

Curso de Métodos Cuantitativos para la toma de decisiones IF-7200

**Proyecto**

**Pronóstico del clima en Kansas**

Estudiantes

María Fernanda Sandi Calderon - B77168

Alan Fabián Zúñiga Díaz - B78597

Esteban Rosales Mora - B96967

Erick Vasquez - B98334

Aldahir Chaves Mora - B92175

Docente

Luis Delgado

Liberia, julio del 2022

**Resumen**

El presente trabajo muestra de forma investigativa una proyección respecto al clima en el Estado de Kansas en Estados Unidos, en donde se presentarán diferentes datos extraídos del Aeropuerto Internacional de Kansas. Estos datos inician del año 2009 al 2022 (junio). Es muy importante destacar que se encuentran diferentes tipos de datos, como por ejemplo: la cantidad de precipitación, la temperatura mínima, máxima y otros más, por lo que se podrá realizar un pronóstico a cualquiera de estos tipos de datos. Por todo esto, este documento promete poner en marcha la aplicación de los distintos conceptos de ciencia de datos de una manera práctica y sencilla de comprender.

***Palabras claves****:* ciencia de datos, estado de Kansas, pronóstico.

**Índice**

[Introducción](#_heading=h.te7ezhk6pllm) **3**

[Marco Teórico](#_heading=h.e3f3wbxkbj25) **4**

[Inteligencia artificial en análisis de datos](#_heading=h.lb1gn7m721en) 4

[Estudio de pronósticos de climas](#_heading=h.uom80p2ww9fa) 4

[Estudio en Kansas](#_heading=h.gg3oryw93n6s) 6

[Regresión lineal](#_heading=h.ii678d76tz50) 11

[Desarrollo de herramientas Machine Learning](#_heading=h.d79a2g16iocz) **13**

[Machine Learning en Python](#_heading=h.pe2v7zaazbul) 13

[Como entrenar un modelo](#_heading=h.qmij6n4kg5rw) 13

[Python](#_heading=h.30wzkifhzuu1) 23

[Pronóstico de Kansas con Machine Learning](#_heading=h.vkitsdlktwnx) 24

[Explicación de los datos](#_heading=h.6944n8wjhla0) 24

[Desarrollo de pronóstico](#_heading=h.410otcapfagu) 24

[Análisis general del entrenamiento de Kansas con machine learning](#_heading=h.ye8ikfaimdjr) 28

[Conclusión](#_heading=h.erz2jgcrnrb) **30**

[Bibliografía](#_heading=h.79m8jg4pig1q) **31**

# Introducción

El propósito principal de este trabajo de investigación es la explicación teórica y práctica sobre cómo se puede aplicar la ciencia de datos por medio del pronóstico de diferentes eventos. Se puede definir pronósticos como la predicción de lo que sucederá con un elemento determinado dentro del marco de un conjunto de datos. Existen diferentes lenguajes de programación para poder realizar pronósticos, como R, Julia, MatLab y Python. En nuestro caso, utilizaremos Python dado la facilidad de sintaxis con la que cuenta el lenguaje y por la gran cantidad de material qué existe en libros, foros y blogs de internet.

La ciencia de datos es una rama qué se encarga de entender al negocio y sus datos para agregar valor a la organización con toma de decisiones basada en datos y no en “opiniones” y con ello incorporar datos a los productos de software. Un científico de datos está en el centro de un matemático-estadístico y un programador, por lo que debe ser muy hábil con las matemáticas y saber programar en alguno de los lenguajes mencionados anteriormente.

Es por ello, qué actualmente es un puesto qué es requerido en las grandes organizaciones para facilitar en la toma de decisiones. En nuestro caso, aplicaremos conceptos de ciencia de datos como: Machine Learning y Regresión Lineal para comprender y generar pronósticos adecuados para el clima en el Estado de Kansas.

Estos pronósticos normalmente son usados para prevenir cualquier inconveniente climático. Conociendo ciertos parámetros se puede pronosticar tormentas qué pueden afectar a la población, con ello saberlo de antelación ayuda a alertar y prevenir cualquier inconveniente qué se pueda presentar.

# Marco Teórico

## Inteligencia artificial en análisis de datos

El uso de técnicas de inteligencia artificial en nuestros días va cada vez en aumento, y el uso de este tipo de técnicas, como menciona Noriega (2021) en su trabajo de uso de inteligencia artificial para el análisis de datos. Y poder de esa manera generar pronósticos adecuados y precisos sobre la penetración del mar en zonas costeras, donde se analizan distintos métodos de IA, resultando que el mejor método de pronóstico es usando las redes neuronales artificiales.

También es importante en áreas de producción de energías limpias, donde Barraza et, al (2019) muestra un análisis detallado a divisor modelos de inteligencia artificial empleados para que mediante el análisis. Pueden brindar un pronóstico sobre la velocidad de los vientos y así poder desarrollar proyectos de plantas eólicas, sustentándose en las predicciones del modelo de IA.

Por otro lado, FUENTES-COVARRUBIAS (2018) muestra cómo hacer empleo de sistemas basados en conocimiento(IA). Se pueden llevar al análisis de datos históricos de una determinada región basándose en una serie de datos que describen o califican el clima, para así generar pronósticos que puedan ayudar en la labor agraria, previniendo desastres o evitarlos mediante la correcta prevención debido al pronóstico realizado por los sistemas.

## Estudio de pronósticos de climas

Sin duda, el uso de pronósticos del clima cada vez es más importante para el desarrollo de las actividades diarias. Según (CIIFEN, s.f), en este sentido, los pronósticos tienen como objetivo proporcionar información para planificar y prepararse, de manera activa, ante eventos meteorológicos.

Por otro lado, el empleo de los registros obtenidos por las estaciones meteorológicas o satélites, son la base para estudio y análisis del clima, estos datos son incorporados a modelos computacionales que generan pronósticos.

En este estudio profundizaremos lo que son los pronósticos del clima, cuyo objetivo está establecido en pronósticos a un plazo más largo, entre meses, estaciones o hasta un año. Con el fin de no confundirlo con pronósticos de tiempo que son a un plazo más corto de horas, días o hasta una semana.

Una vez aclarado estos conceptos, la aplicación de pronósticos de tiempo son más acertados que los pronósticos del clima, esto por la gran distancia entre tiempo a analizar, ya que durante este tiempo son muchos los factores que puede influir en el clima y poder desvalorizar el pronóstico por eso es importante tener en cuenta estos factores presentes.

El uso de pronósticos del clima cada vez va más en aumento, y el uso de este tipo de técnicas, como menciona (PreventionWeb, 2009), están relacionados con la predicción e información del clima para la adopción de decisiones, temas cruciales que generan debates acerca de las diversas formas de adaptación a la variabilidad del clima y al cambio climático en los diferentes países y comunidades del mundo.

El uso de la tecnología en los pronósticos son de gran apoyo en las actividades, mejora las decisiones preventivas, como se menciona en el artículo (TecnoHotel, 2020), ya que dependiendo de las condiciones climatológicas, las personas tienen unas necesidades u otras. En conjunto trabajo de los datos meteorológicos y la big data brinda a las personas información que tiene un potencial económico y mejor calidad de vida.

