimos180_dias_min'
- 'temperature_max_ultimos180_dias_median'
- 'temperature_max_ultimos180_dias_median_avg'
- 'temperature_max_ultimos180_dias_range'
- 'temperature_median_ultimos180_dias_avg'
- 'temperature_median_ultimos180_dias_std'
- 'temperature_median_ultimos180_dias_max'
- 'temperature_median_ultimos180_dias_min'
- 'temperature_median_ultimos180_dias_median'

- 'temperature_median_ultimos180_dias_median_avg'
- 'temperature_median_ultimos180_dias_range'
- 'temperature_mad_ultimos180_dias_avg'
- 'temperature_mad_ultimos180_dias_std'
- 'temperature_mad_ultimos180_dias_max'
- 'temperature_mad_ultimos180_dias_min'
- 'temperature_mad_ultimos180_dias_median'
- 'temperature_mad_ultimos180_dias_median_avg'
- 'temperature_mad_ultimos180_dias_range'
- 'temperature_count_ultimos180_dias_avg'
- 'temperature_count_ultimos180_dias_std'
- 'temperature_count_ultimos180_dias_max'
- 'temperature_count_ultimos180_dias_min'
- 'temperature_count_ultimos180_dias_median'
- 'temperature_count_ultimos180_dias_median_avg'
- 'temperature_count_ultimos180_dias_range'
- 'uvIndex_mean_ultimos180_dias_avg'
- 'uvIndex_mean_ultimos180_dias_std'
- 'uvIndex_mean_ultimos180_dias_max'
- 'uvIndex_mean_ultimos180_dias_min'
- 'uvIndex_mean_ultimos180_dias_median'
- 'uvIndex_mean_ultimos180_dias_median_avg'
- 'uvIndex_mean_ultimos180_dias_range'
- 'uvIndex_std_ultimos180_dias_avg'
- 'uvIndex_std_ultimos180_dias_std'
- 'uvIndex_std_ultimos180_dias_max'
- 'uvIndex_std_ultimos180_dias_min'
- 'uvIndex_std_ultimos180_dias_median'
- 'uvIndex_std_ultimos180_dias_median_avg'
- 'uvIndex_std_ultimos180_dias_range'
- 'uvIndex_min_ultimos180_dias_avg'
- 'uvIndex_min_ultimos180_dias_std'
- 'uvIndex_min_ultimos180_dias_max'
- 'uvIndex_min_ultimos180_dias_min'
- 'uvIndex_min_ultimos180_dias_median'
- 'uvIndex_min_ultimos180_dias_median_avg'
- 'uvIndex_min_ultimos180_dias_range'
- 'uvIndex_max_ultimos180_dias_avg'
- 'uvIndex_max_ultimos180_dias_std'
- 'uvIndex_max_ultimos180_dias_max'
- 'uvIndex_max_ultimos180_dias_min'
- 'uvIndex_max_ultimos180_dias_median'
- 'uvIndex_max_ultimos180_dias_median_avg'
- 'uvIndex_max_ultimos180_dias_range'

- 'uvIndex_median_ultimos180_dias_avg'
- 'uvIndex_median_ultimos180_dias_std'
- 'uvIndex_median_ultimos180_dias_max'
- 'uvIndex_median_ultimos180_dias_min'
- 'uvIndex_median_ultimos180_dias_median'
- 'uvIndex_median_ultimos180_dias_median_avg'
- 'uvIndex_median_ultimos180_dias_range'
- 'uvIndex_mad_ultimos180_dias_avg'
- 'uvIndex_mad_ultimos180_dias_std'
- 'uvIndex_mad_ultimos180_dias_max'
- 'uvIndex_mad_ultimos180_dias_min'
- 'uvIndex_mad_ultimos180_dias_median'
- 'uvIndex_mad_ultimos180_dias_median_avg'
- 'uvIndex_mad_ultimos180_dias_range'
- 'uvIndex_count_ultimos180_dias_avg'
- 'uvIndex_count_ultimos180_dias_std'
- 'uvIndex_count_ultimos180_dias_max'
- 'uvIndex_count_ultimos180_dias_min'
- 'uvIndex_count_ultimos180_dias_median'
- 'uvIndex_count_ultimos180_dias_median_avg'
- 'uvIndex_count_ultimos180_dias_range'
- 'cloudCover_mean_ultimos180_dias_avg'
- 'cloudCover_mean_ultimos180_dias_std'
- 'cloudCover_mean_ultimos180_dias_max'
- 'cloudCover_mean_ultimos180_dias_min'
- 'cloudCover_mean_ultimos180_dias_median'
- 'cloudCover_mean_ultimos180_dias_median_avg'
- 'cloudCover_mean_ultimos180_dias_range'
- 'cloudCover_std_ultimos180_dias_avg'
- 'cloudCover_std_ultimos180_dias_std'
- 'cloudCover_std_ultimos180_dias_max'
- 'cloudCover_std_ultimos180_dias_min'
- 'cloudCover_std_ultimos180_dias_median'
- 'cloudCover_std_ultimos180_dias_median_avg'
- 'cloudCover_std_ultimos180_dias_range'
- 'cloudCover_min_ultimos180_dias_avg'
- 'cloudCover_min_ultimos180_dias_std'
- 'cloudCover_min_ultimos180_dias_max'
- 'cloudCover_min_ultimos180_dias_min'
- 'cloudCover_min_ultimos180_dias_median'
- 'cloudCover_min_ultimos180_dias_median_avg'
- 'cloudCover_min_ultimos180_dias_range'
- 'cloudCover_max_ultimos180_dias_avg'
- 'cloudCover_max_ultimos180_dias_std'

- 'cloudCover_max_ultimos180_dias_max'
- 'cloudCover_max_ultimos180_dias_min'
- 'cloudCover_max_ultimos180_dias_median'
- 'cloudCover_max_ultimos180_dias_median_avg'
- 'cloudCover_max_ultimos180_dias_range'
- 'cloudCover_median_ultimos180_dias_avg'
- 'cloudCover_median_ultimos180_dias_std'
- 'cloudCover_median_ultimos180_dias_max'
- 'cloudCover_median_ultimos180_dias_min'
- 'cloudCover_median_ultimos180_dias_median'
- 'cloudCover_median_ultimos180_dias_median_avg'
- 'cloudCover_median_ultimos180_dias_range'
- 'cloudCover_mad_ultimos180_dias_avg'
- 'cloudCover_mad_ultimos180_dias_std'
- 'cloudCover_mad_ultimos180_dias_max'
- 'cloudCover_mad_ultimos180_dias_min'
- 'cloudCover_mad_ultimos180_dias_median'
- 'cloudCover_mad_ultimos180_dias_median_avg'
- 'cloudCover_mad_ultimos180_dias_range'
- 'cloudCover_count_ultimos180_dias_avg'
- 'cloudCover_count_ultimos180_dias_std'
- 'cloudCover_count_ultimos180_dias_max'
- 'cloudCover_count_ultimos180_dias_min'
- 'cloudCover_count_ultimos180_dias_median'
- 'cloudCover_count_ultimos180_dias_median_avg'
- 'cloudCover_count_ultimos180_dias_range'
- 'precipIntensity_mean_ultimos180_dias_avg'
- 'precipIntensity_mean_ultimos180_dias_std'
- 'precipIntensity_mean_ultimos180_dias_max'
- 'precipIntensity_mean_ultimos180_dias_min'
- 'precipIntensity_mean_ultimos180_dias_median'
- 'precipIntensity_mean_ultimos180_dias_median_avg'
- 'precipIntensity_mean_ultimos180_dias_range'
- 'precipIntensity_std_ultimos180_dias_avg'
- 'precipIntensity_std_ultimos180_dias_std'
- 'precipIntensity_std_ultimos180_dias_max'
- 'precipIntensity_std_ultimos180_dias_min'
- 'precipIntensity_std_ultimos180_dias_median'
- 'precipIntensity_std_ultimos180_dias_median_avg'
- 'precipIntensity_std_ultimos180_dias_range'
- 'precipIntensity_min_ultimos180_dias_avg'
- 'precipIntensity_min_ultimos180_dias_std'
- 'precipIntensity_min_ultimos180_dias_max'
- 'precipIntensity_min_ultimos180_dias_min'

- 'precipIntensity_min_ultimos180_dias_median'
- 'precipIntensity_min_ultimos180_dias_median_avg'
- 'precipIntensity_min_ultimos180_dias_range'
- 'precipIntensity_max_ultimos180_dias_avg'
- 'precipIntensity_max_ultimos180_dias_std'
- 'precipIntensity_max_ultimos180_dias_max'
- 'precipIntensity_max_ultimos180_dias_min'
- 'precipIntensity_max_ultimos180_dias_median'
- 'precipIntensity_max_ultimos180_dias_median_avg'
- 'precipIntensity_max_ultimos180_dias_range'
- 'precipIntensity_median_ultimos180_dias_avg'
- 'precipIntensity_median_ultimos180_dias_std'
- 'precipIntensity_median_ultimos180_dias_max'
- 'precipIntensity_median_ultimos180_dias_min'
- 'precipIntensity_median_ultimos180_dias_median'
- 'precipIntensity_median_ultimos180_dias_median_avg'
- 'precipIntensity_median_ultimos180_dias_range'
- 'precipIntensity_mad_ultimos180_dias_avg'
- 'precipIntensity_mad_ultimos180_dias_std'
- 'precipIntensity_mad_ultimos180_dias_max'
- 'precipIntensity_mad_ultimos180_dias_min'
- 'precipIntensity_mad_ultimos180_dias_median'
- 'precipIntensity_mad_ultimos180_dias_median_avg'
- 'precipIntensity_mad_ultimos180_dias_range'
- 'precipIntensity_count_ultimos180_dias_avg'
- 'precipIntensity_count_ultimos180_dias_std'
- 'precipIntensity_count_ultimos180_dias_max'
- 'precipIntensity_count_ultimos180_dias_min'
- 'precipIntensity_count_ultimos180_dias_median'
- 'precipIntensity_count_ultimos180_dias_median_avg'
- 'precipIntensity_count_ultimos180_dias_range'
- 'precipProbability_mean_ultimos180_dias_avg'
- 'precipProbability_mean_ultimos180_dias_std'
- 'precipProbability_mean_ultimos180_dias_max'
- 'precipProbability_mean_ultimos180_dias_min'
- 'precipProbability_mean_ultimos180_dias_median'
- 'precipProbability_mean_ultimos180_dias_median_avg'
- 'precipProbability_mean_ultimos180_dias_range'
- 'precipProbability_std_ultimos180_dias_avg'
- 'precipProbability_std_ultimos180_dias_std'
- 'precipProbability_std_ultimos180_dias_max'
- 'precipProbability_std_ultimos180_dias_min'
- 'precipProbability_std_ultimos180_dias_median'
- 'precipProbability_std_ultimos180_dias_median_avg'

