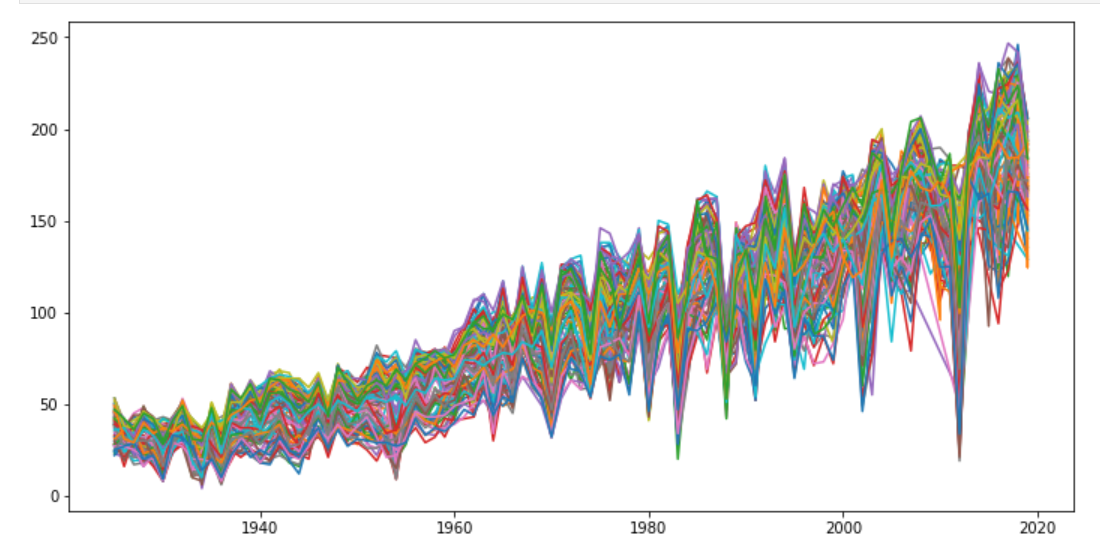
Análisis sobre el rendimiento del cultivo de maíz en Illinois

Los rendimientos de los cultivos no están exentos de sufrir los múltiples efectos nocivos de los que es responsable el cambio climático. El clima extremo es más común ahora que hace siglos, razón por la cual, sequías, inundaciones, altas y bajas temperaturas, entre otros, se requiere modelar dicho comportamiento.

La hipótesis inicial del proyecto se centraba en la predicción de cosechas sobre una región dentro de España, después de cerciorarnos de la poca información sobre las cosechas que hay actualmente disponible y la inexistente normalización en el formato de los datos se decidió elegir Illinois como la localización para nuestro proyecto.

Para poner en contexto el proyecto es imprescindible conocer antes varios aspectos sobre el estado de Illinois: Este estado se caracteriza por su terreno poco accidentado en general, y por su clima inestable así como por varias catástrofes climáticas acontecidas en los últimos años. La agricultura es una importante fuente de ingresos de Illinois especialmente en la producción del maíz y la soja.

Se puede observar en el gráfico mostrado que el rendimiento de la producción del maíz se ha visto incrementado a lo largo de los años, al mismo tiempo podemos constatar que han habido bajadas en la producción provocadas por eventos climáticos extremos.

Tendencia de la producción del 1940 al 2020

Este contexto tan particular de la tendencia de la producción en Illinois nos ayudó a comprender que realizando un proyecto basándonos en la comprensión y categorización de los datos así como el posterior tratado de los mismos con un modelo de predicción nos daría como resultado la posibilidad de predecir la distribución de la producción en Illinois a futuro y con ello se podría ayudar a todo el territorio a prevenir y contrarrestar los picos de baja producción causados por las catástrofes climáticas u otros factores climáticos.

Pasos del proyecto

Para conseguir la predicción del rendimiento de las cosechas hemos priorizado la limpieza y gestión de los datos más óptima. Las dos series de datos históricos principales con los que hemos trabajado han sido : la temperatura a 2m del suelo y el Índice de vegetación EVI. El siguiente paso para analizar la serie temporal de yield es necesario aislar el componente aleatorio, aplicando la diferenciación regular para obtener una nueva serie, con un dato menos que la original, en la que se ha eliminado la tendencia.

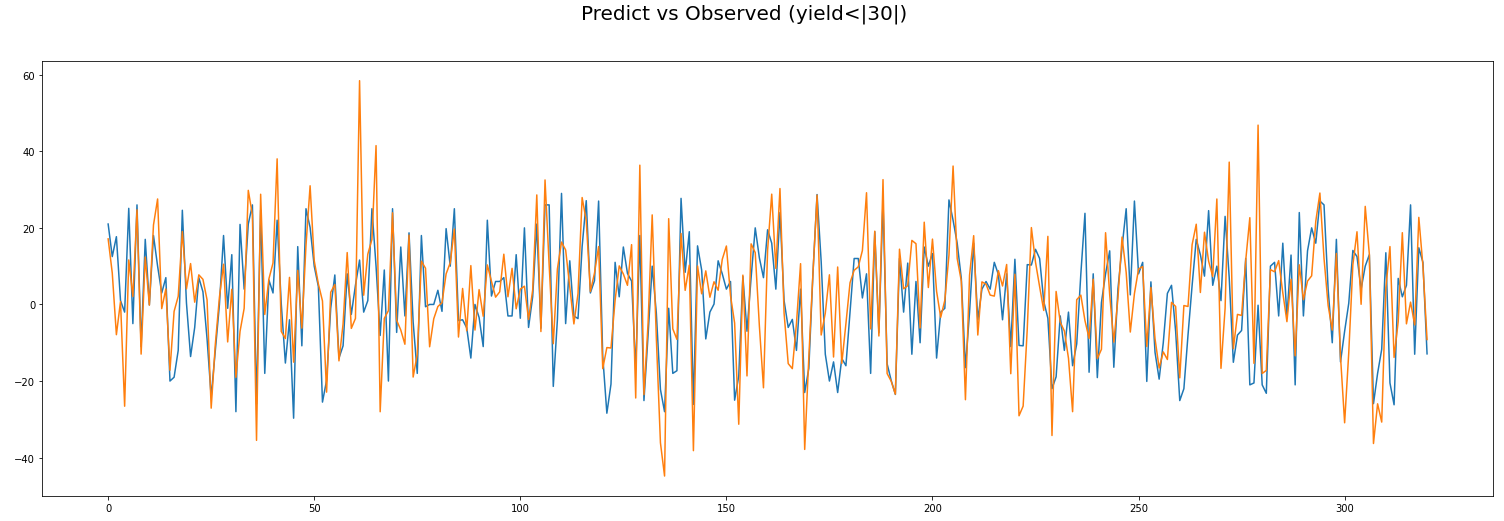
El siguiente paso es el alineamiento de las muestras ya que las muestras de temperatura y de índice de vegetación no se realizaron en el mismo día. Para alinear ambas capturas de información interpolamos la serie de EVI de esta forma generamos un nuevo set de capturas para que ambas series coincidan.