## Estudio en Kansas

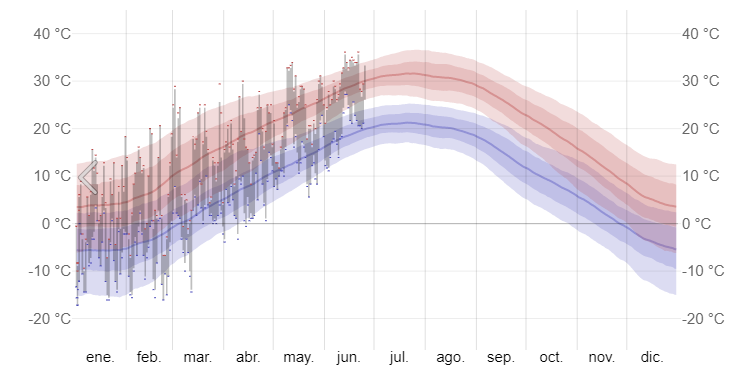
Kansas es un estado de Estados Unidos, donde su capital se denomina Topeka y su ciudad más poblada es Wichita. Kansas se encuentra ubicado en la región Medio Oeste del país, limitado al norte con Nebraska, al este con Misuri, al sur con Oklahoma y al oeste con Colorado.

El clima de este estado se puede caracterizar en tres tipos:

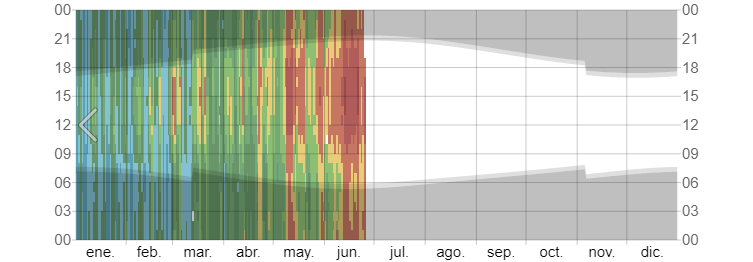
* Estepa continental húmeda.
* Estéril semiárido.
* Subtropical húmeda.

Para Kansas se investigó acerca de diferentes estudios que se hayan realizado y existen estudios detallados sobre el clima a lo largo del tiempo en donde se muestran diferentes gráficas que detallan los diferentes estudios y resultados que datan del año 2014 hasta la fecha actual. Dando un rango de estudio cercano a los 10 años, lo cual era un rango estimado a encontrar.

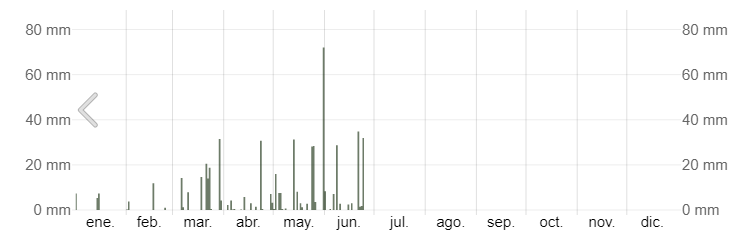
La cantidad de gráficos que se encuentran en la página como antes se ha mencionado van desde el 2014 hasta la fecha actual, por lo que para efectos de presentar los diferentes gráficos requeridos para el presente proyecto, se utilizarán únicamente los estudios del año 2022.

Temperatura máxima y mínima promedio 2022

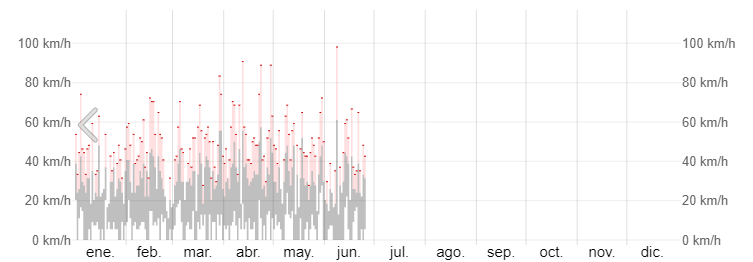
([El tiempo en Kansas City en 2022 (Kansas, Estados Unidos) - Weather Spark](https://es.weatherspark.com/h/y/9825/2022/Datos-hist%C3%B3ricos-meteorol%C3%B3gicos-de-2022-en-Kansas-City-Kansas-Estados-Unidos#Figures-Temperature))

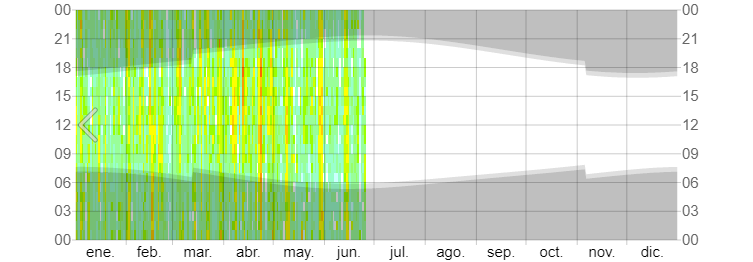
Temperatura por Hora 2022

([El tiempo en Kansas City en 2022 (Kansas, Estados Unidos) - Weather Spark](https://es.weatherspark.com/h/y/9825/2022/Datos-hist%C3%B3ricos-meteorol%C3%B3gicos-de-2022-en-Kansas-City-Kansas-Estados-Unidos#Figures-Temperature))

Precipitación diaria 2022

([El tiempo en Kansas City en 2022 (Kansas, Estados Unidos) - Weather Spark](https://es.weatherspark.com/h/y/9825/2022/Datos-hist%C3%B3ricos-meteorol%C3%B3gicos-de-2022-en-Kansas-City-Kansas-Estados-Unidos#Figures-Temperature))

Velocidad del viento en 2022

Presión atmosférica por hora en 2022

([El tiempo en Kansas City en 2022 (Kansas, Estados Unidos) - Weather Spark](https://es.weatherspark.com/h/y/9825/2022/Datos-hist%C3%B3ricos-meteorol%C3%B3gicos-de-2022-en-Kansas-City-Kansas-Estados-Unidos#Figures-Temperature))

Otros estudios que se buscaron en Kansas fueron acerca de la existencia de ciencia de datos y machine learning con respecto al clima, pero de estos no se encontró información relacionada.

Información encontrada relacionada a la ciencia de datos y Machine Learning se encontró un centro de investigación llamado *“Center for Artificial Intelligence and Data Science Department of Computer Science, Kansas State University”* de la universidad Kansas State University*.*

En este departamento se encuentran diferentes proyectos dedicados a el Machine Learning y Data Science como lo son:

* The ML&DS Lab aims to design algorithms and to develop tools for getting insights from large amounts of data, in particular, social media, user behavior, security and bioinformatics data. Specifically, we are interested in problems with a limited amount of labeled data, but for which large amounts of unlabeled data are readily available. We focus on the design of semi-supervised and domain adaptation approaches, and use our approaches to learn classifiers for a variety of classification problems ‎(<https://caids.cs.ksu.edu/>).
* The Multiagent & Cooperative Reasoning Lab brings together researchers with various expertise to solve interesting problems in the area of distributed, intelligent, and self-adaptive systems. Application areas are varied and include adaptive network security, multiagent systems, cooperative robotics, sensor networks, and a wide variety of networked systems (<https://caids.cs.ksu.edu/>).
* The Data Semantics (DaSe) Laboratory is an internationally visible research group focusing on semantic web (knowledge graphs, ontologies, data management) and artificial intelligence (knowledge representation, automated reasoning, neuro-symbolic integration). Use-inspired research draws from collaboration experience in a wide range of disciplines, including the geosciences, agriculture, and the digital humanities.(<https://caids.cs.ksu.edu/>).
* The research emphasis of the KSU KDD Lab is in the areas of applied artificial intelligence (AI) and knowledge-based software engineering (KBSE) for decision support systems. More specifically, we are interested in machine learning, data mining and knowledge discovery from large spatial and temporal databases, human-computer intelligent interaction (HCII), and high-performance computation in learning and optimization. (<https://caids.cs.ksu.edu/>).
* The mission of the Data-Driven Discovery lab is to develop methodology and tools that turn large and complex data into knowledge and scientific discoveries. By combining elements from machine learning, soft computing, and computational statistics, the lab follows scientific and experimental practices, with the goal of creating bridges between data science foundations and the needs of experimentalists and discovery-driven researchers. The lab works in a broad range of scientific domains, including astronomy, humanities, social science, and the life/medical sciences. (<https://caids.cs.ksu.edu/>).