- 'precipProbability_std_ultimos180_dias_range'
- 'precipProbability_min_ultimos180_dias_avg'
- 'precipProbability_min_ultimos180_dias_std'
- 'precipProbability_min_ultimos180_dias_max'
- 'precipProbability_min_ultimos180_dias_min'
- 'precipProbability_min_ultimos180_dias_median'
- 'precipProbability_min_ultimos180_dias_median_avg'
- 'precipProbability_min_ultimos180_dias_range'
- 'precipProbability_max_ultimos180_dias_avg'
- 'precipProbability_max_ultimos180_dias_std'
- 'precipProbability_max_ultimos180_dias_max'
- 'precipProbability_max_ultimos180_dias_min'
- 'precipProbability_max_ultimos180_dias_median'
- 'precipProbability_max_ultimos180_dias_median_avg'
- 'precipProbability_max_ultimos180_dias_range'
- 'precipProbability_median_ultimos180_dias_avg'
- 'precipProbability_median_ultimos180_dias_std'
- 'precipProbability_median_ultimos180_dias_max'
- 'precipProbability_median_ultimos180_dias_min'
- 'precipProbability_median_ultimos180_dias_median'
- 'precipProbability_median_ultimos180_dias_median_avg'
- 'precipProbability_median_ultimos180_dias_range'
- 'precipProbability_mad_ultimos180_dias_avg'
- 'precipProbability_mad_ultimos180_dias_std'
- 'precipProbability_mad_ultimos180_dias_max'
- 'precipProbability_mad_ultimos180_dias_min'
- 'precipProbability_mad_ultimos180_dias_median'
- 'precipProbability_mad_ultimos180_dias_median_avg'
- 'precipProbability_mad_ultimos180_dias_range'
- 'precipProbability_count_ultimos180_dias_avg'
- 'precipProbability_count_ultimos180_dias_std'
- 'precipProbability_count_ultimos180_dias_max'
- 'precipProbability_count_ultimos180_dias_min'
- 'precipProbability_count_ultimos180_dias_median'
- 'precipProbability_count_ultimos180_dias_median_avg'
- 'precipProbability_count_ultimos180_dias_range'
- 'currently_cloudCover_ultimos180_dias_avg'
- 'currently_cloudCover_ultimos180_dias_std'
- 'currently_cloudCover_ultimos180_dias_max'
- 'currently_cloudCover_ultimos180_dias_min'
- 'currently_cloudCover_ultimos180_dias_median'
- 'currently_cloudCover_ultimos180_dias_median_avg'
- 'currently_cloudCover_ultimos180_dias_range'
- 'currently_uvIndex_ultimos180_dias_avg'

- 'currently_uvIndex_ultimos180_dias_std'
- 'currently_uvIndex_ultimos180_dias_max'
- 'currently_uvIndex_ultimos180_dias_min'
- 'currently_uvIndex_ultimos180_dias_median'
- 'currently_uvIndex_ultimos180_dias_median_avg'
- 'currently_uvIndex_ultimos180_dias_range'
- 'currently_temperature_ultimos180_dias_avg'
- 'currently_temperature_ultimos180_dias_std'
- 'currently_temperature_ultimos180_dias_max'
- 'currently_temperature_ultimos180_dias_min'
- 'currently_temperature_ultimos180_dias_median'
- 'currently_temperature_ultimos180_dias_median_avg'
- 'currently_temperature_ultimos180_dias_range'
- 'currently_dewPoint_ultimos180_dias_avg'
- 'currently_dewPoint_ultimos180_dias_std'
- 'currently_dewPoint_ultimos180_dias_max'
- 'currently_dewPoint_ultimos180_dias_min'
- 'currently_dewPoint_ultimos180_dias_median'
- 'currently_dewPoint_ultimos180_dias_median_avg'
- 'currently_dewPoint_ultimos180_dias_range'
- 'currently_humidity_ultimos180_dias_avg'
- 'currently_humidity_ultimos180_dias_std'
- 'currently_humidity_ultimos180_dias_max'
- 'currently_humidity_ultimos180_dias_min'
- 'currently_humidity_ultimos180_dias_median'
- 'currently_humidity_ultimos180_dias_median_avg'
- 'currently_humidity_ultimos180_dias_range'
- 'currently_pressure_ultimos180_dias_avg'
- 'currently_pressure_ultimos180_dias_std'
- 'currently_pressure_ultimos180_dias_max'
- 'currently_pressure_ultimos180_dias_min'
- 'currently_pressure_ultimos180_dias_median'
- 'currently_pressure_ultimos180_dias_median_avg'
- 'currently_pressure_ultimos180_dias_range'
- 'currently_precipIntensity_ultimos180_dias_avg'
- 'currently_precipIntensity_ultimos180_dias_std'
- 'currently_precipIntensity_ultimos180_dias_max'
- 'currently_precipIntensity_ultimos180_dias_min'
- 'currently_precipIntensity_ultimos180_dias_median'
- 'currently_precipIntensity_ultimos180_dias_median_avg'
- 'currently_precipIntensity_ultimos180_dias_range'
- 'currently_precipProbability_ultimos180_dias_avg'
- 'currently_precipProbability_ultimos180_dias_std'
- 'currently_precipProbability_ultimos180_dias_max'

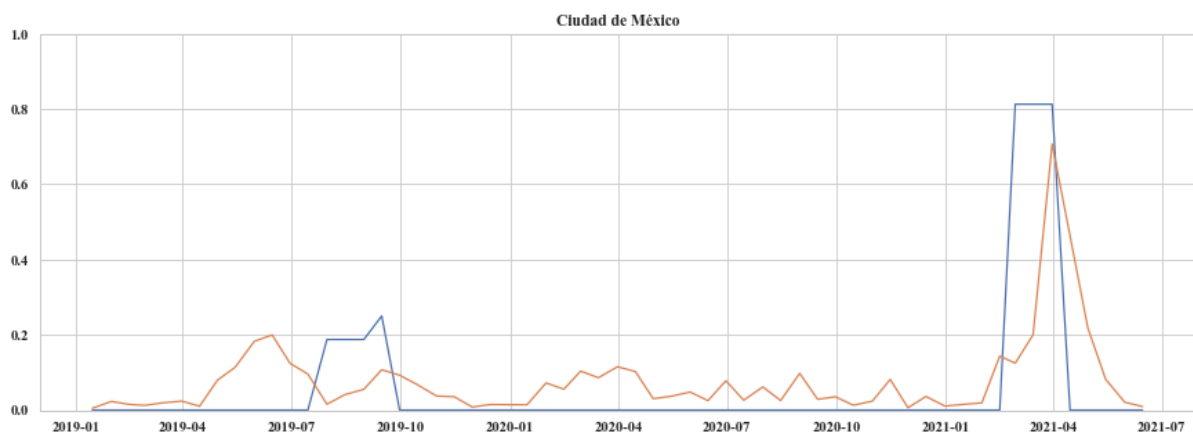
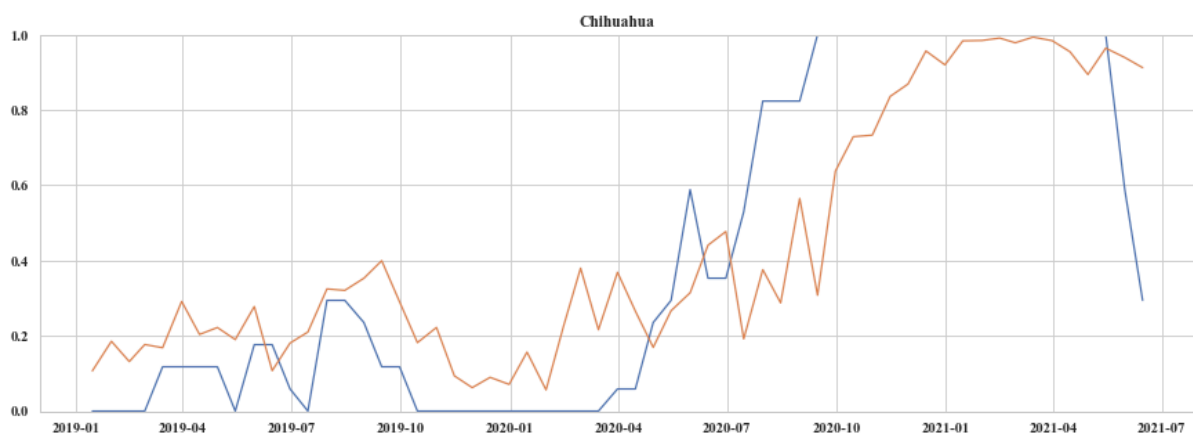
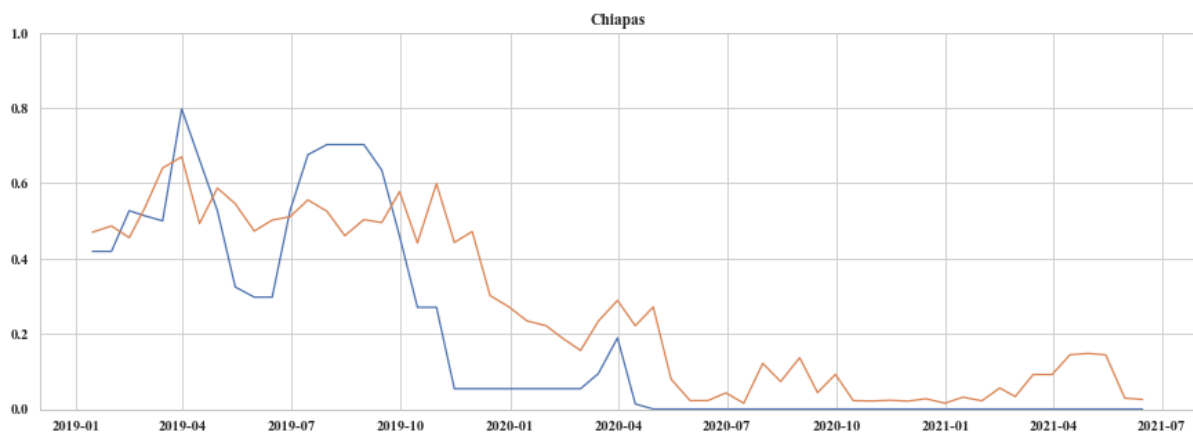
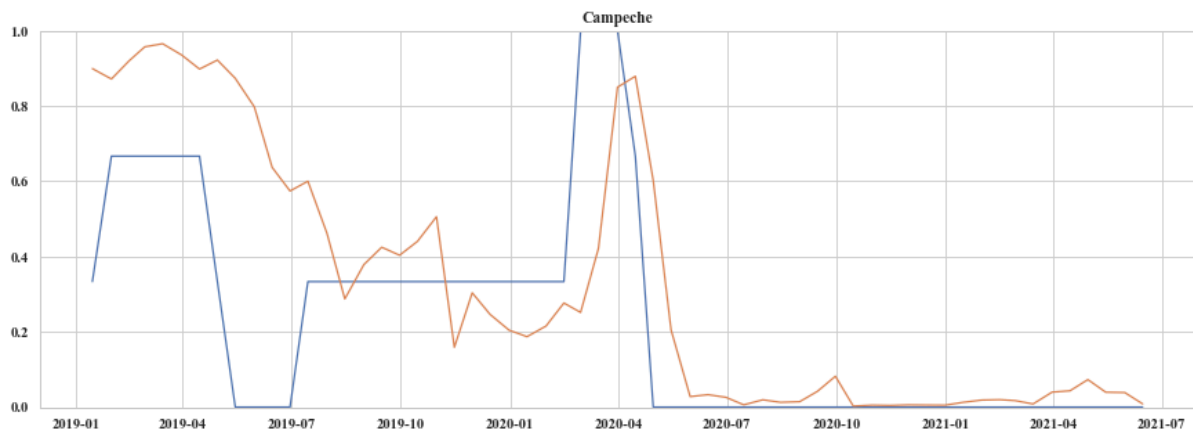
- 'currently_precipProbability_ultimos180_dias_min'
- 'currently_precipProbability_ultimos180_dias_median'
- 'currently_precipProbability_ultimos180_dias_median_avg'
- 'currently_precipProbability_ultimos180_dias_range'
- 'currently_ozone_ultimos180_dias_avg'
- 'currently_ozone_ultimos180_dias_std'
- 'currently_ozone_ultimos180_dias_max'
- 'currently_ozone_ultimos180_dias_min'
- 'currently_ozone_ultimos180_dias_median'
- 'currently_ozone_ultimos180_dias_median_avg'
- 'currently_ozone_ultimos180_dias_range'
- 'currently_cloudCover'
- 'currently_uvIndex'
- 'currently_temperature'
- 'currently_dewPoint'
- 'currently_humidity'
- 'currently_pressure'
- 'currently_icon'
- 'currently_precipIntensity'
- 'currently_precipProbability'
- 'currently_ozone'
- 'pressure_mean'
- 'pressure_std'
- 'pressure_min'
- 'pressure_max'
- 'pressure_median'
- 'pressure_mad'
- 'pressure_count'
- 'humidity_mean'
- 'humidity_std'
- 'humidity_min'
- 'humidity_max'
- 'humidity_median'
- 'humidity_mad'
- 'humidity_count'
- 'dewPoint_mean'
- 'dewPoint_std'
- 'dewPoint_min'
- 'dewPoint_max'
- 'dewPoint_median'
- 'dewPoint_mad'
- 'dewPoint_count'
- 'temperature_mean'
- 'temperature_std'