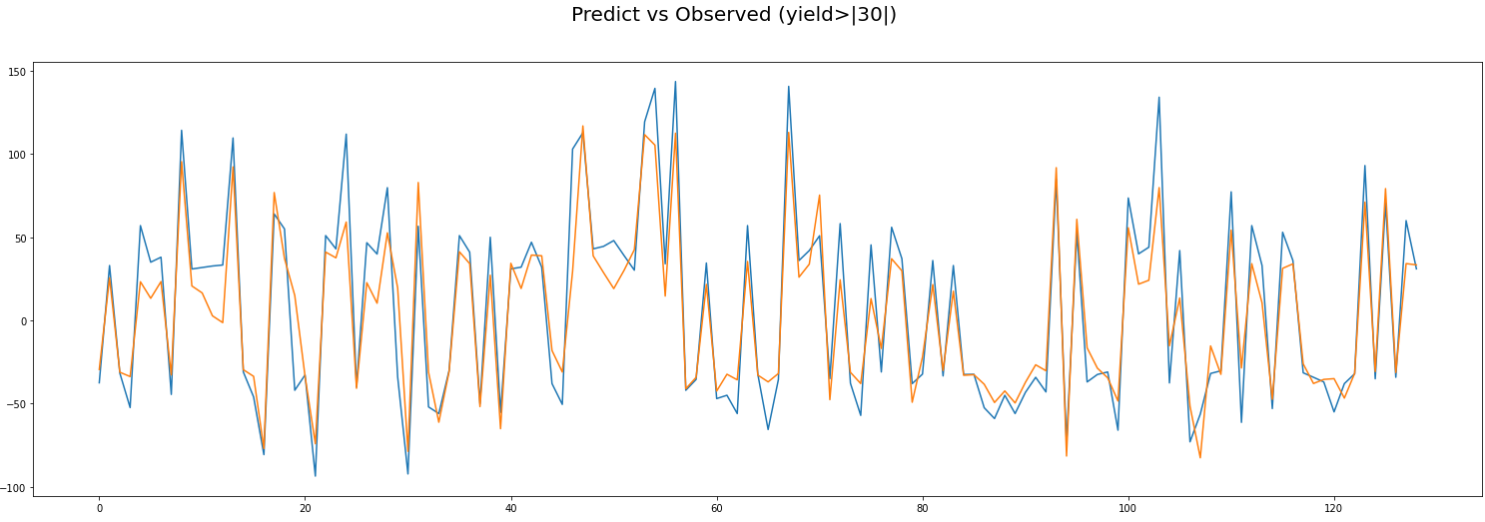
Después de realizar todos estos pasos, indispensables para poder realizar el modelo de predicción, hemos conseguido tener los datos ya limpios para proceder con el mismo. Llegados a este punto hemos explorado varios modelos lineales (LinearRegression, Lasso, Ridge) y el resultado no ha sido el esperado ya que nos hemos encontrado con mucho overfitting. Al final hemos seleccionado un RandomForestRegressor porque consideramos que era el modelo más adecuado para este caso, finalmente para la selección de los hiper parámetros utilizamos el RandomizedSearchCV.

Conclusión

Este proyecto es un punto de partida para el seguimiento, evolución y mejora del rendimiento de producción de cosechas de maíz en Illinois. Nos complace haber encontrado algunas ideas interesantes de los conjuntos de datos proporcionados: una relación entre el rendimiento del cultivo y el tiempo, la temperatura y el valor EVI.

Como resultado del modelo predictivo podemos observar que un error de modelo (varianza) independiente de la magnitud de la producción y que cuanto mayor es la producción menos se nota el error de la predicción del modelo. Según este resultado comprendemos que añadiendo más variables como las precipitaciones, composición del terreno, grado de heladas o catástrofes climaticas podríamos conseguir un porcentaje de error menor.





Como punto de mejora del proyecto podríamos destacar la optimización en el cálculo de interpolación para mejorar la velocidad de procesamiento del código. Es un ejemplo de que con unos datos y variables óptimas se pueden desarrollar modelos de predicción útiles para localizaciones geográficamente equivalentes.

Finalmente, nos gustaría agradecer a Saturdays AI por brindarnos la oportunidad de realizar este proyecto y conocer a otras personas involucradas en el campo de la IA, así como probar nuestras habilidades en ML y ampliar nuestros horizontes en nuevos temas previamente desconocidos para nosotros.