Todos los proyectos mencionados y explicados fueron extraídos de la página de Kansas State University, en donde se puede apreciar que es una universidad que dedica gran esfuerzo y dedicación a promover el interés de estas tecnologías en los jóvenes. Además la página nos muestra diferentes imágenes de los jóvenes y sus profesores que se encuentran en desarrollo de los diferentes proyectos.

## Regresión lineal

La regresión lineal no ayuda a pronosticar o predecir el comportamiento de una variable dependiente en específico a partir de una segunda variable conocida como independiente. Mediante el ajuste de ecuaciones lineal, este método estadístico, relaciona las variables antes vistas. Según nos menciona el escritor Joaquín Amat Rodrigo, según en el contexto qué nos encontramos en el desarrollo de la ecuación lineal; se conoce como regresión lineal simple cuando solo contamos con una variable independiente y se llama regresión lineal múltiple cuando hay más de una. (Amat Rodrigo, 2020)

En Python, existen dos de las librerías más utilizados para el análisis de datos, él primero es **statsmodels** el cual es muy eficiente en regresión logística; la segunda librería y es la qué se utilizara en el proyecto, es la **scikit-learn** la cual es mucho más rápido para la regresión lineal qué la librería anterior, de igual forma, esta está diseñada principalmente para el Machine Learning.

Es importante mencionar qué cuando los datos graficados no muestran un comportamiento lineal, se debe emplear los modelos de regresión poligonal, el cual es otro tipo de regresión, como la simple y múltiple. Sin embargo, nosotros no veremos este tipo de regresión datas al alcance del proyecto.

Cuando usamos Regresión Lineal Simple, se necesitaran dos parámetros, los cuales son **“w”** y **“b”**. Estos valores son los parámetros para poder realizar el proceso de aprendizaje, el cual se deben estimar **“w”** y **“b”**. El escritor Jose Martínez Heres nos explica sobre el aprendizaje en la regresión lineal, “el aprendizaje consiste en encontrar cuáles son los mejores parámetros (coeficientes) para los datos que tengamos. Los mejores coeficientes serán los que minimicen alguna medida de error”. (Martínez, 2020)

Mencionado lo anterior, la fórmula qué se utiliza es: ***y = wx + b***

Donde **“x”** es la variable a querer estimar. Por ejemplo, si contamos con los siguientes valores para w = 0.0918 y b = 1.2859; y deseamos conocer el resultado para x = 5, tendríamos qué la ecuación para final seria: ***y = 0.0918 \* 5 + 1.2859 = 1.7449.*** (Martínez, 2020)

Lo anterior, fue un ejemplo para comprender cómo es el funcionamiento de la Regresión Lineal Simple. A continuación, veremos los pasos para disminuir el error de la ecuación por medio del método mínimos cuadrados.

1. Método de mínimos cuadrados
2. Predicción
3. Condiciones de aceptación

# Desarrollo de herramientas Machine Learning

## Machine Learning en Python

El machine learning o bien aprendizaje automático es un derivado de la ciencia que estudia la inteligencia artificial, este en particular busca que el sistema aprendan por medio de entrenar un modelo. Gibran (2021) nos explica el fin de efectuar estos entrenamientos, los cuales son analizar los datos para encontrar patrones y así proceder predicciones. Este es el caso de la investigación, se va a entrenar un modelo para realizar predicciones en el clima de Kansas.

Además, en machine learning existen dos tipos de aprendizajes, el supervisado y el no supervisado. El aprendizaje supervisado se dispone de los datos, el cual es un valor continuo, mientras que el aprendizaje no supervisado. En este, no se dispone de datos etiquetados, es decir el algoritmo va aprendiendo de los datos, en este caso se desconoce la respuesta esperada en el momento que se entrena un modelo(García G. 2021).

**Algoritmos para aprendizaje supervisado:**

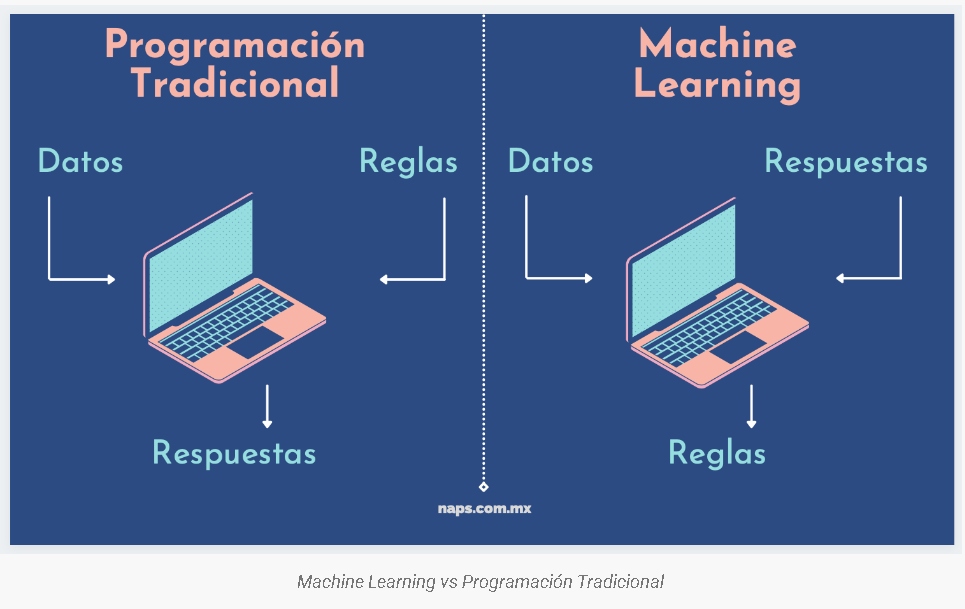
* Regresión Lineal.
* Árbol de Decisión.
* Ramdom Forest.
* Nayve Bayes.

**Algoritmos para aprendizaje no supervisado:**

* SDV (se usa para reducción dimensional).
* PCA (igual emplea para reducción dimensional).
* K-Means

### *Como entrenar un modelo*

En general, para entrenar un modelo de machine learning, se realiza de forma diferente a la programación tradicional que se ha estudiado, en machine learning Gibran (2021), nos dice que el sistema como tal primero se entrena con datos de entrada y datos de respuesta esperados; a esto le llamamos modelo.

*Fuente:García G. 2021*

**Pasos para entrenar un modelo**

En primera instancia se debe definir los datos de entrenamiento y también se debe definir cuál algoritmo se va a utilizar para el entrenamiento. Juaquin Amat (2020), nos explica que la librería de Scikit learn lo hace mediante el uso del objeto ***estimador***, el cual almacena el nombre del algoritmo que se va a emplear en el entrenamiento del modelo. Los cuales hacen uso de los parámetros ***fit(X, y)*** y ***predict(T)***, los cuales tienen la funcionalidad de aprender de los datos y predecir los resultados del modelo. En el siguiente apartado se explica los pasos a seguir:

1. **Entrenamiento**

En este primer paso, Juaquin Amat (2020), dice que se debe ajustar el modelo, en este caso vamos a realizar con el ejemplo de predicción de precios de una vivienda. Entonces se ajusta el modelo con regresión lineal con la regulación de ***ridge***, además, se debe hacer la importación de ***sklearn.linear\_model.Ridge.***  Luego, para asegurarnos que el preprocesado se hace con los datos seleccionados para el entrenamiento, se utiliza el comando ***pipeline***. Aca se muestra un ejemplo del primer paso:

****

*Fuente: (Amat J., 2020)*

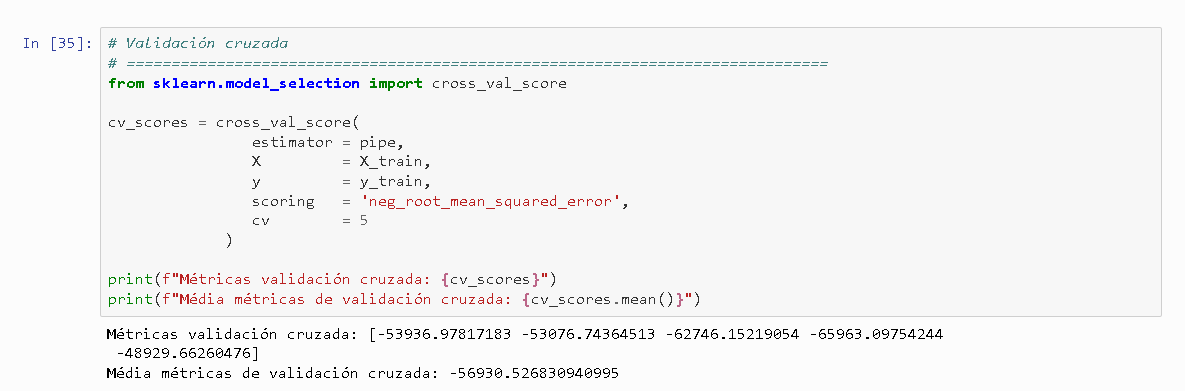
1. **Validación**

Se sabe que la finalidad de entrenar un modelo es predecir una respuesta, es normal encontrar errores en las predicciones, el cual es un medio para conocer cómo está aprendiendo el modelo, aunque este parámetro no es tan realista. Por ello, para conseguir una predicción más certera, Juaquin Amat (2020), nos explica cómo conseguirla. En donde se aplican estrategias de validación, las cuales son basadas en ***resampling.*** También ***Scikit-learn*** posee un módulo para estrategias de validación ***sklearn.model\_selection.***

Para ello:

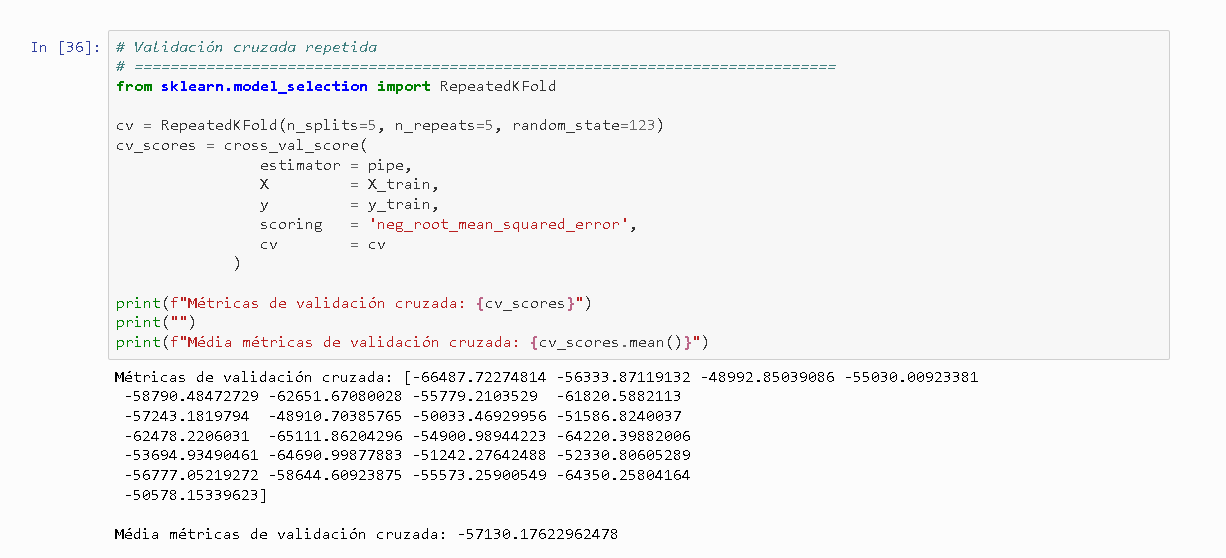
* Primero recibe un argumento ***estimador,*** el cual puede venir de un modelo o bien de pipeline.
* En la mayoría de las métricas se devuelve una respuesta negativa, en donde esté más cerca del valor 0, mejor será el ajuste.

Un ejemplo para una validación es usar la función ***cross\_val\_score(),*** el cual es por defecto de ***KFold:***



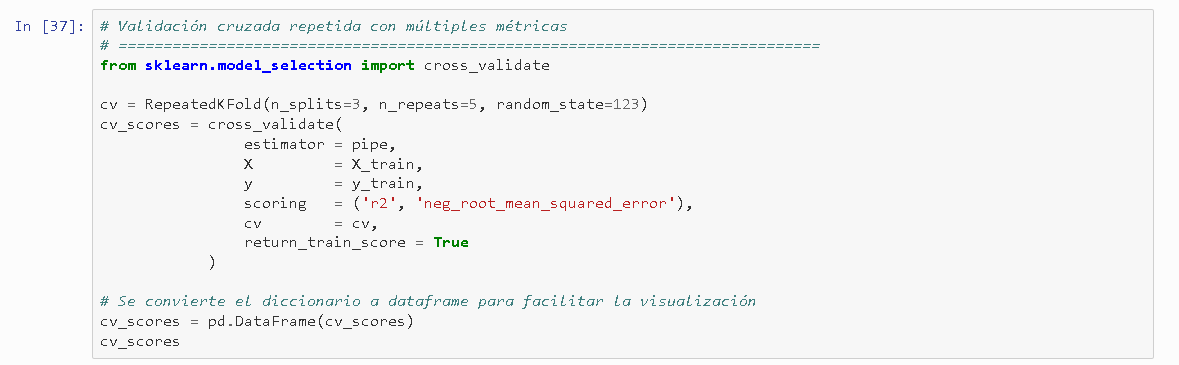
*Fuente: (Amat J., 2020)*

Además, existen otras formas de implementar validaciones cruzadas, entre ellas tenemos: *KFold*, *RepeatedKFold*, *LeaveOneOut*, *LeavePOut*, *ShuffleSplit.* Lo cual es posible si se utilizan las funciones auxiliares de ***sklearn.model\_selection,*** y usando índices de argumento ***cv***. Todos estos métodos funcionan internamente diferente, pero todos tienen en común la finalidad de ajustar y evaluar el modelo que se está entrenando para en cada iteración obtener la estimación de error, esto para llegar al valor real del error del test; ejemplo:

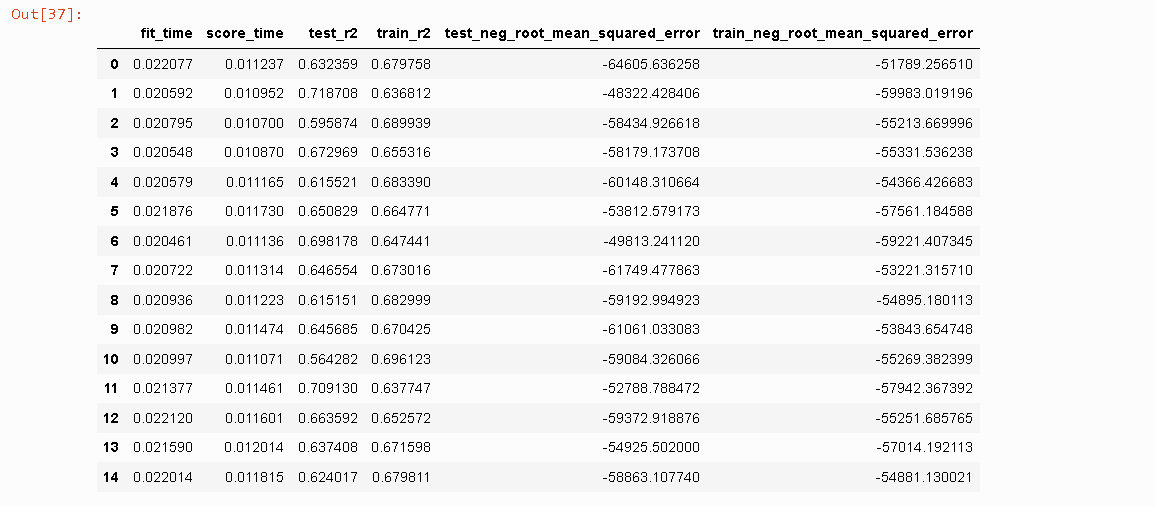


*Fuente: (Amat J., 2020)*

Las funciones: ***cross\_validate*** y ***cross\_val\_score***, son de uso similar, pero en el caso de ***cross\_val\_score***, este se usa para estimar varias métricas a la vez en donde los devuelve los resultados en un diccionario, ejemplo:



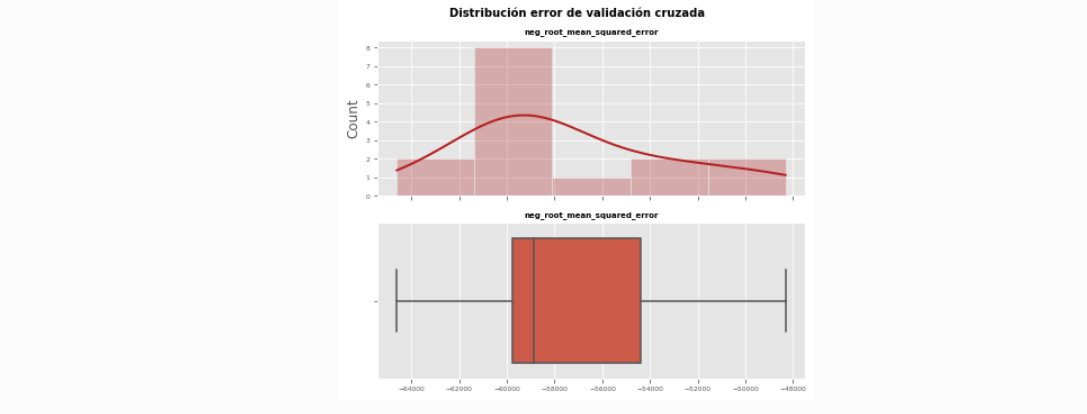
*Fuente: (Amat J., 2020)*



*Fuente: (Amat J., 2020)*



*Fuente: (Amat J., 2020)*

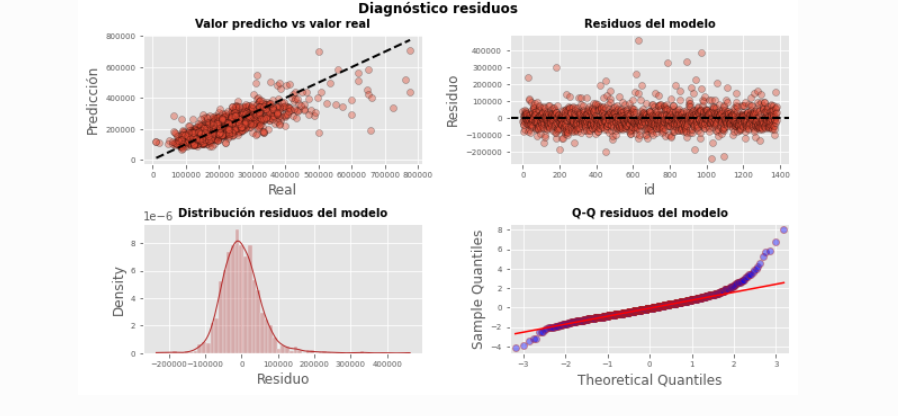


*Fuente: (Amat J., 2020)*

La función ***cross\_val\_predict,*** es utilizada para que de respuesta la predicción de cada partición, lo cual sirve de ayuda para evaluar y diagnosticar los residuos de datos del modelo, ejemplo:



*Fuente: (Amat J., 2020)*



*Fuente: (Amat J., 2020)*

Cuando se realiza una validación cruzada repetida para validar el modelo que se entrena, se debe considerar que:

1. Existe un costo computacional para ajustar reiterativamente un modelo.
2. El subconjunto de datos es diferente en cada iteración.
3. Se producen muchas particiones de datos.

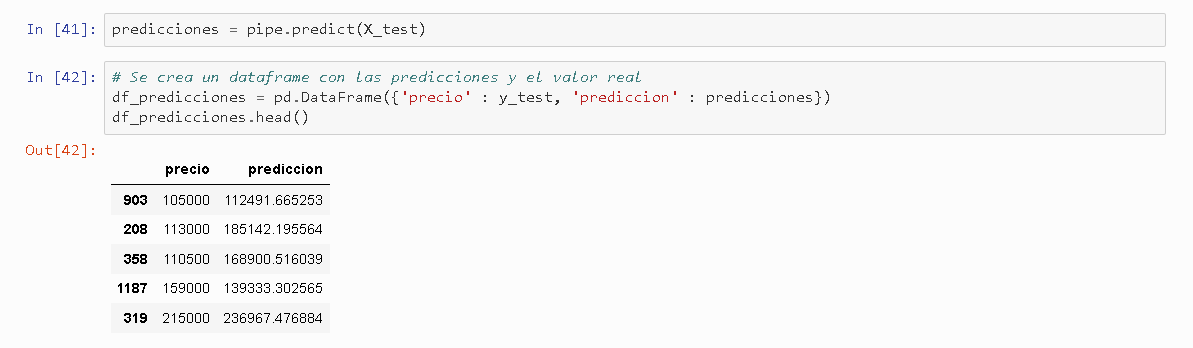
Para paralizar el proceso de las funciones ***cross\_val\_score()*** y ***cross\_val\_predict***, se usa el argumento de ***n\_jobs***. Y con el argumento de ***root\_mean\_squared\_error***, nos da el promedio estimado del error, ejemplo:



*Fuente: (Amat J., 2020)*

1. **Predicción**

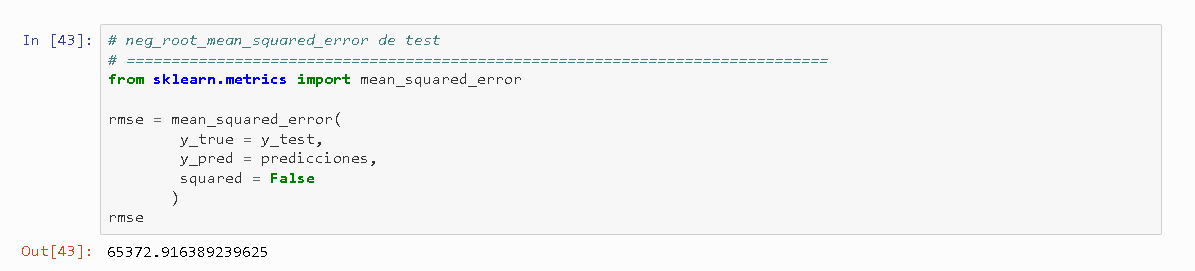
Ya para esta parte, se entiende que el modelo ya fue entrenado con los argumentos estimadores y pipeline. Con la función de ***.predict()***, se pueden obtener las predicciones futuras del modelo. Una recomendación que nos brinda Juaquin Amat (2020), es usar el argumento ***pipeline***, ya que con este se realiza las transformaciones que está aprendiendo el modelo de una forma automática en el entrenamiento, como se muestra en el ejemplo:

****

*Fuente: (Amat J., 2020)*

1. **Error test**

Es importante, una vez realizado el entrenamiento y las predicciones, conocer el error del modelo, esto para ver cuán efectivo ha sido el proceso. Juaquin Amat (2020)nos explica una función para conocer el error de del modelo que se está entrenando el módulo de ***sklearn.metrics*** posee métricas automatizadas para evaluar la eficiencia de las predicciones de nuestro modelo. Como podemos ver en el ejemplo que se adjunta, el error de este modelo es de 65372 y en el modelo que se ha estudiado era de 56735. Lo cual no es un error con una distancia relativamente pequeña en el modelo, podemos considerar que el entrenamiento del modelo es de buena calidad según las predicciones.