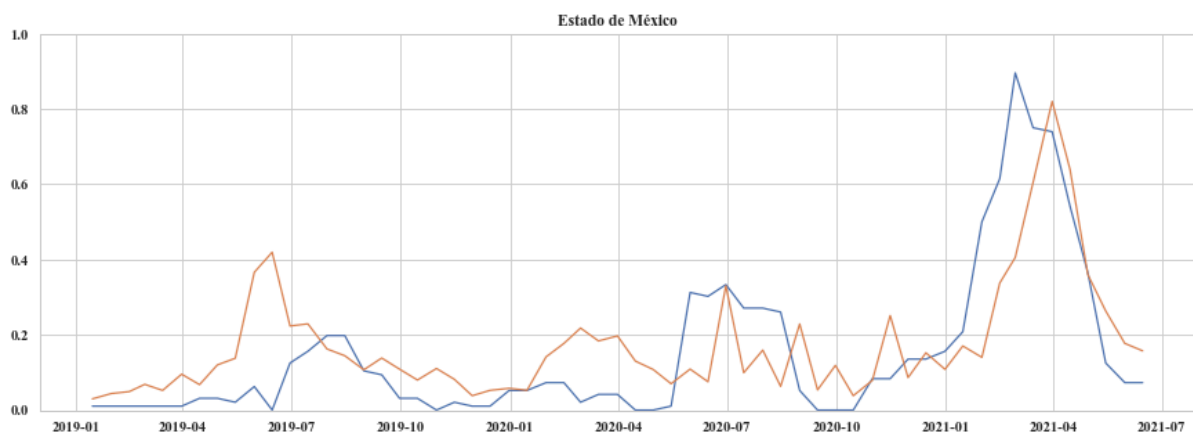
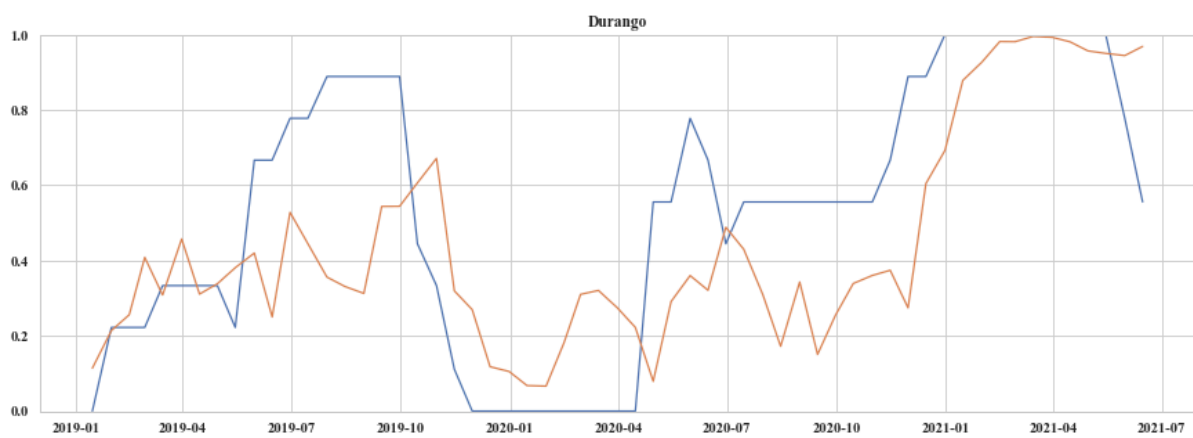
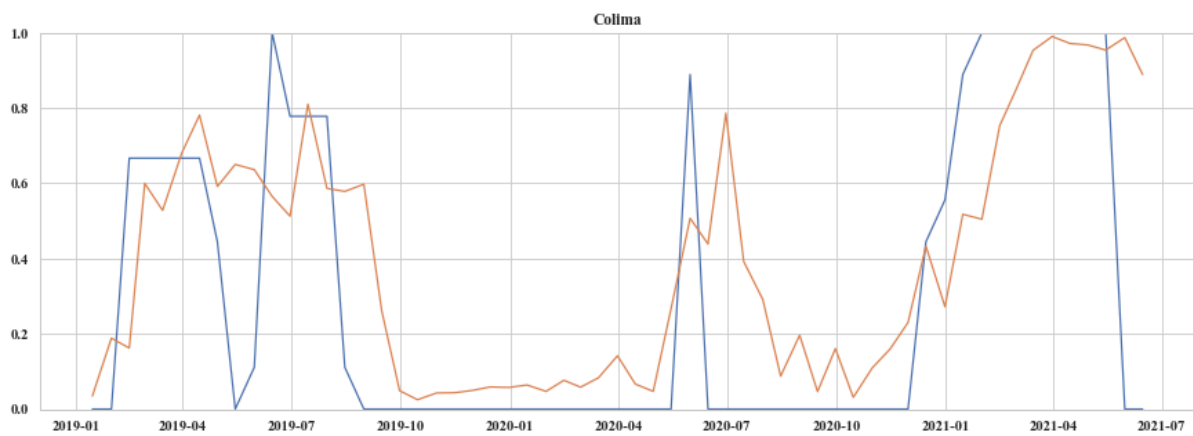
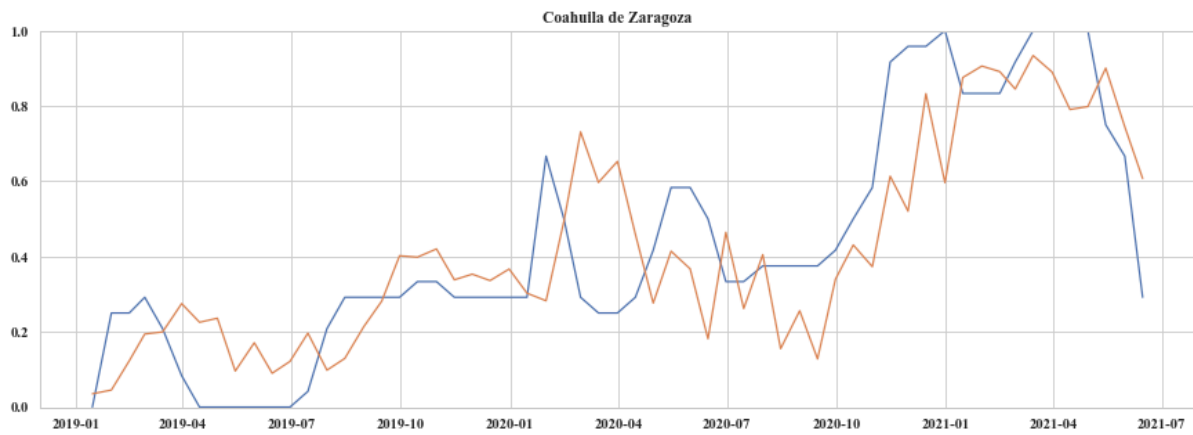
- 'temperature_min'
- 'temperature_max'
- 'temperature_median'
- 'temperature_mad'
- 'temperature_count'
- 'uvIndex_mean'
- 'uvIndex_std'
- 'uvIndex_min'
- 'uvIndex_max'
- 'uvIndex_median'
- 'uvIndex_mad'
- 'uvIndex_count'
- 'cloudCover_mean'
- 'cloudCover_std'
- 'cloudCover_min'
- 'cloudCover_max'
- 'cloudCover_median'
- 'cloudCover_mad'
- 'cloudCover_count'
- 'precipIntensity_mean'
- 'precipIntensity_std'
- 'precipIntensity_min'
- 'precipIntensity_max'
- 'precipIntensity_median'
- 'precipIntensity_mad'
- 'precipIntensity_count'
- 'precipProbability_mean'
- 'precipProbability_std'
- 'precipProbability_min'
- 'precipProbability_max'
- 'precipProbability_median'
- 'precipProbability_mad'
- 'precipProbability_count'
- 'month_of_year'
- 'week_of_year'