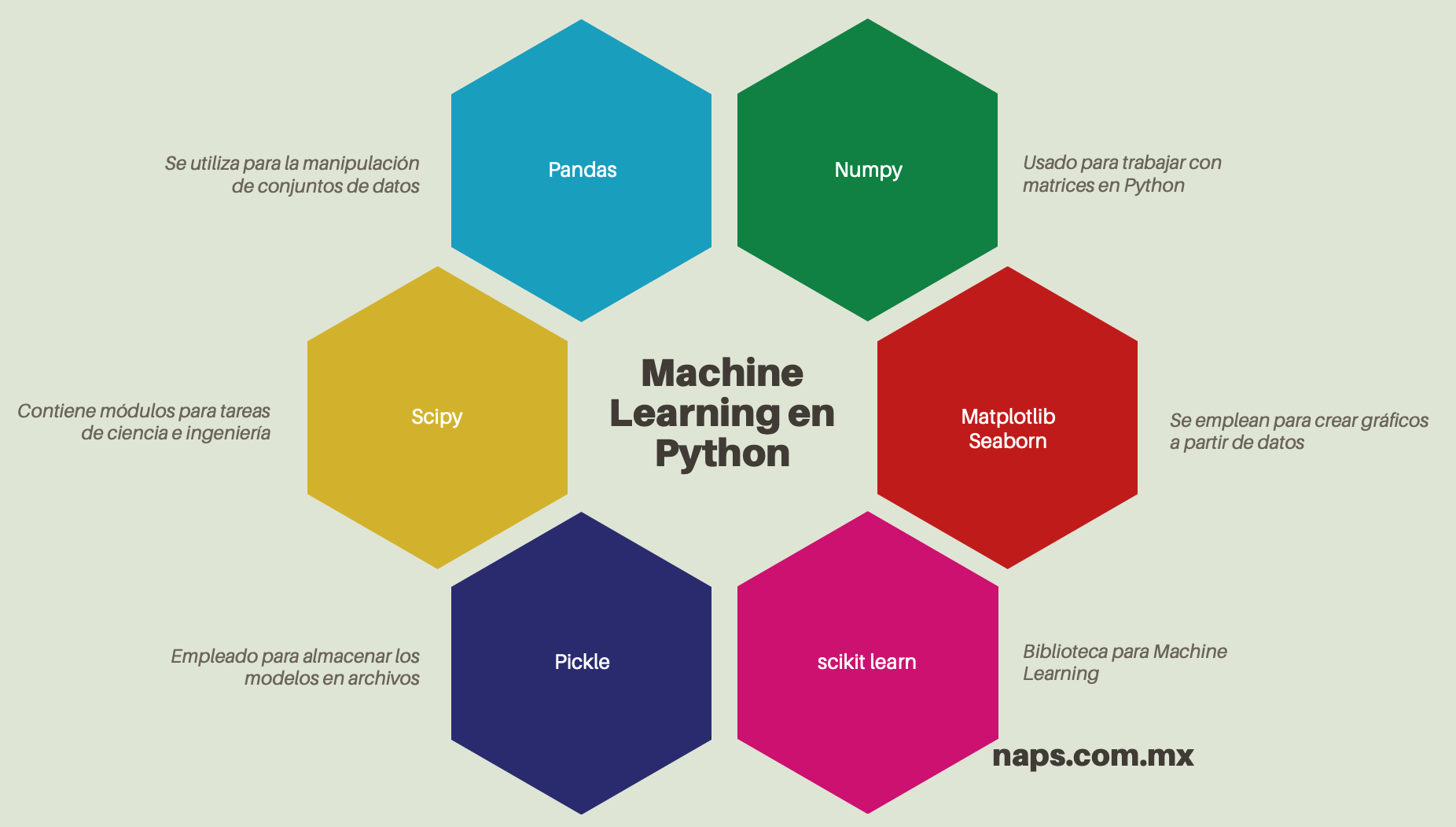
****

*Fuente: (Amat J., 2020)*

### *Python*

Para hacer el entrenamiento de modelo en machine learning, se recomiendan las siguientes librerías(García G. 2021):

1. **Pandas:** Se utiliza para obtener y manipular un conjunto de datos.
2. **Numpy:** Se emplea para trabajar el conjunto de datos en forma de matrices, esto para mejorar el rendimiento del modelo.
3. **Matplotlib / Seaborn:** Con esta librería se crean gráficos, en donde podemos ver el comportamiento de los datos y cómo se relacionan.
4. **Scikit Learn:** Esta es una librería principal de machine learning en Python, ya que en esta se encuentran algoritmos para entrenar el modelo.
5. **Pickle:** Este se utiliza para almacenar los modelos en un archivo y así poder utilizarlos más adelante si es necesario.
6. **Scipy:** Esta librería se emplea para realizar tareas de modelos referentes a ciencia y tecnología.



*Fuente:García G. 2021*

Ventajas de utilizar Python (Tokyo School, 2020):

* La curva de aprendizaje es corta, ya que es un lenguaje simple y rápido para programar.
* Amplia documentación,
* Es un lenguaje ordenado y posee la característica de código limpio.
* Posee una amplia comunidad la cual ha ido creciendo.

## Pronóstico de Kansas con Machine Learning

### Explicación de los datos

Los datos usados para el análisis provienen de la estación cerca del aeropuerto intencional de la ciudad de Kansas, los cuales al provenir de una zona cercana a un aeropuerto contiene un cantidad de datos considerables de diversas condiciones climáticas.

Los datos provenientes de la base de datos son los siguientes:

**WSF2:** Velocidad del viento máxima cada 2 minutos.

**FMTM:** Tiempo de viento más rápido de 5 segundos o más rápido que de un minuto.

**WSF5:** velocidad máxima cada 5 segundos.

**SNOW:** nevada.

**PRCP:** Precipitaciones(lluvias).

**SNWD:** Profundidad de la nieve.

**WDF2:**  Dirección máxima del viento cada 5 minutos.

**AWND:** Promedio de la velocidad del viento.

**WDF5:** Dirección máxima del viento cada 5 segundos.

**PGTM:** Tiempo de rafagas.

**TMAX:**  Temperatura máxima.

**TAVG:** Promedio de la temperatura.

**TMIN:** Temperatura mínima.

A partir de los datos mencionados anteriormente se desarrolla un modelo de pronóstico usando las herramientas que python proporciona.