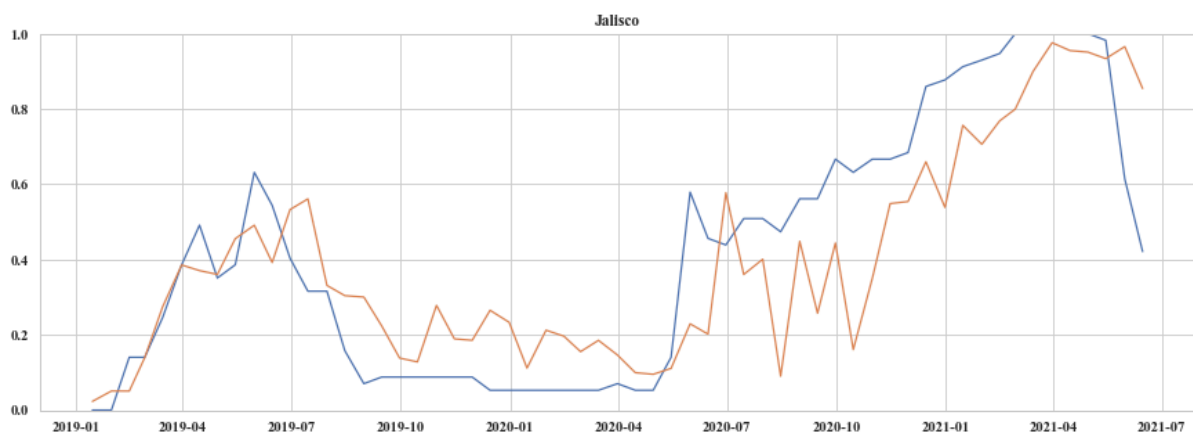
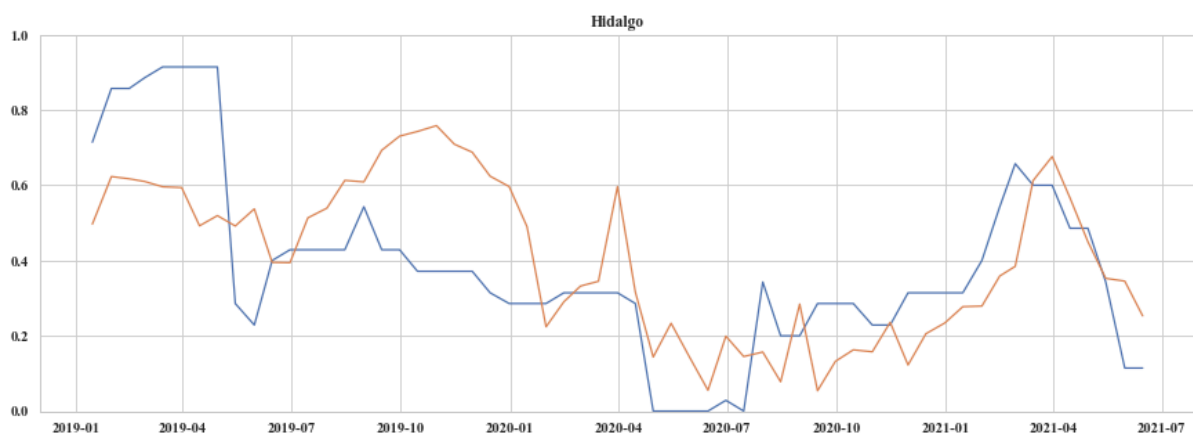
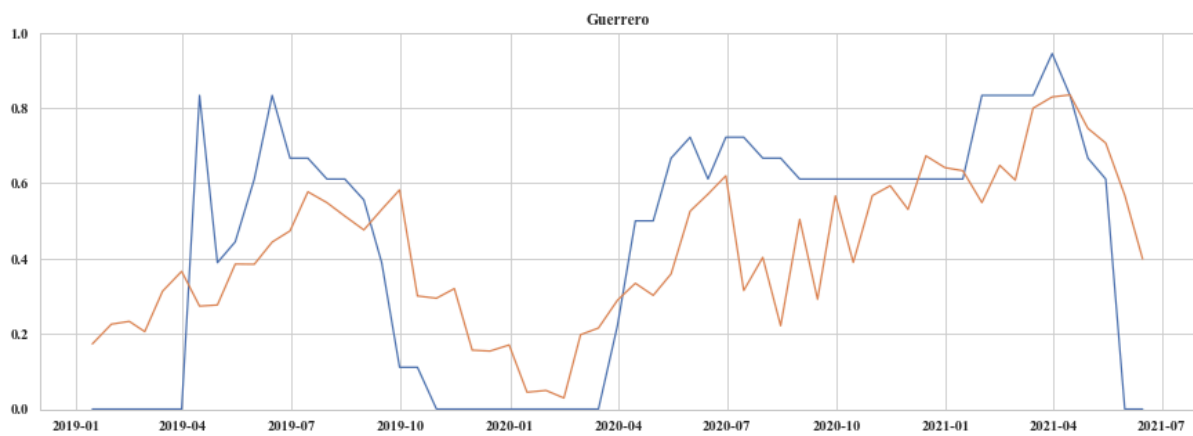
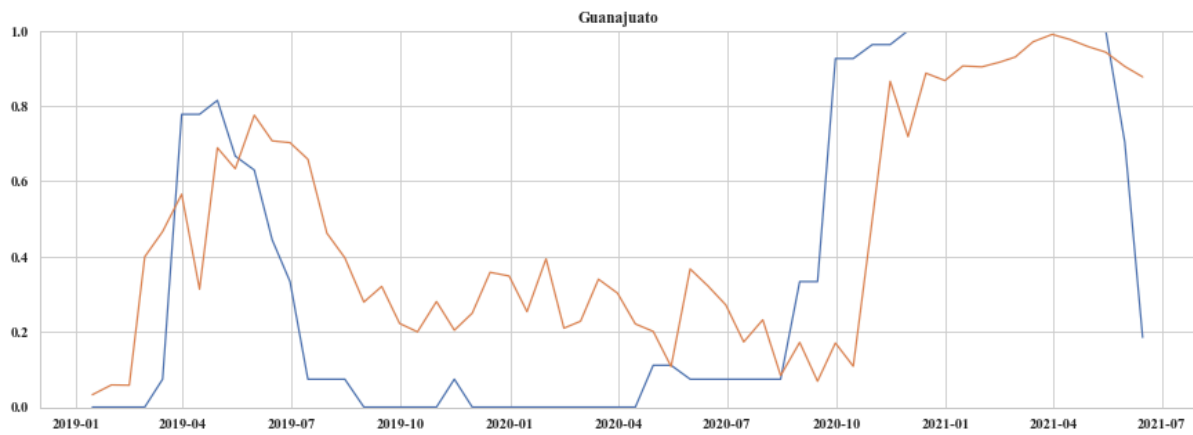
Anexo 4

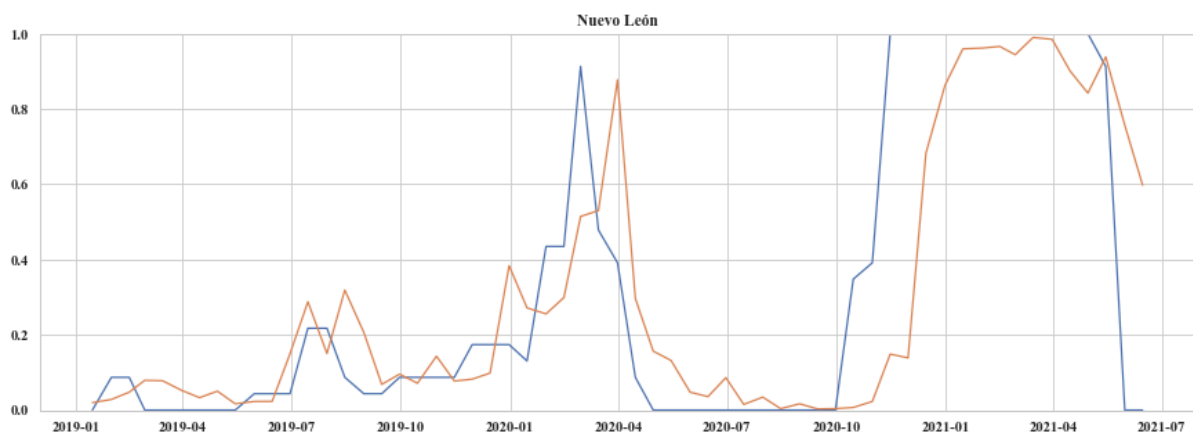
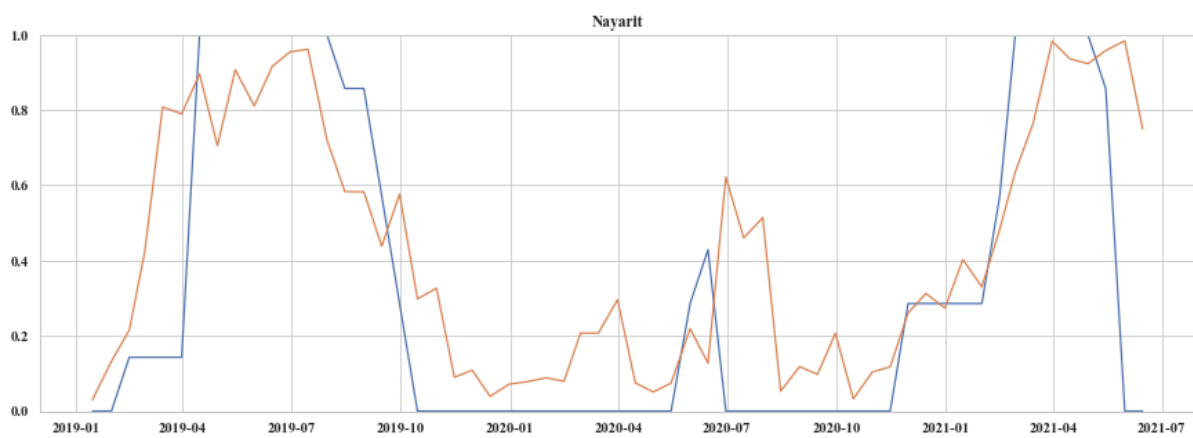
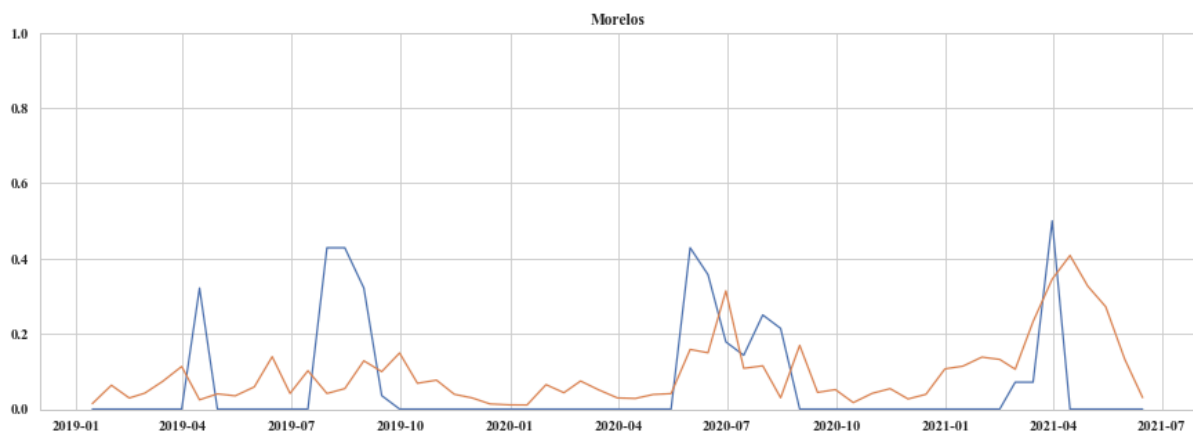
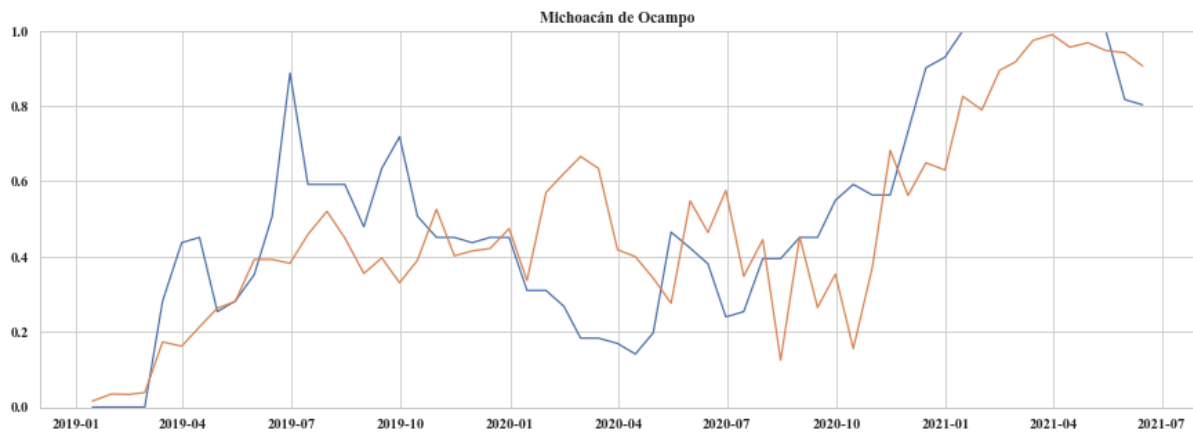
Agrupación de score promedio respecto a la variable de respuesta por entidad federativa

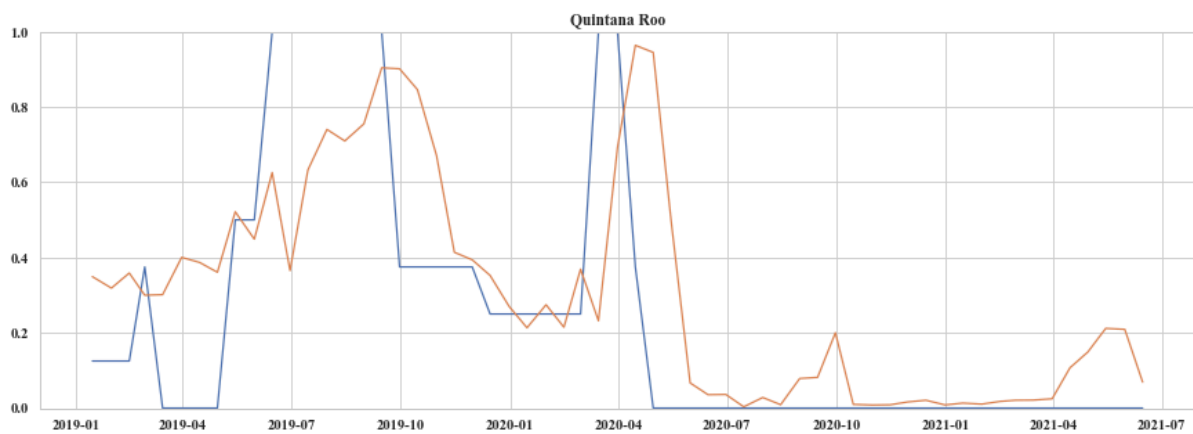
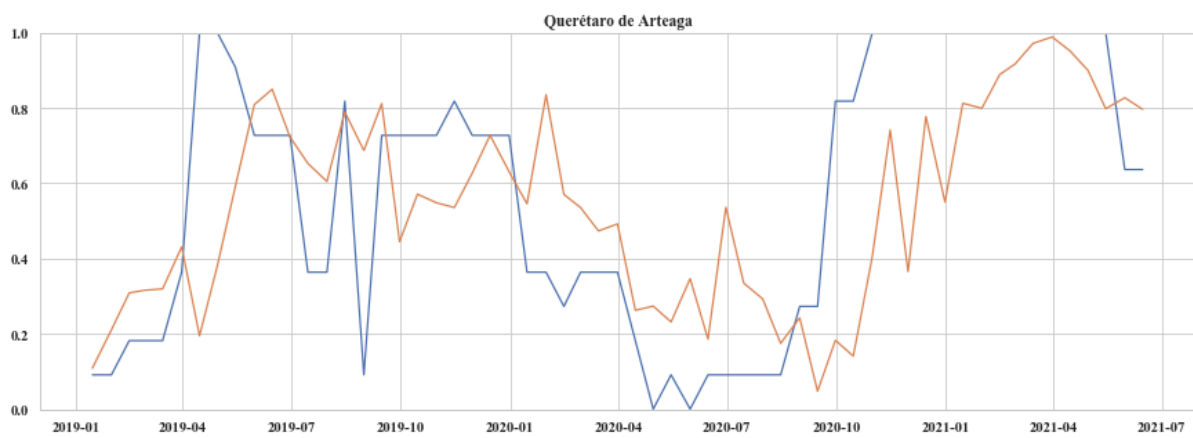
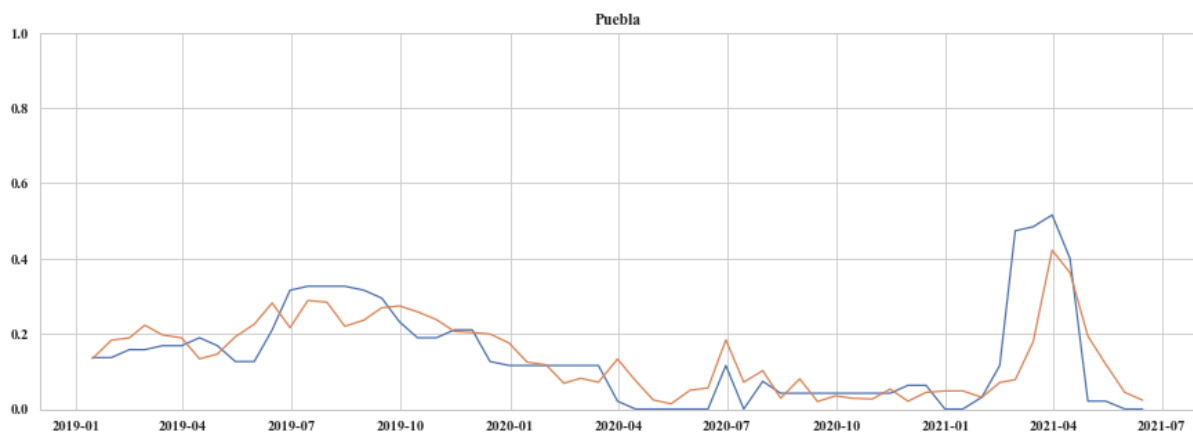
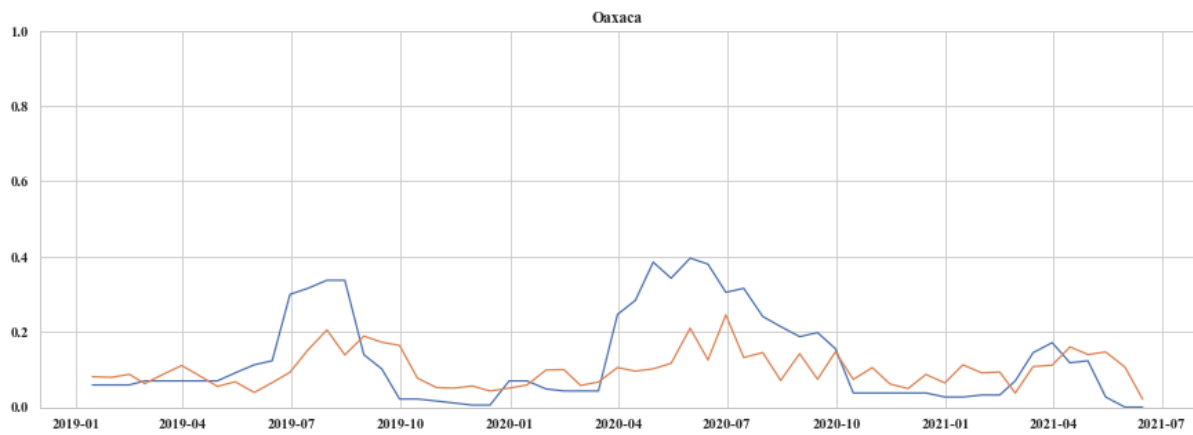