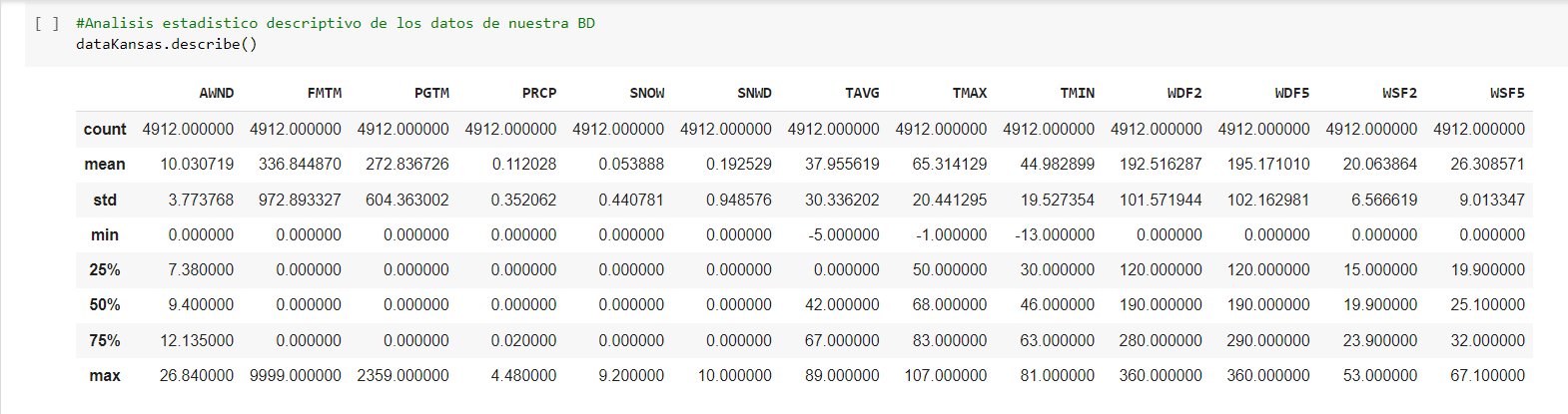
### Desarrollo de pronóstico

Lo primero qué haremos es importar las librerías necesarias para poder trabajar, en este caso utilizaremos Pandas, Numpy qué son dos de las librerías más usadas para el análisis de datos. Luego de ello importaremos los datos a analizar y lo mostraremos, en este caso, es un Excel de datos del Aeropuerto de la ciudad de Kansas.

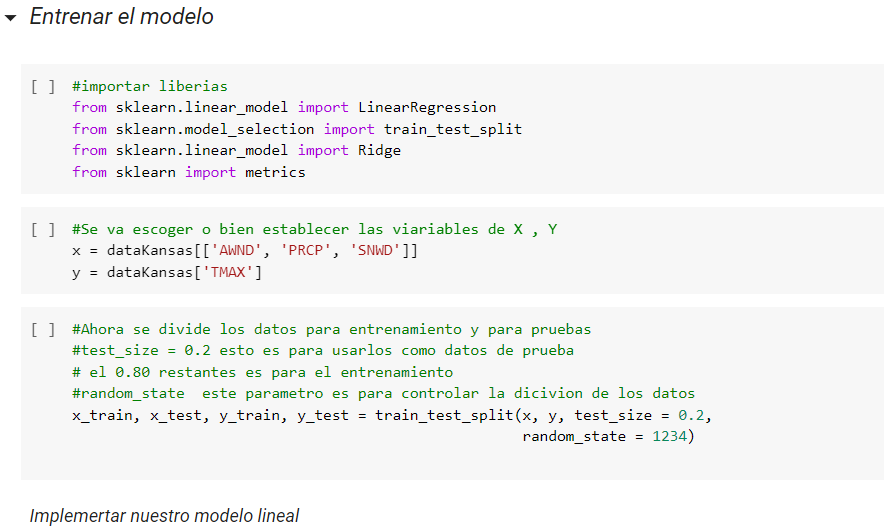


A continuación, debemos saber de todas las columnas, cuáles contiene datos ausentes, lo cual se realizará con el método isna().

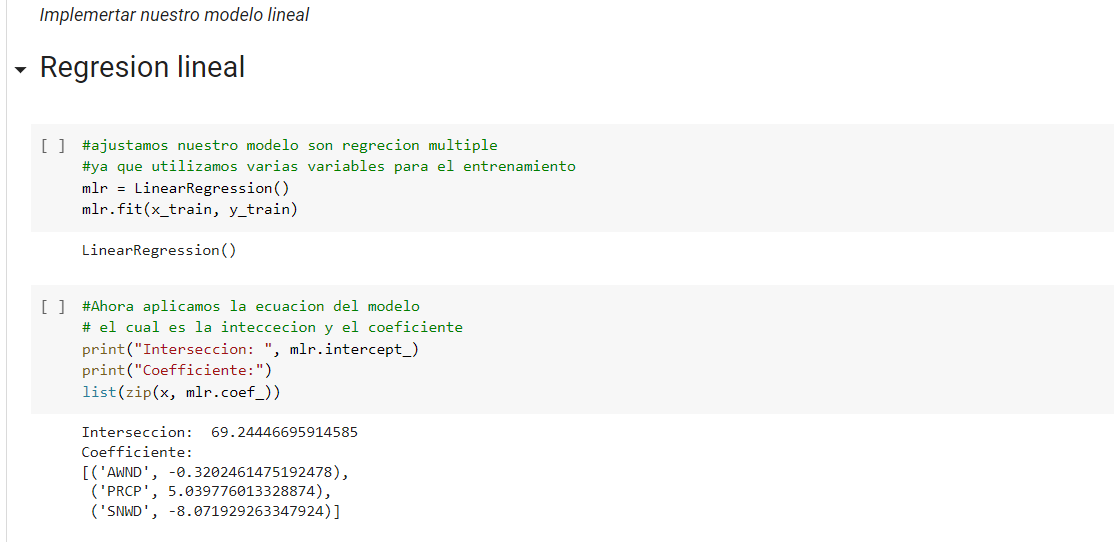


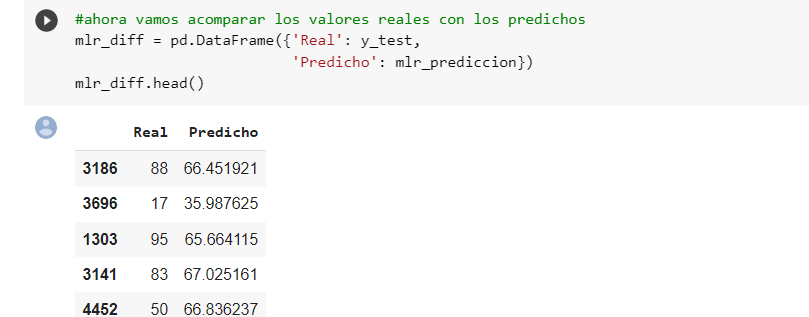
Una vez eliminados los datos ausentes, vamos a realizar un análisis estadístico descriptivo de los datos. 

Luego entrenaremos el modelo para poder crear un pronóstico, creando el modelo por medio scikit-learn. Para ello se importan las librerías necesarias y designamos las variables X & Y para las columnas a pronosticar.



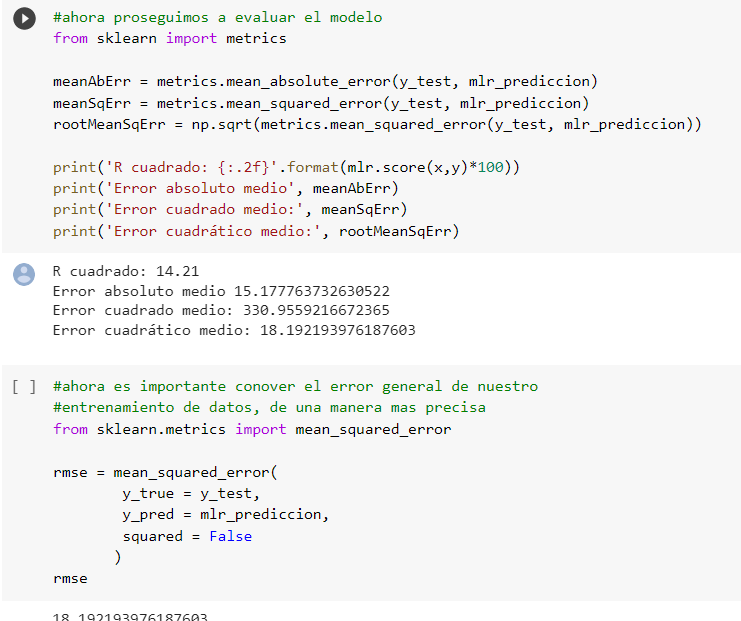
Continuando, implementaremos el algoritmo de regresión lineal para el pronóstico de los datos. Una vez apliquemos la fórmula para la predicción de nuestras pruebas, se imprimirá los datos de la predicción.