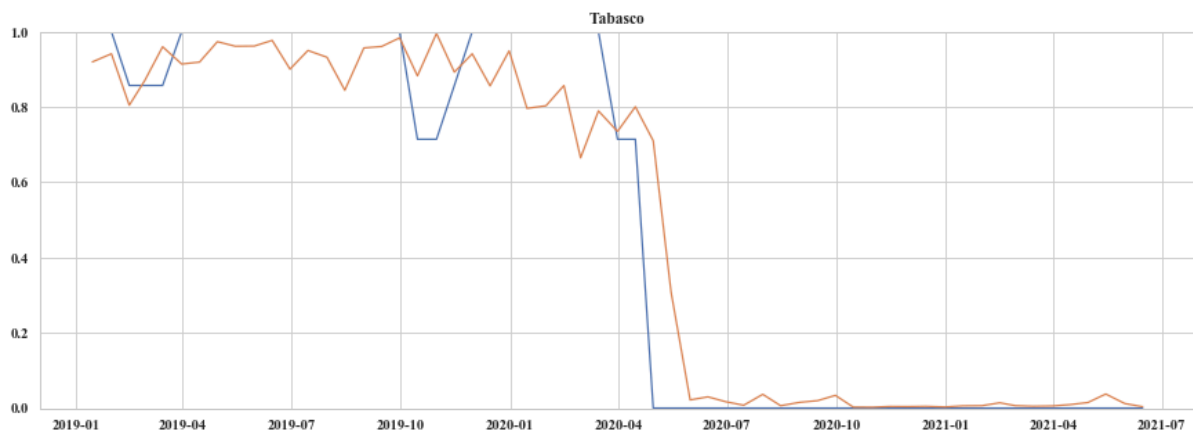
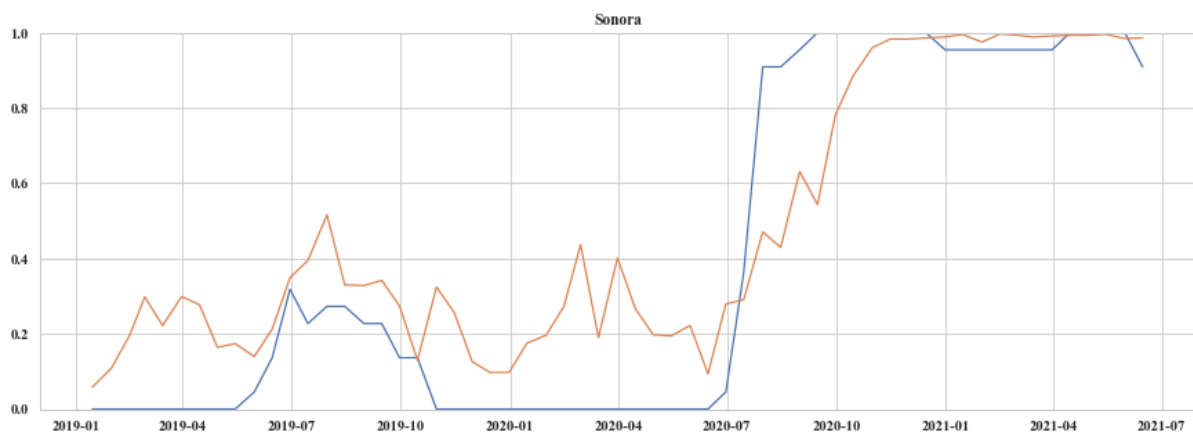
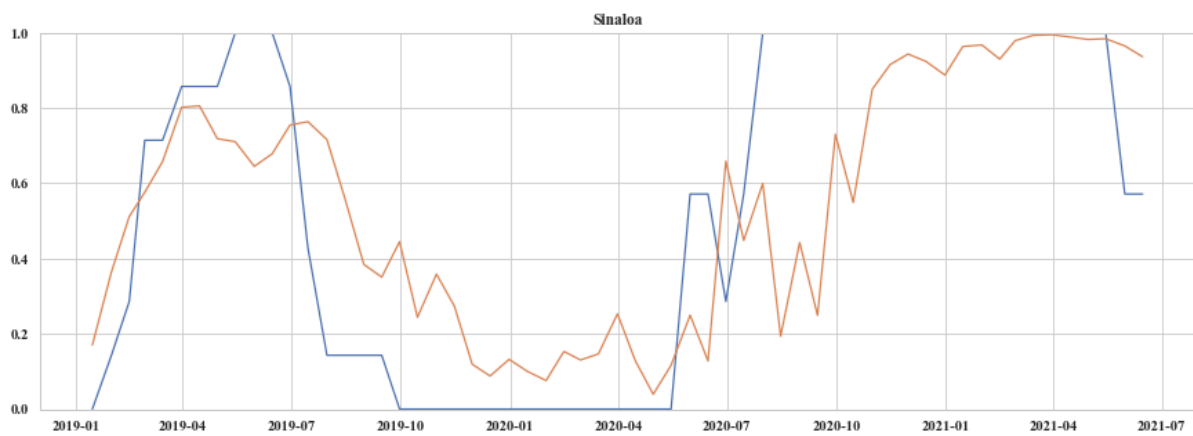
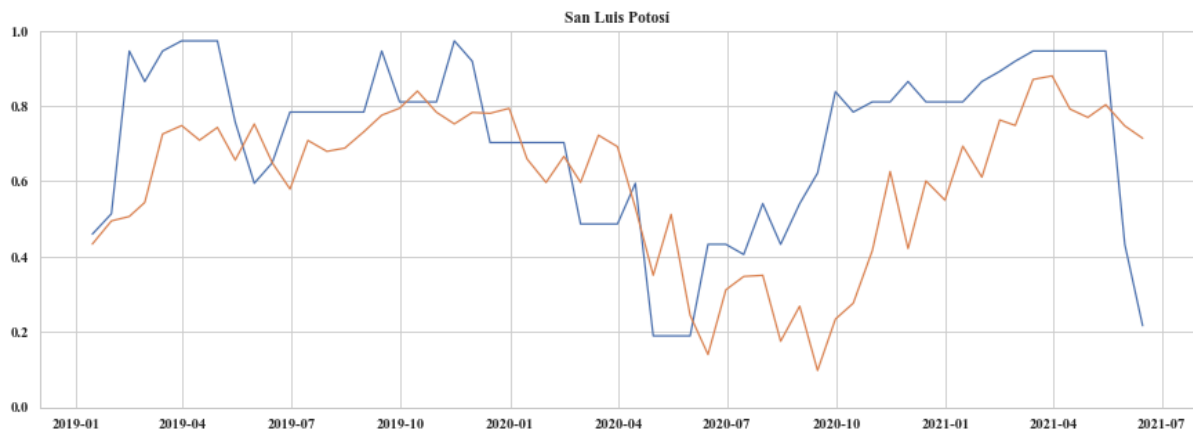


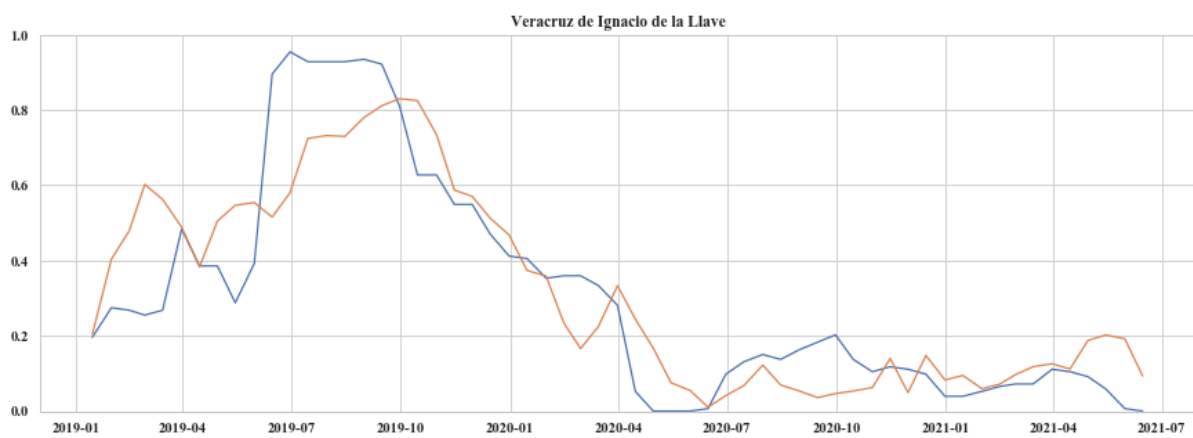
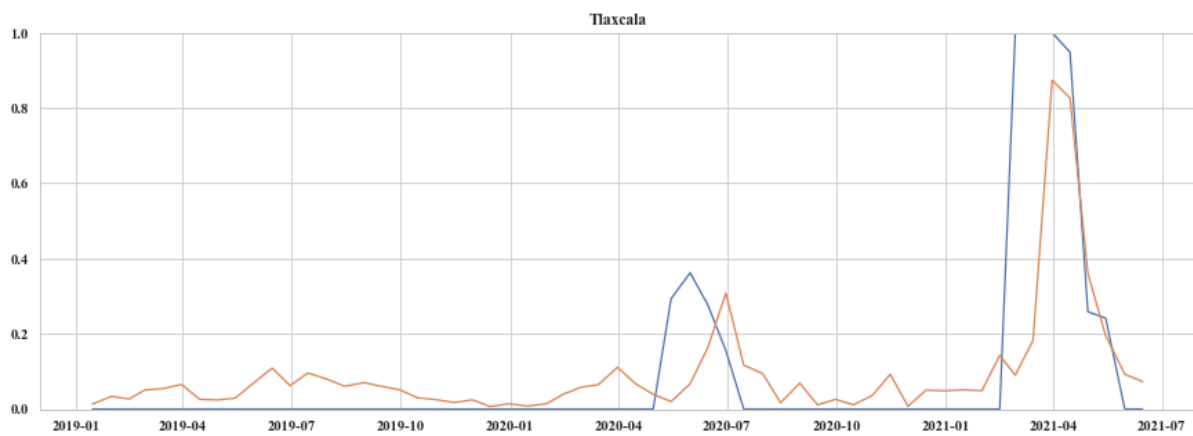
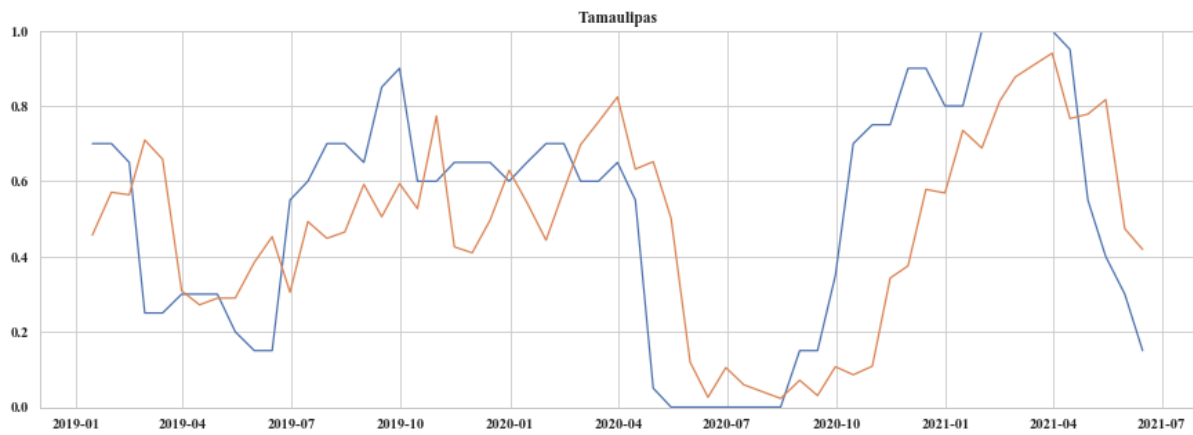


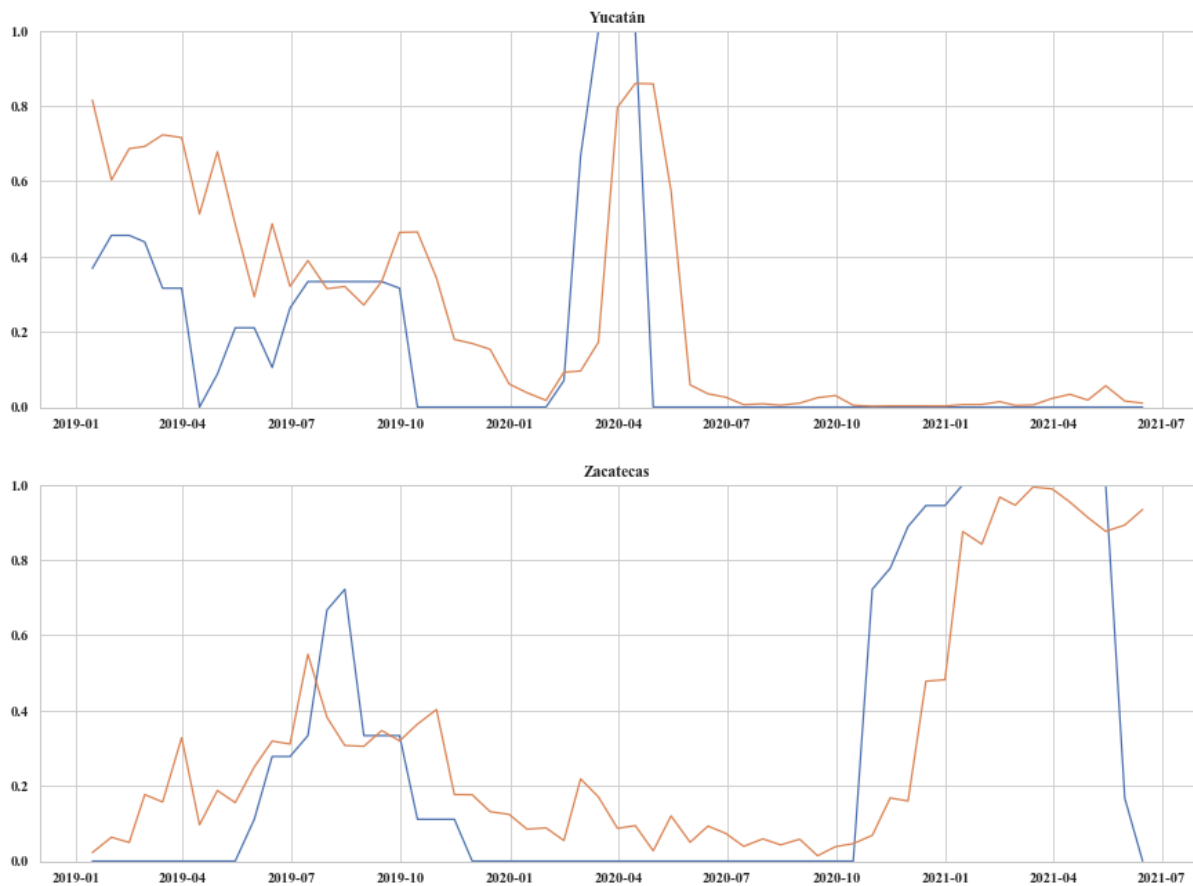






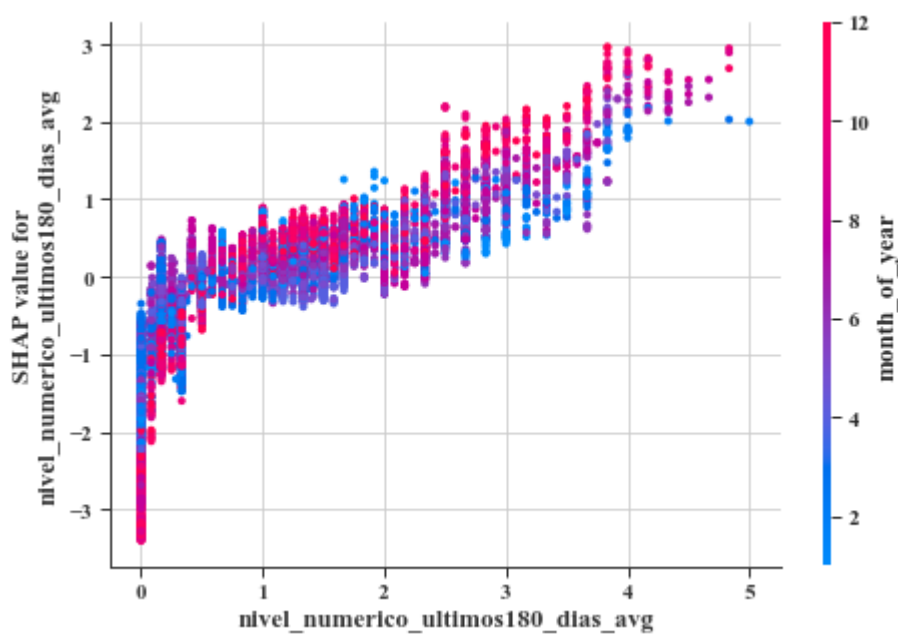


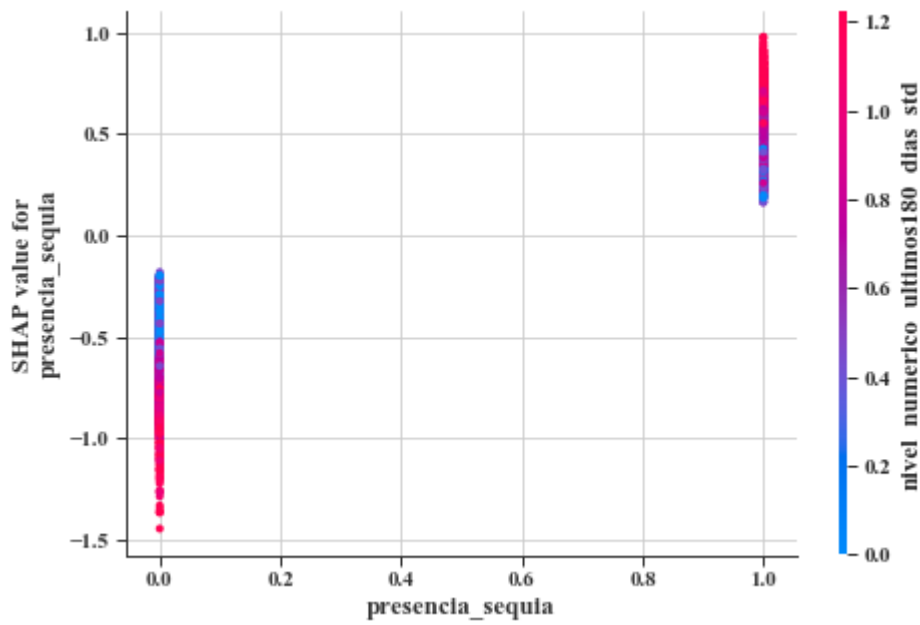
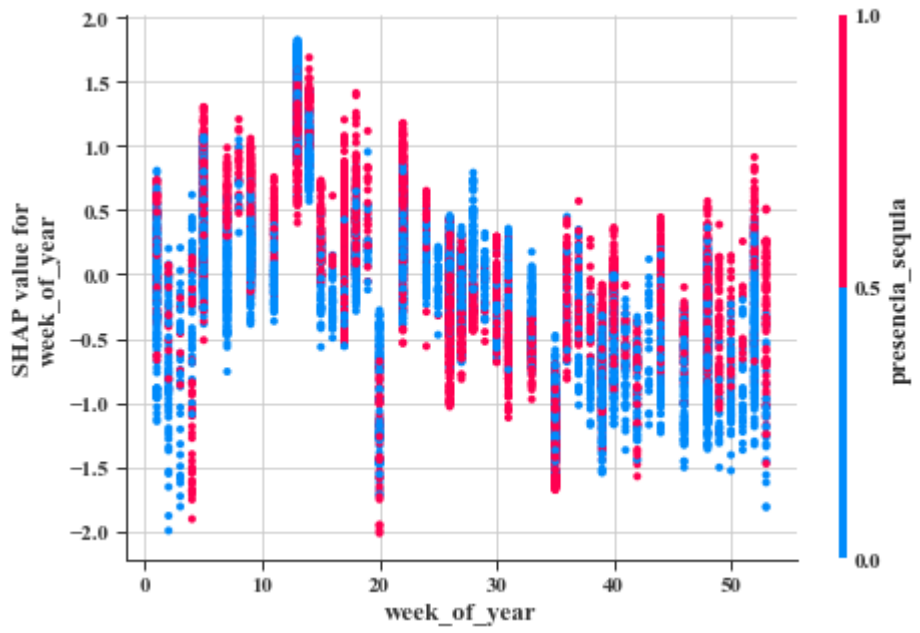


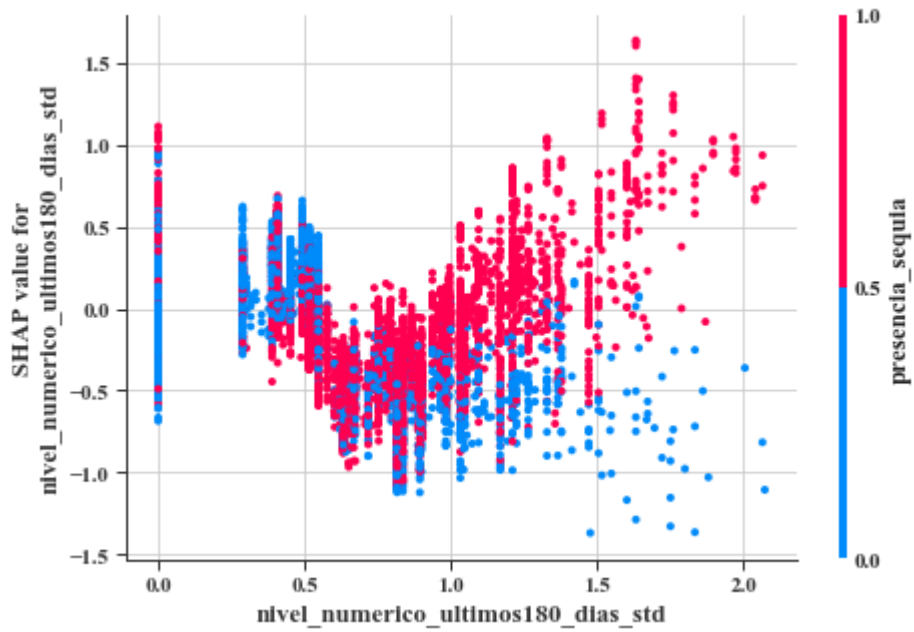
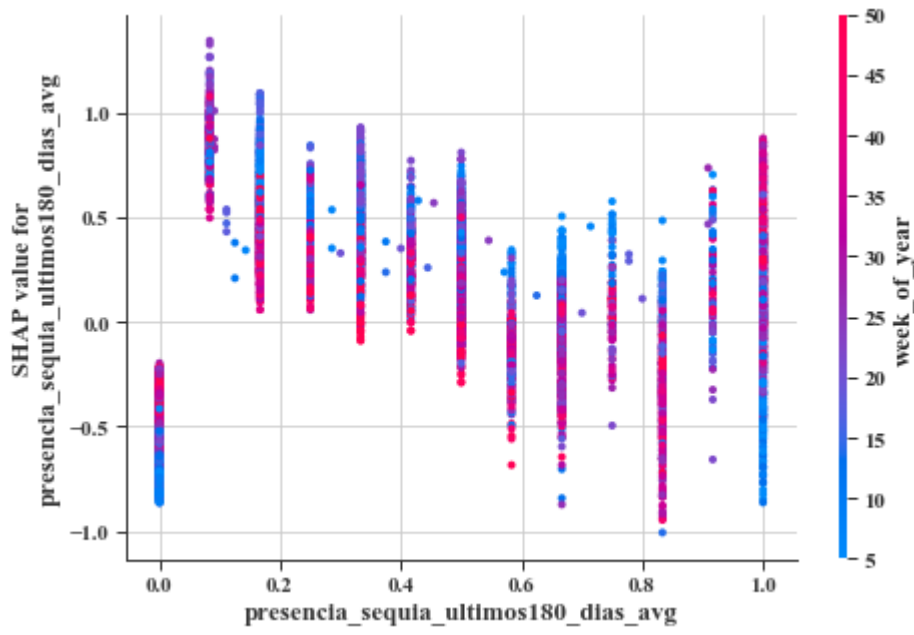