El siguiente paso es comparar los datos reales con los datos qué a predicho nuestro modelo. 

## 

Y por último, pero no menos importante, se realizarán los pasos para analizar el margen de error con el que cuenta las predicciones realizadas por nuestro modelo.



### Análisis general del entrenamiento de Kansas con machine learning

***R Squared:*** El valor de R Square es 14.21 esto nos indica siempre que esté entre 0% y 100% se ajusta al modelo de regresión lineal, por lo tanto, en este caso podemos concluir que el r cuadrado se ajusta al modelo.

**El error absoluto medio:** Esta métrica nos dice que es la diferencia absoluta entre los valores reales y los valores pronosticados, cuando menor sea este valor o más se acerca a cero, nos indica que mejor el rendimiento de nuestro modelo, el error absoluto medio obtenido es de 0.15 el cual está bastante bien.

**El error cuadrático cuadrado:** Este valor lo obtenemos cuando se calcula con el promedio del cuadrado de la diferencia obtenida entre los valores originales y el pronosticado del modelo. Este de igual manera entre más cerca de cero se encuentre nos indica un buen rendimiento en el caso del entrenamiento de Kansas nos da 3.30, lo cual no es muy óptimo.

**El error cuadrático medio:** Este se calcula con la desviación estándar que se genera de la predicción de nuestro conjunto de datos. Este tiene la misma característica que mientras más cerca del cero del resultado mejor es nuestro análisis del rendimiento. El error cuadrático medio obtenido es de 0.18, el cual es muy óptimo.

**Análisis de los resultados:** Se infiere que el resultado del entrenamiento en Kansas es regular, ya que los resultados se encuentran bastante bien.

**Análisis de la ecuación del modelo de regresión:**

En este procedimiento, se puede inferir que en la intersección nos da un valor de 69,24, lo cual nos indica que si se mantuviera AWND, PRCP Y SNWD en cero el promedio estimado para TMAX es de 69,29.

En cambio, si se presenta un aumento en TMAX se presentaría un pronóstico para:

* AWND: -0.3202461475192478
* PRCP: 5.039776013328874
* SNWD: -8.071929263347924

# 

# Conclusión

Para concluir con esta investigación sobre el pronóstico del clima en el estado de Kansas, se abarcaron temas sobre IA, los beneficios que pueden brindar en el estudio de datos y como el entrenamiento del Machine Learning puede generar pronósticos adecuados y precisos. La automatización de procesos utiliza estas herramientas en el análisis con el objetivo de reducir el tiempo y costos, en donde la simulación de los datos generan patrones y así poder obtener predicciones.

Normalmente, los pronósticos son usados para predecir un evento que podría ocurrir de acuerdo a los patrones presentados con antelación entre más información es analizada más precisa va a ser la aproximación, en nuestro caso surge de la prevención de cualquier inconveniente climático, conociendo ciertos parámetros se puede mitigar este tipo de eventos.

Estas predicciones o pronósticos son de gran importancia para el aeropuerto de Kansas y alrededores, con el estudio realizado de los factores climáticos en la ciudad de Kansas, los pronósticos obtenidos serán de gran ayuda en la toma de decisiones del aeropuerto con respecto a los vuelos y aterrizajes que se realizan.

En los cuales se podrá plantear una serie de mitigaciones de acuerdo a tormentas o mal clima, de igual forma esta información se puede ajustar con el objetivo de obtener beneficios del buen clima.

# Bibliografía

Amat Rodrigo, J. (2020, October 30). *Regresión lineal con python*. Cienciadedatos.net.

Retrieved June 13, 2022, from <https://www.cienciadedatos.net/documentos/py10-regresion-lineal-python.html>

Barraza, J. R. V., Rangel, H. R., González-Huitrón, V., Lara-Alvarez, C. A., & Flores, J. J.

(2019). Pronóstico de la velocidad del viento mediante técnicas de Inteligencia Artificial. Res. Comput. Sci., 148(7), 181-195.

FUENTES-COVARRUBIAS, R., FUENTES-COVARRUBIAS, A. G., CORTESQUIROZ, J.

A., & DE JESUS-JUAREZ, J. G. (2018). Sistema basado en conocimiento para la predicción del clima para usos agrícolas Knowledge Based system for predicting climate for agriculture uses. Cómputo Aplicado, 2(8), 1-11.

Noriega, H. A. M. (2021). Evaluación de técnicas de Inteligencia Artificial para el pronóstico

de inundaciones costeras por penetración del mar. Serie Científica de la Universidad de las Ciencias Informáticas, 14(9), 16-27.

Martinez, J. (2020, October 2). *Regresión Lineal: teoría y ejemplos en Python*.

IArtificial.net. Retrieved June 13, 2022, from <https://www.iartificial.net/regresion-lineal-con-ejemplos-en-python/>

*Pronósticos Climáticos | CIIFEN*. (2021, abril). CIIFEN. Recuperado 20 de junio de 2022, de

<https://ciifen.org/pronosticos-climaticos/>

Garcia, G. (2021, Septiembre 4). Introducción al machine learning en Python. Naps

Tecnologia y Educacion. Recuperado 22 de junio de 2022, de <https://naps.com.mx/blog/introduccion-al-machine-learning-en-python/>

Tokyo School ( 2020, Noviembre 24). Fundamentos del Machine learning con Python.

Tokyo School. Recuperado 23 de junio de 2022, de <https://www.tokioschool.com/noticias/fundamentos-machine-learning-python/>

Amat, J. (2020, Octubre). Machine learning con Python y Scikit-learn. Ciencia de Datos.

Recuperado 23 de junio de 2022, de <https://www.cienciadedatos.net/documentos/py06_machine_learning_python_scikitlearn.html>

Importancia de la información y predicción climática para el beneficio de la población. (2009, 30 junio). PreventionWeb. <https://www.preventionweb.net/news/importancia-de-la-informacion-y-prediccion-climatica-para-el-beneficio-de-la-poblacion#:%7E:text=La%20informaci%C3%B3n%20y%20las%20predicciones%20clim%C3%A1ticas%20sirven%20como%20base%20para,cient%C3%ADfica%20para%20planificar%20sus%20actividades>.

Th, R. (2020, 18 febrero). *Las 7 ventajas que ofrecen las predicciones meteorológicas y el Big Data*. TecnoHotel. <https://tecnohotelnews.com/2015/04/las-7-ventajas-que-ofrecen-las-predicciones-meteorologicas-y-el-big-data/>

Weather Spark. (2022, 25 junio). *El clima en Kansas City, el tiempo por mes, temperatura promedio (Kansas, Estados Unidos)* Recuperado de <https://es.weatherspark.com/y/9825/Clima-promedio-en-Kansas-City-Kansas-Estados-Unidos-durante-todo-el-a%C3%B1o>

*Center for Artificial Intelligence and Data Science*. (s. f.). Center for Artificial Intelligence and Data Science. Recuperado 25 de junio de 2022, de <https://caids.cs.ksu.edu/>