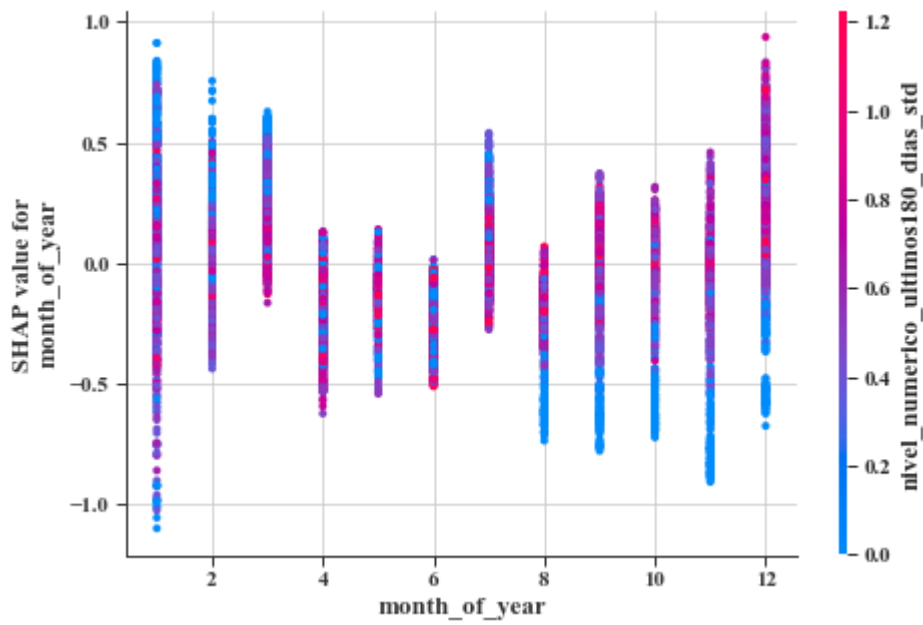
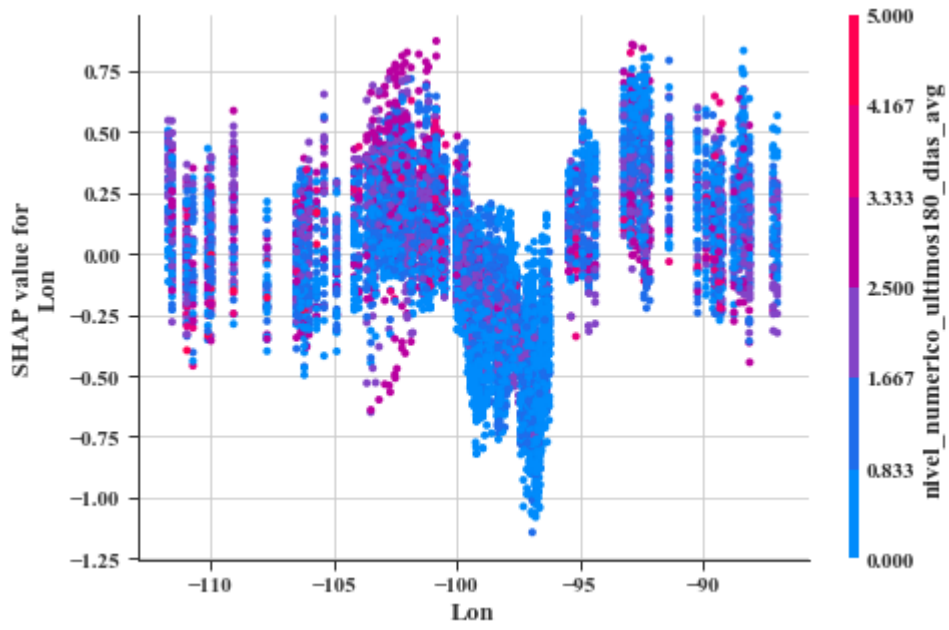
Anexo 5

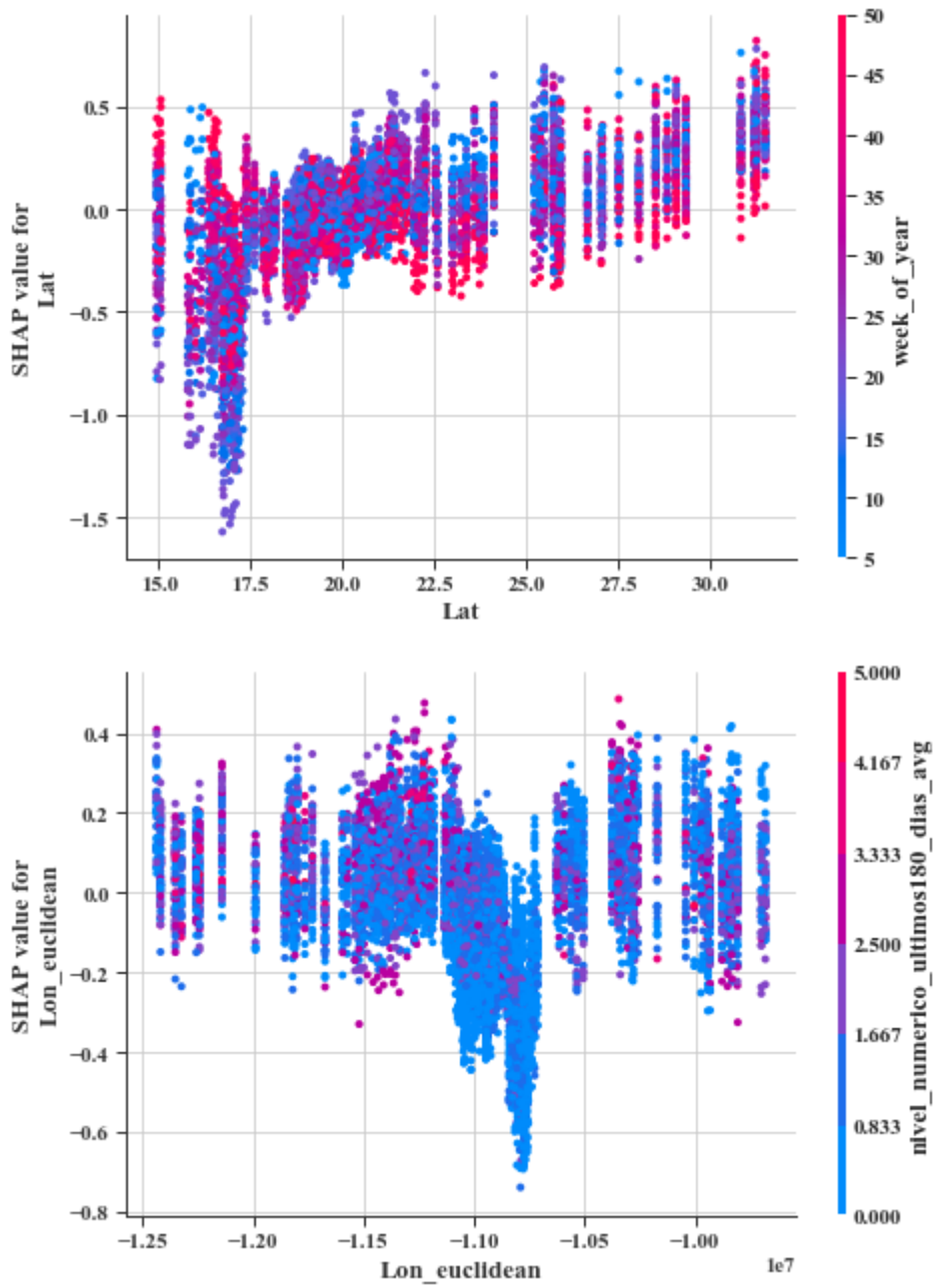
Explicabilidad de las primeras 10 variables más importantes del modelo simple mediante la técnica SHAP

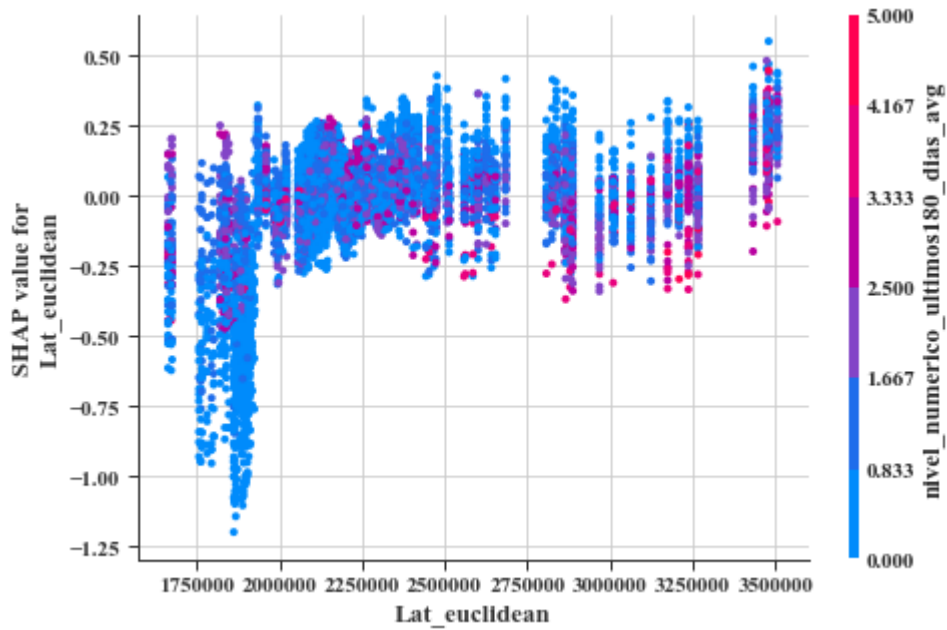












Explicabilidad de las primeras 10 variables más importantes del modelo complejo mediante la técnica SHAP

